Journal of Classification 20:221-255 (2003)
DOI: 10.1007/s00357-003-0013-5

Maximum Likelihood Estimation and Model Comparison for Mixtures of Structural Equation Models with Ignorable Missing Data

Sik-Yum Lee

The Chinese University of Hong Kong

Xin-Yuan Song

The Chinese University of Hong Kong, and Zhongshan University, China

Abstract: The objective of this paper is to develop the maximum likelihood approach for analyzing a finite mixture of structural equation models with missing data that are missing at random. A Monte Carlo EM algorithm is proposed for obtaining the maximum likelihood estimates. A well-known statistic in model comparison, namely the Bayesian Information Criterion (BIC), is used for model comparison. With the presence of missing data, the computation of the observed-data likelihood function value involved in the BIC is not straightforward. A procedure based on path sampling is developed to compute this function value. It is shown by means of simulation studies that ignoring the incomplete data with missing entries gives less accurate ML estimates. An illustrative real example is also presented.

Keywords: Mixture; MAR missing data; MCEM algorithm; Gibbs sampler; Path sampling; BIC.

This research is fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, CUHK 4243/03H. The authors are grateful to reviewers for valuable comments on the earlier draft of the paper, to ICPSR and the relevant funding agency for allowing the use of the data, and to Esther Tam for preparation of the manuscript.

Author's Address: Prof. S.Y. Lee, Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong; Facs.: 852-26037251; e-mail: sylee@sparc2.sta.cuhk.edu.hk.

1. Introduction

In general, mixture models arise in many contexts and have been found to be useful for modeling heterogeneity (Pettit & Smith, 1985), and density estimation (Roeder & Wasserman, 1997). As analysis of mixture models is non-trivial, it has received a lot of attention in statistics (see, McLachlan & Barford, 1998; Titterington et al., 1985; among others). Structural equation models (SEMs) are well-recognized to be very useful in behavioral, psychological and social sciences. As the data often come from a heterogeneous group of populations, developments of rigorous methods for analyzing a finite mixture of SEMs have received a great deal of attention. Arminger and Stein (1997) used a two-stage method to analyze mixtures of conditional distributions with covariance structures. Jedidi, Jagpal and DeSarbo (1997a) analyzed a finite mixture of multivariate regression and simultaneous equation models; while Jedidi, Jagpal and DeSarbo (1997b) considered the estimation of a general finite mixture of structural equation models. Yung (1997) investigated a mixture of confirmatory factor analysis models and proposed an approximated scoring algorithm and an EM algorithm to solve the likelihood equation. Dolan and van der Maas (1998) used a quasi-Newton algorithm and inferred the estimation by changing the degree of separation and the sample sizes. Arminger, Stein and Wittenberg (1999) discussed ML analysis for mixtures of conditional mean- and covariance- structure models. Three estimation strategies on the basis of the EM algorithm were proposed. Zhu and Lee (2001) developed a Bayesian analysis to a finite mixture of LISREL models, using the idea of augmenting the observed data with latent variables and allocation variables. However, very limited work has been done to handle mixture of SEMs with missing data.

In practice, missing data are very common in behavioral, psychological and social researches. In general, it is well-recognized that the information of the missing data should be taken into account for achieving correct results. The approach of replacing the missing entries by estimates obtained from the sample means or the predicted values by regression on the basis of the fully observed data creates dependent observations which are very difficult to handle. Moreover, for mixture models, as the component memberships of the observations are not identified, one does not know which part of the data should be used to compute the mean estimates or the predicted values from regression. Existing specific methods in common SEMs with missing data proposed by Lee (1986), Allison (1987), Jamshidiam and Bentler (1999), and Song and Lee (2002) are not for analyzing mixture SEMs.

In this paper, a maximum likelihood (ML) approach is developed for analyzing mixtures of SEMs with missing data which are missing at random (MAR) with an ignorable mechanism (see, Little & Rubin, 1987). We will study the impact of ignoring incomplete data on estimation. We consider the

ML approach because it is well-recognized as an important statistical method and has optimal properties such as consistency, efficiency and asymptotic normality. Moreover, it is the foundation of many important statistical methods, for example, the Bayesian Information Criterion (BIC) for model comparison (see, Schwarz, 1978; Kass & Raftery , 1995), and global and local influence analyses for model diagnostics (see, Cook, 1977, 1986). We do not pursue a Bayesian approach (Diebolt & Robert, 1994) because it involves a great deal of effort in programming the full conditional distributions in the posterior analysis corresponding to the large number of unknown parameters in the model, see Zhu and Lee (2001). However, it is not our intention to claim that the ML approach is better or to compare it with the Bayesian approach. In the literature, there are many papers to compare these two important approaches, and it seems that a definite conclusion is not yet arrived. Like other areas in statistics, our purpose is to develop the ML approach as a complementary method to the Bayesian approach.

It is well-known that the EM algorithm (Dempster et al. 1977) is a powerful tool for ML estimation with missing data. Due to the complexities of the mixture SEMs, it is difficult and inefficient to directly compute the conditional expectations in the E-step of the algorithm. The Gibbs sampler will be implemented to generate observations from the conditional distributions for approximating these expectations and the E-step. Hence, it can be regarded as a Monte Carlo EM (MCEM) algorithm (Wei & Tanner, 1990). The M-step is completed by a sequence of conditional maximization steps (see, Meng & Rubin, 1993), and convergence of the algorithm is monitored by the recent method given in Shi and Copas (2002). Standard errors estimates are evaluated by the Louis (1982) formula. In situations where there are a large number of distinct missing patterns associated with the missing data, direct computation of the final observed-data log-likelihood function, which is essential in obtaining the BIC, is very tedious and inefficient. A procedure on the basis of path sampling (Gelman & Meng, 1998) is introduced to compute this value.

The paper is organized as follows. Section 2 presents a finite mixture of SEMs with missing data. The ML estimation is developed in Section 3. Components in the implementation of the MCEM algorithm are discussed. A procedure for computing the final observed-data likelihood via path sampling, and model comparison on the basis of BIC are discussed in Section 4. Results obtained from simulation studies are reported in Section 5. Section 6 presents an illustrative example with a real data set, and Section 7 gives a summary.

2. Model Description

A mixture SEMs for a $p \times 1$ random vector $\mathbf{y}_i$ is defined as follows:

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad i = 1, \cdots, n, \tag{1}$$

where K is the number of components, π_k's are component probabilities which are nonnegative and sum to 1.0, $f_k(\mathbf{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is a multivariate normal density function with an unknown mean vector $\boldsymbol{\mu}_k$ and a covariance matrix $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\sigma}_k)$ that is a matrix function of an unknown parameter vector $\boldsymbol{\sigma}_k$.

Suppose that the random vector $\mathbf{y}$ conditional on the *k*-th component satisfies the following measurement model:

$$\mathbf{y} = \boldsymbol{\mu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\omega}_k + \boldsymbol{\epsilon}_k, \tag{2}$$

where $\boldsymbol{\mu}_k$ is an $p \times 1$ intercept vector, $\boldsymbol{\Lambda}_k$ is a $p \times q$ factor loading matrix, $\boldsymbol{\omega}_k$ is a $q \times 1$ random vector of latent variables, and $\boldsymbol{\epsilon}_k$ is a $p \times 1$ random vector of error measurements with distribution $N(\mathbf{0}, \boldsymbol{\Psi}_k)$, which is independent of $\boldsymbol{\omega}_k$, and $\boldsymbol{\Psi}_k$ is a diagonal matrix. To handle more complex situations, the latent vector $\boldsymbol{\omega}_k$ is partitioned into $(\boldsymbol{\eta}_k', \boldsymbol{\xi}_k')'$, where $\boldsymbol{\eta}_k$ is a $q_1 \times 1$ vector, $\boldsymbol{\xi}_k$ is a $q_2 \times 1$ vector, and $q_1 + q_2 = q$. The structural equation of this k-th component model is defined as (see, Jöreskog & Sörbom, 1996):

$$\boldsymbol{\eta}_k = \mathbf{B}_k \boldsymbol{\eta}_k + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k + \boldsymbol{\delta}_k, \tag{3}$$

where $\mathbf{B}_k$ and $\boldsymbol{\Gamma}_k$ are $q_1 \times q_1$ and $q_1 \times q_2$ matrices of unknown parameters such that $\mathbf{B}_{0k} = (\mathbf{I}_{q_1} - \mathbf{B}_k)$ is nonsingular and $|\mathbf{I}_{q_1} - \mathbf{B}_k|$ is a constant that does not depend on elements of $\mathbf{B}_k$. Random vectors $\boldsymbol{\xi}_k$ and $\boldsymbol{\delta}_k$ are independently distributed as $N(\mathbf{0}, \boldsymbol{\Phi}_k)$ and $N(0, \boldsymbol{\Psi}_{\delta k})$, respectively; and $\boldsymbol{\Psi}_{\delta k}$ is a diagonal matrix. The structural equation defined in (3) describes the casuals effects of $\boldsymbol{\eta}_k$ and $\boldsymbol{\xi}_k$ on $\boldsymbol{\eta}_k$. Let $\boldsymbol{\Pi}_k = (\mathbf{B}_k, \boldsymbol{\Gamma}_k)$, and $\boldsymbol{\sigma}_k$ be the parameter vector that contains the free unknown parameters in $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Psi}_k$, $\boldsymbol{\Pi}_k$, $\boldsymbol{\Psi}_{\delta k}$ and $\boldsymbol{\Phi}_k$. The covariance structure of $\boldsymbol{\omega}_k$ is given by

$$\boldsymbol{\Sigma}_{\omega k} = \begin{bmatrix} \mathbf{B}_{0k}^{-1}(\boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_k' + \boldsymbol{\Psi}_{\delta k})(\mathbf{B}_{0k}^{-1})' & \mathbf{B}_{0k}^{-1}\boldsymbol{\Gamma}_k \boldsymbol{\Phi}_k \\ \boldsymbol{\Phi}_k \boldsymbol{\Gamma}_k'(\mathbf{B}_{0k}^{-1})' & \boldsymbol{\Phi}_k \end{bmatrix}, \tag{4}$$

and $\boldsymbol{\Sigma}_k(\boldsymbol{\sigma}_k) = \boldsymbol{\Lambda}_k \boldsymbol{\Sigma}_{\omega k} \boldsymbol{\Lambda}_k' + \boldsymbol{\Psi}_k$. Any of these unknown parameter matrices can be set invariant across components.

As the mixture model defined in (1) is invariant with respect to permutation of labels $k = 1, \cdots, K$, adoption of an unique labeling for identifiability is important. On the basis of the suggestions by Crawford (1994), Roeder and Wasserman (1997), and Zhu and Lee (2001), the ordering $\mu_{11} < \cdots < \mu_{K1}$ is imposed to eliminate label switching, where μ_{k1} is the first element of the mean

vector $\boldsymbol{\mu}_k$. For each $k = 1, \cdots, K$, structural parameters in the covariance matrix $\boldsymbol{\Sigma}_k$ corresponding to the model defined by (2) and (3) are not identified. A common method in structural equation modeling for identifying the covariance matrix is to fix appropriate elements in $\boldsymbol{\Lambda}_k$, $\mathbf{B}_k$, and/or $\boldsymbol{\Gamma}_k$ at preassigned values that are chosen on problem-by-problem basis. See the following sections for more concrete examples. For clear discussion of the proposed method, we let $\boldsymbol{\theta}$ be the vector which contains all unknown parameters in π_k, $\boldsymbol{\mu}_k$, $\boldsymbol{\Lambda}_k$, $\boldsymbol{\Psi}_k$, $\mathbf{B}_k$, $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Psi}_{\delta k}$, $k = 1, \cdots, K$, that define an identified model.

3. ML Estimation of Mixture SEMs with Missing Data

To deal with the missing data problem, let $\mathbf{y}_i = \{\mathbf{y}_{i,\text{obs}}, \mathbf{y}_{i,\text{mis}}\}$, where $\mathbf{y}_{i,\text{obs}}$ represents the observed components of $\mathbf{y}_i$, whilst $\mathbf{y}_{i,\text{mis}}$ represents the missing entries. We assume that missing data are MAR with an ignorable mechanism (Little & Rubin, 1987). For a fully observed data point $\mathbf{y}_i$, $\mathbf{y}_{i,\text{mis}}$ does not exist. Let $\mathbf{Y}_{\text{obs}} = \{\mathbf{y}_{i,\text{obs}}, i = 1, \cdots, n\}$, $\mathbf{Y}_{\text{mis}} = \{\mathbf{y}_{i,\text{mis}}, i = 1, \cdots, n\}$, and $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$. Since for each $i = 1, \cdots, n$, the number of missing entries and their positions in $\mathbf{y}_i$ can be varied, there may be a lot of missing patterns in $\mathbf{Y}_{\text{obs}}$; moreover, the sample sizes in the patterns can be very small and very different. As very limited work has been done to analyze mixtures of SEM with missing data, the main objective of this paper is to develop a procedure for obtaining the ML estimates of $\boldsymbol{\theta}$ and standard errors on the basis of the observed-data $\mathbf{Y}_{\text{obs}}$.

3.1 Standard Method of Treating Missing Data to Mixture SEMs

In the standard method of dealing missing data in covariance structure modeling (Finkbeiner, 1979; Lee, 1986), for each $i = 1, \cdots, n$, $\mathbf{y}_{i,\text{obs}}$ is related with $\mathbf{y}_i$ by $\mathbf{y}_{i,\text{obs}} = \mathbf{P}_i \mathbf{y}_i$, where $\mathbf{P}_i$ is an appropriate matrix of zeros and ones. If $\mathbf{y}_i$ contains no missing data, $\mathbf{P}_i$ is the identity matrix. The observed data are equal to $\mathbf{Y}_{\text{obs}} = \{\mathbf{P}_i \mathbf{y}_i, i = 1, \cdots, n\}$. Applying this standard method to a K-component mixture SEM as defined in (1),

$$f(\mathbf{y}_{i,\text{obs}}) = \sum_{k=1}^{K} \pi_k f_k(\mathbf{P}_i \mathbf{y}_i | \mathbf{P}_i \boldsymbol{\mu}_k, \mathbf{P}_i \boldsymbol{\Sigma}_k \mathbf{P}_i'),$$

and the observed-data log-likelihood function becomes

$$L_o(\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_k f_k(\mathbf{P}_i \mathbf{y}_i | \mathbf{P}_i \mu_k, \mathbf{P}_i \boldsymbol{\Sigma}_k \mathbf{P}_i').$$

As 'log' cannot be interchanged with the summation sign, this function can only be expressed as a sum of n density functions of non-identically distributed

observations with different dimensions. To apply gradient methods such as the approximate-scoring algorithm (Yung, 1997) or the quasi-Newton algorithm (Dolan & van der Maas, 1998) in maximizing $L_o(\mathbf{Y}_{\rm obs}, \boldsymbol{\theta})$, we have to compute the gradient vector of $L_o(\mathbf{Y}_{\rm obs}, \boldsymbol{\theta})$, which involves n first derivatives corresponding to $\log f(\mathbf{y}_{i,\rm obs}), i = 1, \cdots, n$, in each iteration. The computational burden is rather heavy if n is large. Moreover, corresponding to different data sets, $\mathbf{P}_i, i = 1, \cdots, n$ can be different. From an implementation point of view, it is rather difficult to write a program that can handle general missing data sets with arbitrary missing patterns.

To reduce the computational burden, the standard EM algorithm (Dempster et al., 1977) has been applied to mixture SEMs with fully observed data (Yung, 1997; Jedidi et al., 1997). In the application, a grouping variable (indicator variable, component membership) w_i for $\mathbf{y}_i$ is introduced as a latent allocation variable. Naturally, it assumed that w_i is independently drawn from the following distribution

$$p(w_i = k) = \pi_k, \quad k = 1, \cdots, K, \tag{5}$$

and given w_i, observations are drawn independently from the respective subpopulations. In the EM algorithm of Jedidi et al. (1997), only $\mathbf{W} = (w_1, \cdots, w_n)$ are treated as missing data. Let $L_{1c}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta})$ be the complete-data likelihood. At the r-th iteration of the EM algorithm, the E-step evaluates $Q_1(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\theta}^{(r)}) = E\{L_{1c}(\mathbf{W}, \mathbf{Y}; \boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta}^{(r)}\}$, where the expectation is taken with respect to the conditional distribution of $\mathbf{W}$ given $\mathbf{Y}$ and $\boldsymbol{\theta}^{(r)}$. The M-step maximizes $Q_1(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\theta}^{(r)})$ with respect to $\boldsymbol{\theta}$. The objective function, $Q_1(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\theta}^{(r)})$, to be maximized in the M-step can be expressed as

$$Q_1(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\theta}^{(r)}) = \sum_{k=1}^{K} Q_{1k}(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k|\mathbf{Y}_k, \boldsymbol{\mu}_k^{(r)}, \boldsymbol{\Sigma}_k^{(r)}),$$

where Q_{1k} is the conditional expectation with respect to the k-th component likelihood and $\mathbf{Y}_k$ is the collection of observations from component k. The maximization is done via a common multiple-group approach with a quasi-Newton algorithm. The presence of missing data gives no trouble in evaluating $Q_1(\boldsymbol{\theta}|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)})$ in the E-step. However, to deal with the missing patterns, the multiple-group problem in the M-step will be more complicated. Let $D \leq min(2^p, n)$ be the number of distinct missing patterns, and $\mathbf{P}_d$ be the matrix as defined above to relate $\mathbf{y}_{i,\rm obs}$ with $\mathbf{y}_i$ in the pattern d, for $d = 1, \cdots, D$. The objective function to be maximized at the M-step is given by

$$\begin{aligned} &Q_1(\boldsymbol{\theta}|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)}) \\ &\quad = \textstyle\sum_{d=1}^{D} \sum_{k=1}^{K} Q_{1k}(\pi_k, \mathbf{P}_d\boldsymbol{\mu}_k, \mathbf{P}_d\boldsymbol{\Sigma}_k\mathbf{P}_d'|\mathbf{Y}_{dk}, \mathbf{P}_d\boldsymbol{\mu}_k^{(r)}, \mathbf{P}_d\boldsymbol{\Sigma}_k^{(r)}\mathbf{P}_d'), \end{aligned}$$

where $\mathbf{Y}_{dk}$ contains the observed data points in the pattern d under component k. When D is large and/or the numbers of data points for some $\mathbf{Y}_{dk}$ are small, maximizing $Q_1(\boldsymbol{\theta}|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)})$ is not practical.

In applying the EM algorithm to a mixture of confirmatory factor analysis model (see equations (1) and (2)) with latent factors $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_n)$, Yung (1997) treated $\mathbf{W}$ and $\boldsymbol{\Omega}$ as missing data. Let $L_{2c}(\mathbf{Y}, \mathbf{W}, \boldsymbol{\Omega}; \boldsymbol{\theta})$ be the complete-data likelihood. At the E-step of r-th EM iteration, it requires to evaluate the conditional expectation, $E\{L_{2c}(\mathbf{Y}, \mathbf{W}, \boldsymbol{\Omega}; \boldsymbol{\theta})|\mathbf{Y}, \boldsymbol{\theta}^{(r)}\}$, with respect to $\mathbf{W}$ and $\boldsymbol{\Omega}$ given $\mathbf{Y}$ and $\boldsymbol{\theta}^{(r)}$. Yung (1997) derived expression for this conditional expectation in the context of the simple factor analysis model with fully observed data; and obtained closed form solution in the M-step under certain situations. For more general SEMs (see equations (2) and (3)) with missing data, $L_{2c}(\mathbf{Y}_{\rm obs}, \mathbf{W}, \boldsymbol{\Omega}; \boldsymbol{\theta})$ is a summation of D more complicated functions, and the conditional distribution of $\mathbf{W}$ and $\boldsymbol{\Omega}$ given $\mathbf{Y}_{\rm obs}$ (not $\mathbf{Y}$) and $\boldsymbol{\theta}^{(r)}$ is much more complicated. Hence, if D is large or the form of $\mathbf{Y}_{\rm obs}$ is complicated, evaluating $Q_2(\boldsymbol{\theta}|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)}) = E\{L_{2c}(\mathbf{Y}_{\rm obs}, \mathbf{W}, \boldsymbol{\Omega}; \boldsymbol{\theta})|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)}\}$ is difficult. Moreover, the form of $Q_2(\boldsymbol{\theta}|\mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)})$ can be complex, it is not clear whether or not the M-step can be completed via closed-form solution. Motivated by the above difficulties of the standard methods in treating the MAR missing data for mixture of NSEMs, we develop a MCEM algorithm for ML estimation of the model that is more efficient and reliable.

3.2 ML Estimation via a MCEM Algorithm

As the likelihood function $p(\mathbf{Y}_{\rm obs}|\boldsymbol{\theta})$ based on the observed data is rather complicated due to presence of missing data and latent variables, obtaining the ML solution by directly working with $p(\mathbf{Y}_{\rm obs}|\boldsymbol{\theta})$ or by applying the above mentioned standard methods is difficult. Inspired by the key idea of data augmentation (see, Tanner & Wong, 1987), we augment $\mathbf{Y}_{\rm obs}$ with $(\mathbf{Y}_{\rm mis}, \boldsymbol{\Omega})$ in the ML estimation. Let $\mathbf{X} = (\mathbf{Y}_{\rm obs}, \mathbf{Y}_{\rm mis}, \boldsymbol{\Omega})$, the complete-data likelihood function is:

$$\begin{aligned} L_c(\mathbf{X}; \boldsymbol{\theta}) &= p(\mathbf{X}|\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K} \pi_k (2\pi)^{-n(p+q)/2} (|\boldsymbol{\Psi}_k||\boldsymbol{\Psi}_{\delta k}||\boldsymbol{\Phi}_k|)^{-n/2} |\mathbf{B}_{0k}|^n \\ &\quad \prod_{i=1}^{n} \exp\left\{(\mathbf{y}_i - \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i)' \boldsymbol{\Psi}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k - \boldsymbol{\Lambda}_k \boldsymbol{\omega}_i)\right\} \times \\ &\quad \exp\left\{\boldsymbol{\xi}_i' \boldsymbol{\Phi}_k^{-1} \boldsymbol{\xi}_i + (\boldsymbol{\eta}_i - \boldsymbol{\Pi}_k \boldsymbol{\omega}_i)' \boldsymbol{\Psi}_{\delta k}^{-1} (\boldsymbol{\eta}_i - \boldsymbol{\Pi}_k \boldsymbol{\omega}_i)\right\}. \end{aligned} \quad (6)$$

Note that a key difference between Yung's (1997) approach and our approach

in using the EM algorithm is that we additionally augment $\mathbf{Y}_{\rm mis}$ in the analysis, so that our complete-data likelihood $L_c(\mathbf{X};\boldsymbol{\theta})$ does not involve any $\mathbf{P}_i$. Consequently, the ML estimation based on the complete-data set $\mathbf{X}$ is comparatively easier with the following EM algorithm: At the rth iteration with a current value $\boldsymbol{\theta}^{(r)}$:

E-step: Evaluate

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) &= E\{L_c(\mathbf{X},\boldsymbol{\theta})|\mathbf{Y}_{\rm obs},\boldsymbol{\theta}^{(r)}\} \\ &= E\{L_c(\mathbf{Y}_{\rm obs},\mathbf{Y}_{\rm mis},\boldsymbol{\Omega},\boldsymbol{\theta})|\mathbf{Y}_{\rm obs},\boldsymbol{\theta}^{(r)}\}, \end{aligned}$$

where the expectation is taken with respect to the joint conditional distribution of $(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega})$ given $\mathbf{Y}_{\rm obs}$ and $\boldsymbol{\theta}^{(r)}$.

M-step: Maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ to update $\boldsymbol{\theta}^{(r)}$ to $\boldsymbol{\theta}^{(r+1)}$.

Due to the existence of the missing data, it is still difficult to directly evaluate the conditional expectations involved in the E-step. Inspired by the idea given in Wei and Tanner (1990), the E-step is approximated by a sufficiently large number of observations simulated from the conditional distribution of $(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega})$ given $\mathbf{Y}_{\rm obs}$ and $\boldsymbol{\theta}^{(r)}$. A common tool in statistical computing for simulating observations from $p(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega}|\mathbf{Y}_{\rm obs},\boldsymbol{\theta}^{(r)})$ is the Gibbs sampler (Geman & Geman, 1984), which iteratively generates observations from $p(\mathbf{Y}_{\rm mis}|\mathbf{Y}_{\rm obs},\boldsymbol{\Omega},\boldsymbol{\theta}^{(r)})$ and $p(\boldsymbol{\Omega}|\mathbf{Y}_{\rm obs},\mathbf{Y}_{\rm mis},\boldsymbol{\theta})$. However, owing to the complexity of the current mixture model, the above marginal conditional distributions are complicated. Hence, utilizing the common idea in handling mixture models (see, for example Zhu & Lee, 2001; among others), we further augment the grouping variable w_i for $\mathbf{y}_i$ (see equation (5)) in the analysis. Now, the data set $(\mathbf{Y}_{\rm obs},\mathbf{Y}_{\rm mis},\boldsymbol{\Omega})$ will be further augmented with the vector $\mathbf{W}=(w_1,\cdots,w_n)$ in the ML analysis. In the following, we will briefly describe the Gibbs sampler to simulate observations from $p(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega},\mathbf{W}|\mathbf{Y}_{\rm obs},\boldsymbol{\theta})$ for approximating the conditional expectations in the E-step. Thus the proposed EM algorithm can be regarded as a Monte Carlo EM (MCEM) algorithm (Wei & Tanner, 1990).

Let $h(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega})$ be a generic function of $\mathbf{Y}_{\rm mis}$ and $\boldsymbol{\Omega}$. In the evaluation of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$, its conditional expectation given $\mathbf{Y}_{\rm obs}$ and $\boldsymbol{\theta}$ is approximated by

$$\hat{E}\{h(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega})|\mathbf{Y}_{\rm obs},\boldsymbol{\theta}\}=\frac{1}{T}\sum_{a=1}^{T}h(\mathbf{Y}_{\rm mis}^{(a)},\boldsymbol{\Omega}^{(a)}), \tag{7}$$

where $\{(\mathbf{Y}_{\rm mis}^{(a)},\boldsymbol{\Omega}^{(a)});a=1,\cdots,T\}$ is the corresponding subset of a sufficiently large sample $\{(\mathbf{Y}_{\rm mis}^{(a)},\boldsymbol{\Omega}^{(a)},\mathbf{W}^{(a)});a=1,\cdots,T\}$ simulated from the joint conditional distribution $p(\mathbf{Y}_{\rm mis},\boldsymbol{\Omega},\mathbf{W}|\mathbf{Y}_{\rm obs},\boldsymbol{\theta})$. The following Gibbs

sampler (Geman & Geman, 1984) is used to sample these observations. At the ath iteration within the Gibbs sampler with current values $\mathbf{Y}_{\text{mis}}^{(a)}$, $\boldsymbol{\Omega}^{(a)}$ and $\mathbf{W}^{(a)}$: (i) generate $\mathbf{W}^{(a+1)}$ from $p(\mathbf{W}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}^{(a)}, \boldsymbol{\Omega}^{(a)}, \boldsymbol{\theta})$, (ii) generate $\mathbf{Y}_{\text{mis}}^{(a+1)}$ from $p(\mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\Omega}^{(a)}, \mathbf{W}^{(a+1)}, \boldsymbol{\theta})$, and (iii) generate $\boldsymbol{\Omega}^{(a+1)}$ from $p(\boldsymbol{\Omega}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}^{(a+1)}, \mathbf{W}^{(a+1)}, \boldsymbol{\theta})$. Marginal conditional distributions for every iteration in the Gibbs sampler are briefly discussed as follows.

Consider the conditional distribution $p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta})$. As $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$ is given, the model becomes one without missing data. Hence, it follows from the results in the literature (see, for example, Zhu & Lee, 2001) that

$$p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\Omega}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(w_i|\mathbf{y}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}), \quad \text{and}$$

$$p(w_i = k|\mathbf{y}_i, \boldsymbol{\omega}_i, \boldsymbol{\theta}) = \frac{\pi_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}, \tag{8}$$

where $f_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the probability density function of $N[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k(\boldsymbol{\sigma}_k)]$.

As $\boldsymbol{\omega}_i$ and $\boldsymbol{\epsilon}_i$ are mutually independent, $\mathbf{y}_i$ are also mutually independent. We have

$$p(\boldsymbol{\Omega}|\mathbf{Y}, \mathbf{W}, \boldsymbol{\theta}) \propto p(\mathbf{Y}|\boldsymbol{\Omega}, \mathbf{W}, \boldsymbol{\theta}) p(\boldsymbol{\Omega}|\mathbf{W}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{y}_i|\boldsymbol{\omega}_i, w_i, \boldsymbol{\theta}) p(\boldsymbol{\omega}_i|w_i, \boldsymbol{\theta}).$$

Hence, the conditional distribution of $\boldsymbol{\Omega}$ can be obtained from

$$p(\boldsymbol{\omega}_i|\mathbf{y}_i, w_i = k, \boldsymbol{\theta}) \sim N[\ \mathbf{A}_k^{-1} \boldsymbol{\Lambda}_k' \boldsymbol{\Psi}_k^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_k), \mathbf{A}_k^{-1}\ ],$$

where $\mathbf{A}_k = \boldsymbol{\Sigma}_{\omega k}^{-1} + \boldsymbol{\Lambda}_k' \boldsymbol{\Psi}_k^{-1} \boldsymbol{\Lambda}_k$, and $\boldsymbol{\Sigma}_{\omega k}$ is given by (4).

For $i = 1, \cdots, n$, since $\mathbf{y}_i$ are mutually independent, $\mathbf{y}_{i,\text{mis}}$ are also mutually independent. Since $\boldsymbol{\Psi}_k$ is diagonal, $\mathbf{y}_{i,\text{mis}}$ is conditionally independent with $\mathbf{y}_{i,\text{obs}}$ given $\boldsymbol{\omega}_i$. Let p_i be the dimension of $\mathbf{y}_{i,\text{mis}}$, it follows from (1) that

$$p(\mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\Omega}, \mathbf{W}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(\mathbf{y}_{i,\text{mis}}|\boldsymbol{\omega}_i, w_i, \boldsymbol{\theta}), \quad \text{and}$$

$$[\mathbf{y}_{i,\text{mis}}|\boldsymbol{\omega}_i, w_i = k, \boldsymbol{\theta}] \stackrel{D}{=} N[\boldsymbol{\mu}_{i,\text{mis},k} + \boldsymbol{\Lambda}_{i,\text{mis},k} \boldsymbol{\omega}_i, \boldsymbol{\Psi}_{i,\text{mis},k}], \tag{9}$$

where $\boldsymbol{\mu}_{i,\text{mis},k}$ is a $p_i \times 1$ subvector of $\boldsymbol{\mu}_k$ with elements corresponding to observed components deleted, $\boldsymbol{\Lambda}_{i,\text{mis},k}$ is the corresponding $p_i \times q$ submatrix of $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\Psi}_{i,\text{mis},k}$ is the corresponding $p_i \times p_i$ submatrix of $\boldsymbol{\Psi}_k$ with the appropriate rows and columns deleted. Hence, even the form of $\mathbf{Y}_{\text{mis}}$ is complicated

with many distinct missing patterns, its conditional distribution only involves a product of very simple univariate normal distributions. The computational burden for simulating $\mathbf{Y}_{\rm mis}$ is light. If $\mathbf{\Psi}_{i,{\rm mis},k}$ is not diagonal, $\mathbf{y}_{i,{\rm mis}}$ is simulated from the multivariate normal distribution as given in (9), the computational burden is again light.

At the M-step, we need to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ with respect to $\boldsymbol{\theta}$. This is equivalent to solve the following system of equations:

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})}{\partial \boldsymbol{\theta}} = E\left\{ \frac{\partial}{\partial \boldsymbol{\theta}} L_c(\mathbf{X}, \boldsymbol{\theta}) \;\middle|\; \mathbf{Y}_{\rm obs}, \boldsymbol{\theta}^{(r)} \right\} = 0. \tag{10}$$

Let $\mathbf{Y}_k$ and $\mathbf{\Omega}_k$ be the submatrices of $\mathbf{Y}$ and $\mathbf{\Omega}$ respectively such that all the i-th columns with $w_i \neq k$ are deleted, n_k be the number of observations in the kth subpopulation, and $y_{ki(m)}$ be the mth entry of $\mathbf{y}_{ki}$, $\eta_{ki(j)}$ be the jth entry of $\boldsymbol{\eta}_{ki}$, μ_{km} is the mth element of $\boldsymbol{\mu}_k$, and $\mathbf{\Lambda}_{km}$ and $\mathbf{\Pi}_{kj}$ be the mth and the jth rows of $\mathbf{\Lambda}_k$ and $\mathbf{\Pi}_k$, respectively. For $k = 1, \cdots, K$, $m = 1, \cdots, p$, and $j = 1, \cdots, q$, it can be shown that

$$\begin{aligned}
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k} &= \mathbf{\Psi}^{-1} \sum_{i=1}^{n_k} (\mathbf{y}_{ki} - \boldsymbol{\mu}_k - \mathbf{\Lambda}_k \boldsymbol{\omega}_{ki}), \\
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial \mathbf{\Phi}_k} &= \frac{1}{2} \mathbf{\Phi}_k^{-1} \sum_{i=1}^{n_k} (\boldsymbol{\xi}_{ki} \boldsymbol{\xi}_{ki}' - \mathbf{\Phi}_k) \mathbf{\Phi}_k^{-1}, \\
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial \mathbf{\Lambda}_{km}} &= \psi_{km}^{-1} \sum_{i=1}^{n_k} (y_{ki}(m) - \mu_{km} - \mathbf{\Lambda}_{km} \boldsymbol{\omega}_{ki}) \boldsymbol{\omega}_{ki}', \\
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial \mathbf{\Pi}_{kj}} &= \psi_{k\delta j}^{-1} \sum_{i=1}^{n_k} [\eta_{ki(j)} - \mathbf{\Pi}_{kj} \boldsymbol{\omega}_{ki}] \boldsymbol{\omega}_{ki}', \\
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial {\rm diag}(\mathbf{\Psi}_k)} &= \\
&\frac{1}{2} {\rm diag} \left\{ \mathbf{\Psi}_k^{-1} \sum_{i=1}^{n_k} [(\mathbf{y}_{ki} - \boldsymbol{\mu}_k - \mathbf{\Lambda}_k \boldsymbol{\omega}_{ki})(\mathbf{y}_{ki} - \boldsymbol{\mu}_k - \mathbf{\Lambda}_k \boldsymbol{\omega}_{ki})' - \mathbf{\Psi}_k] \mathbf{\Psi}_k^{-1} \right\}, \\
\frac{\partial L_c(\mathbf{X}, \boldsymbol{\theta})}{\partial {\rm diag}(\mathbf{\Psi}_{k\delta})} &= \\
&\frac{1}{2} {\rm diag} \left\{ \mathbf{\Psi}_{k\delta}^{-1} \sum_{i=1}^{n_k} [(\boldsymbol{\eta}_{ki} - \mathbf{\Pi}_k \boldsymbol{\omega}_{ki})(\boldsymbol{\eta}_{ki} - \mathbf{\Pi}_k \boldsymbol{\omega}_{ki})' - \mathbf{\Psi}_{k\delta}] \mathbf{\Psi}_{k\delta}^{-1} \right\}.
\end{aligned} \tag{11}$$

These simultaneous equations cannot be solved in closed form. Based on the idea given in Meng and Rubin (1993), the solution required in the M-step can be

obtained by several computationally simpler conditional maximizations. Conditional on the other parameters, the solution of the individual equation given in (11) can be obtained as follows:

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_k &= \frac{1}{n_k}\sum_{i=1}^{n_k} E[(\mathbf{y}_{ki} - \boldsymbol{\Lambda}_k\boldsymbol{\omega}_{ki})|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}], \quad \hat{\boldsymbol{\Phi}}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} E[\boldsymbol{\xi}_{ki}\boldsymbol{\xi}'_{ki}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}], \\
\hat{\boldsymbol{\Lambda}}'_{km} &= (\sum_{i=1}^{n_k} E[\boldsymbol{\omega}_{ki}\boldsymbol{\omega}'_{ki}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}])^{-1}\sum_{i=1}^{n_k} E[\boldsymbol{\omega}_{ki}(y_{ki(m)} - \hat{\mu}_{km})|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}], \\
\hat{\boldsymbol{\Pi}}'_{kj} &= (\sum_{i=1}^{n_k} E[\boldsymbol{\omega}_{ki}\boldsymbol{\omega}'_{ki}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}])^{-1}\sum_{i=1}^{n_k} E[\boldsymbol{\omega}_{ki}\eta_{ki(j)}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}], \qquad (12) \\
\hat{\psi}_{km} &= \frac{1}{n_k}\sum_{i=1}^{n_k} E[(y_{ki(m)} - \hat{\mu}_{km} - \hat{\boldsymbol{\Lambda}}_{km}\boldsymbol{\omega}_{ki})^2|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}], \\
\hat{\psi}_{k\delta j} &= \frac{1}{n_k}\sum_{i=1}^{n_k} E[(\eta_{ki(j)} - \boldsymbol{\Pi}_{kj}\boldsymbol{\omega}_{ki})^2|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}].
\end{aligned}$$

The conditional expectations involved in (12) are approximated by the corresponding generated observations at E-step, see equation (7). Hence, the M-step can be completed via closed form solution via (12).

3.3 Stopping Rules and Standard Errors

Selection of the sample size T in the Monte Carlo E-step is an important issue in the implementation of the MCEM algorithm. To decrease the Monte Carlo error at the E-step, we need to generate a sufficiently large number of observations, see (7). However, it is inefficient to start with a large T when $\hat{\boldsymbol{\theta}}^{(r)}$ is far from the ML estimate. Hence, it has been suggested that T should be increased from one iteration to the next (see, Wei & Tanner, 1990; McCulloch, 1994; Booth & Hobert, 1999). According to Shi and Copas (2002), we use a new scheme that uses average values of $\boldsymbol{\theta}^{(r)}$ but without iteratively increasing T. Specifically, we calculate the following average 'batch mean' at the rth iteration of the MCEM algorithm:

$$\bar{\boldsymbol{\theta}}^{(r)} = \frac{1}{J}(\boldsymbol{\theta}^{(r-J+1)} + \cdots + \boldsymbol{\theta}^{(r)}), \qquad (13)$$

after the Jth iteration. For a sufficiently large J, the average of the Monte Carlo errors in equation (13) becomes negligible, and Shi and Copas (2002) argued that the MCEM algorithm is converging when the batch means $\bar{\boldsymbol{\theta}}^{(r)}$ are stabilized. They recommended the following standard stopping rule in terms of

the batch means for monitoring convergence: The procedure is stopped when the absolute difference $||\bar{\boldsymbol{\theta}}^{(r)} - \bar{\boldsymbol{\theta}}^{(r-\gamma_0)}||$ or the relative difference

$$\frac{||\bar{\boldsymbol{\theta}}^{(r)} - \bar{\boldsymbol{\theta}}^{(r-\gamma_0)}||}{||\bar{\boldsymbol{\theta}}^{(r-\gamma_0)}|| + \delta_0} \tag{14}$$

is smaller than some predetermined small value δ (e.g. 0.001). In expression (14), δ_0 is selected to ensure that the denominator is not too close to zero. To avoid the danger of stopping too early because of the correlation between successive elements of the Markov chain, a large value of γ_0, such as 5, is suggested. Also, convergence is claimed after the stopping rule is satisfied for several consecutive iterations. At convergence, say at iteration r, the ML estimate is taken to be $\bar{\boldsymbol{\theta}}^{(r)}$. According to Shi and Copas (2002), the overall procedure is roughly equivalent to using the MCEM algorithm with Monte Carlo sample size JT, although in fact the actual sample size is just T for all the iterations.

As a by-product, the standard error of $\hat{\boldsymbol{\theta}}$ can also be calculated quite easily for the MCEM algorithm. Louis (1982) showed that the observed information matrix of $\hat{\boldsymbol{\theta}}$ is

$$\mathbf{H} = -E\{\ddot{\mathbf{L}}(\mathbf{X}, \boldsymbol{\theta})|\mathbf{Y}_{\rm obs}\} - E\{\dot{\mathbf{L}}(\mathbf{X}, \boldsymbol{\theta})\dot{\mathbf{L}}'(\mathbf{X}, \boldsymbol{\theta})|\mathbf{Y}_{\rm obs}\},$$

where $\dot{\mathbf{L}}$ and $\ddot{\mathbf{L}}$ are the first two derivatives of complete-data log-likelihood function $L_c(\mathbf{X}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which can be obtained by straightforward matrix calculus. For the MCEM algorithm, $\mathbf{H}$ can be estimated from the random samples generated. In the rth iteration, an estimate of $\mathbf{H}$ is equal to

$$\hat{\mathbf{H}}^{(r)} = \frac{1}{T}\sum_{a=1}^{T} \mathbf{H}^{(r)}(\mathbf{Y}_{\rm obs}, \mathbf{Y}_{\rm mis}^{a,(r)}, \boldsymbol{\Omega}^{a,(r)}) = \frac{1}{T}\sum_{a=1}^{T} \mathbf{H}^{(r)}(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{a,(r)})$$

where $\mathbf{Z}^{a,(r)} = (\mathbf{Y}_{\rm mis}^{a,(r)}, \boldsymbol{\Omega}^{a,(r)})$, and

$$\mathbf{H}^{(r)}(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{a,(r)}) = \\ -\ddot{\mathbf{L}}(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{a,(r)}, \boldsymbol{\theta}^{(r)}) - \dot{\mathbf{L}}(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{a,(r)}, \boldsymbol{\theta}^{(r)})\dot{\mathbf{L}}(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{a,(r)}, \boldsymbol{\theta}^{(r)})'.$$

When the algorithm is stopped, say at the rth iteration, the final estimate of the information matrix can be estimated by the following average

$$\bar{\mathbf{H}}^{(r)} = \frac{1}{J}(\hat{\mathbf{H}}^{(r-J+1)} + \cdots + \hat{\mathbf{H}}^{(r)}). \tag{15}$$

4. Model Comparison for Mixture of SEMs

Suppose the observed data $\mathbf{Y}_{\rm obs}$ have arisen under two competing mixture SEMs M_h and M_k with different number of components. For $v = h, k$,

let $\boldsymbol{\theta}_v$ be the underlying parameter vector and $p(\mathbf{Y}_{\text{obs}}|\boldsymbol{\theta}_v)$ be the likelihood function of $\mathbf{Y}_{\text{obs}}$ given M_v. The following BIC is used for comparing two competitive models M_1 and M_2:

$$\text{BIC}_{hk} = -2[\log p(\mathbf{Y}_{\text{obs}}|\hat{\boldsymbol{\theta}}_h, M_h) - \log p(\mathbf{Y}_{\text{obs}}|\hat{\boldsymbol{\theta}}_k, M_k)] + (d_h - d_k)\log n, \tag{16}$$

where $\hat{\boldsymbol{\theta}}_v$ is the ML estimate of $\boldsymbol{\theta}_v$ under model M_v, and d_v is the dimension of $\boldsymbol{\theta}_v$. The interpretation of the BIC for model comparison is given in Kass and Raftery (1995) as an approximation of the Bayes factor. In the application of the BIC to model comparison of the mixture of SEMs with missing data, we need to evaluate the observed-data log-likelihood $\log p(\mathbf{Y}_{\text{obs}}|\hat{\boldsymbol{\theta}}_v)$ at ML estimates $\hat{\boldsymbol{\theta}}_v$ under some specific model M_v. Owing to the presence of missing data and latent variables, the computation of this final observed-data log-likelihood value is non-trivial. The main objective in this section is to establish a method for computing these final values associated with specific models of interest, or their ratios so that the BIC can be evaluated for model comparison.

The development is based on the key idea in path sampling (Gelmen & Meng, 1998) for computing (ratios) of normalizing constants of probability models. Recall that $\mathbf{Z} = (\mathbf{Y}_{\text{mis}}, \boldsymbol{\Omega})$. As $p(\mathbf{Z}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}) = p(\mathbf{Z}, \mathbf{Y}_{\text{obs}}|\boldsymbol{\theta})$ $/p(\mathbf{Y}_{\text{obs}}|\boldsymbol{\theta})$, the observed-data likelihood can be treated as the normalizing constant of $p(\mathbf{Z}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta})$ with the complete-data likelihood as the unnormalized density. Consider the following class of densities with a continuous parameter t in [0,1]:

$$z(t) = \int p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|t, \boldsymbol{\theta}) d\mathbf{Z}, \tag{17}$$

where $p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|t, \boldsymbol{\theta})$ is a density function for each t, $p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|b, \boldsymbol{\theta}) =$ $p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|M_b, \boldsymbol{\theta})$ with $p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|M_b, \boldsymbol{\theta})$ denotes the complete-data likelihood under any model M_b, $b = 0, 1$. Hence, the observed-data likelihood under M_b is given by

$$z(b) = \int p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|M_b, \boldsymbol{\theta}) d\mathbf{Z} = p(\mathbf{Y}_{\text{obs}}|M_b, \boldsymbol{\theta}), \quad b = 0, 1.$$

Taking logarithm and then differentiating both sides of (17) with respect to t, it can be shown that (see, Gelman & Meng, 1998)

$$\frac{d\log z(t)}{dt} = \text{E}\left[\frac{d\log p(\mathbf{Y}_{obs}, \mathbf{Z}|t, \boldsymbol{\theta})}{dt}\right], \tag{18}$$

where E denotes the expectation with respect to the sampling distribution $p(\mathbf{Z}|\mathbf{Y}_{\text{obs}}, t, \boldsymbol{\theta})$. Integrating (18) from 0 to 1, we have

$$\lambda = \log\left[\frac{z(1)}{z(0)}\right] = \int_0^1 \text{E}[U(\mathbf{Y}_{\text{obs}}, \mathbf{Z}, t, \boldsymbol{\theta})]dt, \tag{19}$$

where $U(\mathbf{Y}_{\rm obs}, \mathbf{Z}, t, \boldsymbol{\theta}) = d \log p(\mathbf{Y}_{\rm obs}, \mathbf{Z}, t, \boldsymbol{\theta})/dt$. An estimate of λ can be obtained by numerical evaluation of the above integral over t, via the method given in Ogata (1989). Specifically, we select fixed grids $\{t_{(s)}; s = 0, \cdots, S\}$ such that $t_{(0)} = 0 < t_{(1)} < \cdots < t_{(S+1)} = 1$, and compute λ by

$$\lambda = \frac{1}{2} \sum_{s=1}^{S} (t_{(s+1)} - t_{(s)})(\bar{U}_{(s+1)} + \bar{U}_{(s)}), \tag{20}$$

where $\bar{U}_{(s)} = J_1^{-1} \sum_{j=1}^{J_1} U(\mathbf{Y}_{\rm obs}, \mathbf{Z}^{(j)}, t_{(s)}, \boldsymbol{\theta})$, in which $\{\mathbf{Z}^{(j)}, j = 1, \cdots, J_1\}$ are simulated observations from $p(\mathbf{Z}|\mathbf{Y}_{\rm obs}, t_{(s)}, \boldsymbol{\theta})$. It follows from (19) that

$$\log z(1) = \lambda + \log z(0). \tag{21}$$

Hence, if we can link M_1 with a M_0 whose observed-data likelihood $z(0)$ can be computed easily, we can obtain the observed-data likelihood under M_1 via (21).

To give a more specific illustration, we apply the method to compute the observed-data likelihood of a mixture of SEMs. Let M_1 be the model defined in (1)-(3), and

$$\begin{aligned} M_0: \quad & f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_k \tilde{f}_k(\mathbf{y}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad \text{where} \\ & \mathbf{y}_i|_{w_i=k} = \boldsymbol{\mu}_k + \boldsymbol{\epsilon}_k, \ \ i = 1, \cdots, n. \end{aligned}$$

Hence, in the kth component, $\tilde{f}_k$ is the probability density function of the simple distribution $N[\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k]$. The linked model M_t is defined by equation (1), $\mathbf{y} = \boldsymbol{\mu}_k + t\boldsymbol{\Lambda}_k\boldsymbol{\omega} + \boldsymbol{\epsilon}_k$ and equation (3). It corresponds to $z(t) = \int p(\mathbf{Y}_{\rm obs}, \mathbf{Z}|\boldsymbol{\theta}, t) d\mathbf{Z}$, and $t \in [0, 1]$ is defined to specify the linked model M_t for accounting the change of $\boldsymbol{\Lambda}_k$ to $t\boldsymbol{\Lambda}_k$, for $k = 1, \cdots, K$. As a result,

$$z(1) = \int p(\mathbf{Y}_{\rm obs}, \mathbf{Z}|\boldsymbol{\theta}, 1) d\mathbf{Z} = p(\mathbf{Y}_{\rm obs}|\boldsymbol{\theta}, M_1),$$

which is the observed-data likelihood corresponding to M_1; and $z(0) = \int p(\mathbf{Y}_{\rm obs}, \mathbf{Z}|\boldsymbol{\theta}, 0) d\mathbf{Z} = p(\mathbf{Y}_{\rm obs}|\boldsymbol{\theta}, M_0)$ is the observed-data likelihood corresponding to the simple model M_0.

The complete-data log-likelihood function for M_t can be written as

$$L_c(\mathbf{X}, \boldsymbol{\theta}, t) = \log p(\mathbf{Y}_{\rm obs}, \mathbf{Y}_{\rm mis}, \boldsymbol{\Omega}|\boldsymbol{\theta}, t) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i, \boldsymbol{\omega}_i|\boldsymbol{\theta}_k, t) \right\}. \tag{22}$$

By differentiation with respect to t, we have

$$U(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \boldsymbol{\Omega}, \boldsymbol{\theta}, t) = \sum_{i=1}^{n} \frac{\sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i, \boldsymbol{\omega}_i|\boldsymbol{\theta}_k, t)(\mathbf{y}_i - \boldsymbol{\mu}_k - t\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i)'\boldsymbol{\Psi}_k^{-1}\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i}{\sum_{k=1}^{K} \pi_k f_k(\mathbf{y}_i, \boldsymbol{\omega}_i|\boldsymbol{\theta}_k, t)} \quad (23)$$

where

$$\begin{aligned} &f_k(\mathbf{y}_i, \boldsymbol{\omega}_i|\boldsymbol{\theta}_k, t) = \\ &(2\pi)^{-p/2}|\boldsymbol{\Psi}_k|^{-1/2} \exp\{-\tfrac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k - t\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i)'\boldsymbol{\Psi}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k - t\boldsymbol{\Lambda}_k\boldsymbol{\omega}_i)\} \\ &\times(2\pi)^{-q_1/2}|\mathbf{I}_{q_1} - \mathbf{B}_k||\boldsymbol{\Psi}_{\delta k}|^{-1/2} \exp\{-\tfrac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\Pi}_k\boldsymbol{\omega}_i)'\boldsymbol{\Psi}_{\delta k}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\Pi}_k\boldsymbol{\omega}_i)\} \\ &\times(2\pi)^{-q_2/2}|\boldsymbol{\Phi}_k|^{-1/2}\exp\{-\tfrac{1}{2}\boldsymbol{\xi}_i'\boldsymbol{\Phi}_k^{-1}\boldsymbol{\xi}_i\}. \end{aligned}$$

Thus, the observed-data log-likelihood ratio $\lambda = \log[z(1)/z(0)]$ can be estimated via (20) and (23). Also, the observed-data log-likelihood, $\log p(\mathbf{Y}_{\text{obs}}|\boldsymbol{\theta}, M_1) = \log z(1)$, can be obtained via (21). Note that

$$z(0) = \int p(\mathbf{Y}_{\text{obs}}, \mathbf{Z}|\boldsymbol{\theta}, 0)d\mathbf{Z} = \int p(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}|\boldsymbol{\theta})p(\mathbf{Y}_{\text{mis}}|\boldsymbol{\theta})d\mathbf{Y}_{\text{mis}}. \quad (24)$$

It is still not in close form. The following Monte Carlo approximation is used for obtaining an estimate:

$$z(0) = \frac{1}{J_2}\sum_{j=1}^{J_2} p(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}^{(j)}|\boldsymbol{\theta}), \quad (25)$$

where $\{\mathbf{Y}_{\text{mis}}^{(j)}, j = 1, \cdots, J_2\}$ is simulated from $p(\mathbf{Y}_{\text{mis}}|\boldsymbol{\theta})$ by a similar Gibbs sampler as discussed in section 3.1. As the underlying distribution corresponding to M_0 is a finite mixture of the simple $N[\boldsymbol{\mu}_k, \boldsymbol{\Psi}_k]$, simulating $\mathbf{Y}_{\text{mis}}^{(j)}$ is straightforward and fast.

For the competitive model M_2 which is a finite mixture of SEMs with h components, where $h < K$, we can use the above procedure to compute $\log p(\mathbf{Y}|\boldsymbol{\theta}, M_2)$. The BIC for comparing M_1 and M_2 can be obtained via (16) with the observed-data log-likelihoods evaluated at their ML estimates.

5. Simulation Studies

In this section, results of some simulation studies are presented to illustrate the proposed procedure in ML estimation for a mixture SEMs with missing data. The objective is to reveal the accuracy of the ML estimates obtained by our approach using the complete data (COM Data) that contain the fully observed data and the data with missing entries; and those obtained by the listwise

deletion approach that only use the fully observed data (FO Data). The bad effect of ignoring the data with missing entries to the accuracy of estimation is noted.

5.1 Simulation Study 1

The data set in this simulation study is generated from a mixture of SEMs with two components defined by (1), (2) and (3). For each $k = 1, 2$, the model involves six manifest variables which are indicators for three latent variables η, ξ_1 and ξ_2. The structure of the loading matrix in each component is

$$\mathbf{\Lambda}' = \begin{bmatrix} 1.0 & \lambda_{k,21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & \lambda_{k,42} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 & \lambda_{k,63} \end{bmatrix},$$

where the one's and zero's are fixed parameters for achieving an identified covariance structure, whilst $\lambda_{k,21}$, $\lambda_{k,42}$ and $\lambda_{k,63}$ will be treated as unknown parameters which can have different values in different components. In the k-th component, the structural equation that relates η, ξ_1, and ξ_2 is of the following form

$$\eta = \gamma_{k,1}\xi_1 + \gamma_{k,2}\xi_2 + \delta,$$

where $\gamma_{k,1}$ and $\gamma_{k,2}$ will be treated as unknown parameters. The true population values of the unknown parameters are given by $\pi_1 = \pi_2 = 0.5$, $\boldsymbol{\mu}_1 = (0, \cdots, 0)'$, $\boldsymbol{\mu}_2 = (2.0, \cdots, 2.0)'$; $\lambda_{1,21} = \lambda_{1,42} = \lambda_{1,63} = 0.4$, $\lambda_{2,21} = \lambda_{2,42} = \lambda_{2,63} = 0.8$, $\gamma_{1,1} = \gamma_{1,2} = 0.6$, $\gamma_{2,1} = 0.6$, $\gamma_{2,2} = -0.6$, $\phi_{1,11} = \phi_{1,22} = \phi_{2,11} = \phi_{2,22} = 1.0$, $\phi_{1,12} = \phi_{2,12} = 0.3$, $\psi_{1,11} = \cdots = \psi_{1,66} = \psi_{1,\delta} = 0.36$, $\psi_{2,11} = \cdots = \psi_{2,66} = \psi_{2,\delta} = 0.64$. In this 2-component mixture SEMs, the total number of unknown parameters is 44.

Based on the true population values of the parameters and the model defined by (2) and (3), we simulate 400 observations from each of the first and the second component, hence the total sample size is 800. The MAR missing data are created via the following steps: (i) G fully observed data points are randomly selected among the 800 observations, and sample means $\bar{y}_{(1)}$, $\bar{y}_{(2)}$, $\bar{y}_{(3)}$, $\bar{y}_{(4)}$, $\bar{y}_{(5)}$, and $\bar{y}_{(6)}$ are computed on the basis of these data points. (ii) In each and every of the remaining $(800 - G)$ observations, we decide whether its elements $y_{(1)}$, $y_{(2)}$, $y_{(3)}$, $y_{(4)}$, $y_{(5)}$ and/or $y_{(6)}$ are missing or not by randomly generating an observation v from $N[0, 1]$. More specifically, we randomly generate independent observations $v_{(1)}$, $v_{(2)}$, $v_{(3)}$, $v_{(4)}$, $v_{(5)}$ and $v_{(6)}$ from $N[0, 1]$, then $y_{(1)}$ is deleted only if $v_{(1)} > \bar{y}_{(1)} + \bar{y}_{(2)}$, $y_{(2)}$ is deleted only if $v_{(2)} < \bar{y}_{(1)} - \bar{y}_{(2)}$, $y_{(3)}$ is deleted only if $v_{(3)} > \bar{y}_{(3)} + \bar{y}_{(4)}$, $y_{(4)}$ is deleted only if $v_{(4)} < \bar{y}_{(3)} - \bar{y}_{(4)}$, $y_{(5)}$ is deleted only if $v_{(5)} > \bar{y}_{(5)}$, and $y_{(6)}$ is deleted only if $v_{(6)} > \bar{y}_{(6)}$. Hence,

the probability of response depends on the fully observed data points and the missing data are MAR. In this missing data set, a number of $y_{(1)}$, $y_{(2)}$, $y_{(3)}$, $y_{(4)}$, $y_{(5)}$ and/or $y_{(6)}$ are missing at random. To account for different missing proportions, $G = 100$, and 200 are considered.

For each G, ML estimates are obtained using: (A) all the data before the creation of missing data (ALL Data), (B) fully observed data and the incomplete data with missing entries (COM Data), and (C) only the fully observed data (FO Data). In this and all other simulation studies, we set $T = 50$, $J = 50$, and take the stopping rule with $\gamma_0 = 5$, $\delta_0 = 0.001$ and $\delta = 0.001$ (see Section 3.2). The algorithm is terminated after the stopping rule is satisfied for five successive iterations. For example, for the first replication of case (B), the iterations stopped at the 163rd iteration. So, $\bar{\boldsymbol{\theta}}^{(163)}$ is taken as ML estimate in that case. ML estimates in 100 replications in each of the cases (A), (B) and (C) are obtained. Based on 100 replications, the mean of ML estimates (Mean), and root mean squares (Rms) between estimates and true values are computed. Results obtained under (A), (B) and (C) are reported in Tables 1, 2 and 3.

5.2 Simulation Study 2

In this study, a 3-component mixture SEM is considered. The structure of the SEM is the same as in simulation study 1, with the same population values of parameters in the first and second components be given as before, and $\pi_1 = 0.4$, $\pi_2 = 0.3$, and $\pi_3 = 0.3$. The true values of the parameters in the third component are: $\mu_3 = (-2.0, \cdots, -2.0)$, $\lambda_{3,21} = \lambda_{3,42} = \lambda_{3,63} = 0.6$, $\gamma_{3,1} = -0.6$, $\gamma_{3,2} = 0.6$, $\phi_{3,11} = \phi_{3,22} = 1.0$, $\phi_{3,21} = 0.3$, $\psi_{3,11} = \psi_{3,22} = \psi_{3,33} = 0.36$, $\psi_{3,44} = \psi_{3,55} = \psi_{3,66} = 0.64$, and $\psi_{3,\delta} = 0.36$. In this 3-component model, the number of unknown parameters is 66. We simulate 400, 300 and 300 observations respectively from the first, second and third component; hence the total sample size is 1000. To get two missing proportions, $G = 360$ and 180; and the procedure for creating MAR data is the same as in simulation study 1. ML estimates for cases (A), (B) and (C) are obtained by the MCEM algorithm as before. Results obtained under (A) are presented in Table 4. To save space, only results obtained under (B) and (C) with $G = 360$ are reported in Tables 5 and 6. The performances of the estimates obtained with $G = 180$ are slightly worse.

5.3 Simulation Study 3

In this study, a 2-component mixture SEM with two latent variables and a different structure is considered. For each $k = 1, 2$, the structure of the loading

matrix is

$$\mathbf{\Lambda}' = \begin{bmatrix} 1.0 & \lambda_{k,21} & \lambda_{k,31} & 0 & 0 & \lambda_{k,61} & 0 \\ 0 & 0 & 0 & 1.0 & \lambda_{k,52} & \lambda_{k,62} & \lambda_{k,72} \end{bmatrix}.$$

The latent factors are denoted by η and ξ, which are related by the following structural equation $\eta = \gamma_{k,1}\xi + \delta$, where $\gamma_{k,1}$ is an unknown parameter and the unknown variances of ξ and δ are denoted by $\phi_{k,11}$ and $\psi_{k,\delta}$, respectively. The true population values of the unknown parameters are given by: $\pi_1 = \pi_2 = 0.5$; $\boldsymbol{\mu}_1 = (-1.0, \cdots, -1.0)$, $\boldsymbol{\mu}_2 = (1.0, \cdots, 1.0)$; $\lambda_{1,21} = \lambda_{1,31} = 0.5$, $\lambda_{1,52} = 0.7$, $\lambda_{1,61} = 0.3$, $\lambda_{1,62} = \lambda_{1,72} = 0.7$, $\lambda_{2,21} = \lambda_{2,31} = 0.7$, $\lambda_{2,52} = 0.5$, $\lambda_{2,61} = 0.3$, $\lambda_{2,62} = \lambda_{2,72} = 0.5$; $\gamma_{1,1} = 0.6$, $\gamma_{2,1} = -0.6$; $\phi_{1,11} = \phi_{2,11} = 1.0$; $\psi_{1,11} = \cdots = \psi_{1,77} = 0.25$, $\psi_{2,11} = \cdots = \psi_{2,77} = 0.36$; $\psi_{1,\delta} = 0.25$, and $\psi_{2,\delta} = 0.36$. The number of unknown parameters in this model is 48. The total sample size is taken to be 600, with 300 observations are simulated from each component; and $G = 200$, and 100. ML estimates obtained under cases (A), (B) and (C) are reported in Tables 7, 8 and 9 respectively.

5.4 Conclusion of the Simulation Results

Based on the results reported in Tables 1-9, we can draw the following major conclusions: (i) The ML estimates obtained by the proposed method on the basis of the 'COM Data' that utilize all the available fully observed and incomplete data are quite accurate. Comparing Table 1 with 2, 4 with 5, as well as 7 with 8, we see that the empirical performances are quite close to those obtained from using 'ALL Data' in the ideal case. (ii) Comparing Table 2 with 3, 5 with 6, and 8 with 9, we observe that the empirical performances of the proposed approach are better than those of the listwise deletion approach that only uses the fully observed data. (iii) It seems that the above phenomena are not changed with different missing proportions, different number of components, and different structures of the SEMs. To save space, more detailed discussions are not presented.

6. An Illustrative Example

To provide a real example for illustrating the proposed method, a small portion of ICPSR data set collected in the project WORLD VALUES SURVEY 1981-1984 AND 1990-1993 (World Values Study Group, ICPSR Version) is analyzed. The whole data set was collected in 45 societies around the world on various topics relating to work, attitude towards competition, the meaning and purpose of life, family life, religious belief, interest in politics, contemporary social issues; and many others. As an illustration of our proposed method, only the data obtained from the West Germany are used. Eight variables in the

Table 1: Mean and root mean squares of ML estimates obtained with 'ALL Data':

Simulation Study 1

Component 1			Component 2		
Par	Mean	Rms	Par	Mean	Rms
$\pi_1 = 0.5$	0.50	0.02	$\pi_2 = 0.5$	0.50	0.02
$\mu_{1,1} = 0.0$	-0.01	0.08	$\mu_{2,1} = 2.0$	2.01	0.07
$\mu_{1,2} = 0.0$	-0.01	0.05	$\mu_{2,2} = 2.0$	2.01	0.06
$\mu_{1,3} = 0.0$	-0.01	0.07	$\mu_{2,3} = 2.0$	2.01	0.07
$\mu_{1,4} = 0.0$	-0.00	0.04	$\mu_{2,4} = 2.0$	2.00	0.07
$\mu_{1,5} = 0.0$	-0.01	0.07	$\mu_{2,5} = 2.0$	1.99	0.07
$\mu_{1,6} = 0.0$	-0.01	0.04	$\mu_{2,6} = 2.0$	1.99	0.06
$\lambda_{1,21} = 0.4$	0.40	0.05	$\lambda_{2,21} = 0.8$	0.81	0.10
$\lambda_{1,42} = 0.4$	0.40	0.05	$\lambda_{2,42} = 0.8$	0.81	0.11
$\lambda_{1,63} = 0.4$	0.41	0.06	$\lambda_{2,63} = 0.8$	0.82	0.10
$\gamma_{1,1} = 0.6$	0.60	0.07	$\gamma_{2,1} = 0.6$	0.60	0.11
$\gamma_{1,2} = 0.6$	0.62	0.09	$\gamma_{2,2} = -0.6$	-0.61	0.09
$\phi_{1,11} = 1.0$	0.99	0.12	$\phi_{2,11} = 1.0$	1.00	0.19
$\phi_{1,12} = 0.3$	0.29	0.07	$\phi_{2,12} = 0.3$	0.29	0.08
$\phi_{1,22} = 1.0$	0.98	0.14	$\phi_{2,22} = 1.0$	0.99	0.16
$\psi_{1,11} = 0.36$	0.35	0.10	$\psi_{2,11} = 0.64$	0.65	0.14
$\psi_{1,22} = 0.36$	0.36	0.03	$\psi_{2,22} = 0.64$	0.63	0.08
$\psi_{1,33} = 0.36$	0.36	0.09	$\psi_{2,33} = 0.64$	0.63	0.13
$\psi_{1,44} = 0.36$	0.36	0.03	$\psi_{2,44} = 0.64$	0.63	0.09
$\psi_{1,55} = 0.36$	0.38	0.10	$\psi_{2,55} = 0.64$	0.63	0.12
$\psi_{1,66} = 0.36$	0.36	0.03	$\psi_{2,66} = 0.64$	0.64	0.08
$\psi_{1,\delta} = 0.36$	0.36	0.10	$\psi_{2,\delta} = 0.64$	0.64	0.12

Table 2: Mean and root mean squares of ML estimates obtained with 'COM Data' :

Simulation Study 1

Component 1					Component 2				
	G = 200		G = 100			G = 200		G = 100	
Par	Mean	Rms	Mean	Rms	Par	Mean	Rms	Mean	Rms
$\pi_1 = 0.5$	0.49	0.03	0.48	0.03	$\pi_2 = 0.5$	0.51	0.03	0.52	0.03
$\mu_{1,1} = 0.0$	-0.04	0.09	-0.06	0.10	$\mu_{2,1} = 2.0$	1.98	0.07	1.98	0.07
$\mu_{1,2} = 0.0$	-0.04	0.07	-0.05	0.08	$\mu_{2,2} = 2.0$	2.02	0.08	2.03	0.09
$\mu_{1,3} = 0.0$	-0.02	0.07	-0.04	0.08	$\mu_{2,3} = 2.0$	1.96	0.10	1.97	0.10
$\mu_{1,4} = 0.0$	-0.03	0.05	-0.04	0.06	$\mu_{2,4} = 2.0$	2.01	0.08	2.01	0.09
$\mu_{1,5} = 0.0$	-0.05	0.09	-0.07	0.10	$\mu_{2,5} = 2.0$	2.00	0.07	1.98	0.08
$\mu_{1,6} = 0.0$	-0.02	0.05	-0.03	0.06	$\mu_{2,6} = 2.0$	1.98	0.08	1.97	0.08
$\lambda_{1,21} = 0.4$	0.36	0.06	0.36	0.06	$\lambda_{2,21} = 0.8$	0.80	0.11	0.80	0.10
$\lambda_{1,42} = 0.4$	0.37	0.06	0.37	0.07	$\lambda_{2,42} = 0.8$	0.73	0.12	0.71	0.14
$\lambda_{1,63} = 0.4$	0.39	0.06	0.41	0.07	$\lambda_{2,63} = 0.8$	0.82	0.13	0.82	0.12
$\gamma_{1,1} = 0.6$	0.59	0.10	0.59	0.11	$\gamma_{2,1} = 0.6$	0.57	0.12	0.55	0.11
$\gamma_{1,2} = 0.6$	0.61	0.10	0.63	0.12	$\gamma_{2,2} = -0.6$	-0.60	0.10	-0.59	0.11
$\phi_{1,11} = 1.0$	1.02	0.15	1.02	0.18	$\phi_{2,11} = 1.0$	1.09	0.21	1.11	0.23
$\phi_{1,12} = 0.3$	0.26	0.09	0.25	0.11	$\phi_{2,12} = 0.3$	0.32	0.10	0.30	0.09
$\phi_{1,22} = 1.0$	0.93	0.17	0.90	0.19	$\phi_{2,22} = 1.0$	0.99	0.19	0.96	0.17
$\psi_{1,11} = 0.36$	0.34	0.10	0.31	0.11	$\psi_{2,11} = 0.64$	0.65	0.15	0.67	0.16
$\psi_{1,22} = 0.36$	0.37	0.04	0.38	0.05	$\psi_{2,22} = 0.64$	0.70	0.12	0.70	0.12
$\psi_{1,33} = 0.36$	0.34	0.12	0.32	0.12	$\psi_{2,33} = 0.64$	0.58	0.15	0.55	0.18
$\psi_{1,44} = 0.36$	0.38	0.05	0.37	0.04	$\psi_{2,44} = 0.64$	0.72	0.13	0.74	0.13
$\psi_{1,55} = 0.36$	0.38	0.10	0.41	0.14	$\psi_{2,55} = 0.64$	0.64	0.14	0.65	0.13
$\psi_{1,66} = 0.36$	0.36	0.03	0.36	0.03	$\psi_{2,66} = 0.64$	0.67	0.11	0.67	0.11
$\psi_{1,\delta} = 0.36$	0.36	0.11	0.39	0.12	$\psi_{2,\delta} = 0.64$	0.63	0.17	0.62	0.16

Table 3: Mean and root mean squares of ML estimates obtained with 'FO Data' :

Simulation Study 1

Component 1					Component 2				
	G = 200		G = 100			G = 200		G = 100	
Par	Mean	Rms	Mean	Rms	Par	Mean	Rms	Mean	Rms
$\pi_1 = 0.5$	0.50	0.03	0.49	0.04	$\pi_2 = 0.5$	0.50	0.03	0.51	0.04
$\mu_{1,1} = 0.0$	-0.00	0.11	-0.01	0.17	$\mu_{2,1} = 2.0$	2.00	0.11	2.01	0.15
$\mu_{1,2} = 0.0$	-0.00	0.07	0.00	0.09	$\mu_{2,2} = 2.0$	2.02	0.11	1.98	0.15
$\mu_{1,3} = 0.0$	0.01	0.11	-0.02	0.14	$\mu_{2,3} = 2.0$	2.00	0.13	2.01	0.14
$\mu_{1,4} = 0.0$	0.00	0.06	-0.01	0.08	$\mu_{2,4} = 2.0$	2.01	0.12	2.01	0.14
$\mu_{1,5} = 0.0$	0.01	0.10	0.01	0.14	$\mu_{2,5} = 2.0$	2.01	0.12	1.99	0.14
$\mu_{1,6} = 0.0$	-0.00	0.07	-0.01	0.08	$\mu_{2,6} = 2.0$	2.01	0.11	2.00	0.14
$\lambda_{1,21} = 0.4$	0.41	0.07	0.41	0.07	$\lambda_{2,21} = 0.8$	0.82	0.14	0.85	0.22
$\lambda_{1,42} = 0.4$	0.41	0.09	0.41	0.12	$\lambda_{2,42} = 0.8$	0.81	0.18	0.86	0.24
$\lambda_{1,63} = 0.4$	0.39	0.08	0.43	0.11	$\lambda_{2,63} = 0.8$	0.83	0.20	0.80	0.21
$\gamma_{1,1} = 0.6$	0.63	0.16	0.63	0.17	$\gamma_{2,1} = 0.6$	0.63	0.18	0.63	0.19
$\gamma_{1,2} = 0.6$	0.60	0.15	0.63	0.17	$\gamma_{2,2} = -0.6$	-0.64	0.21	-0.58	0.19
$\phi_{1,11} = 1.0$	1.00	0.23	0.99	0.28	$\phi_{2,11} = 1.0$	1.06	0.31	1.03	0.35
$\phi_{1,12} = 0.3$	0.30	0.14	0.29	0.15	$\phi_{2,12} = 0.3$	0.33	0.15	0.32	0.17
$\phi_{1,22} = 1.0$	1.01	0.23	0.98	0.27	$\phi_{2,22} = 1.0$	1.00	0.28	1.06	0.33
$\psi_{1,11} = 0.36$	0.36	0.14	0.31	0.15	$\psi_{2,11} = 0.64$	0.65	0.20	0.63	0.26
$\psi_{1,22} = 0.36$	0.36	0.06	0.36	0.06	$\psi_{2,22} = 0.64$	0.63	0.13	0.60	0.19
$\psi_{1,33} = 0.36$	0.37	0.16	0.35	0.18	$\psi_{2,33} = 0.64$	0.60	0.19	0.62	0.24
$\psi_{1,44} = 0.36$	0.35	0.05	0.34	0.06	$\psi_{2,44} = 0.64$	0.63	0.14	0.58	0.17
$\psi_{1,55} = 0.36$	0.33	0.15	0.38	0.21	$\psi_{2,55} = 0.64$	0.63	0.20	0.58	0.25
$\psi_{1,66} = 0.36$	0.36	0.05	0.33	0.07	$\psi_{2,66} = 0.64$	0.62	0.15	0.63	0.17
$\psi_{1,\delta} = 0.36$	0.33	0.15	0.33	0.19	$\psi_{2,\delta} = 0.64$	0.59	0.21	0.62	0.27

Table 4: Mean and root mean squares of ML estimates obtained with 'All Data': Simulation Study 2

Component 1			Component 2			Component 3		
Par	Mean	Rms	Par	Mean	Rms	Par	Mean	Rms
$\pi_1 = 0.4$	0.40	0.02	$\pi_2 = 0.3$	0.30	0.01	$\pi_3 = 0.3$	0.30	0.02
$\mu_{1,1} = 0.0$	0.00	0.08	$\mu_{2,1} = 2.0$	2.00	0.08	$\mu_{3,1} = -2.0$	-1.99	0.07
$\mu_{1,2} = 0.0$	0.01	0.05	$\mu_{2,2} = 2.0$	2.00	0.07	$\mu_{3,2} = -2.0$	-2.00	0.06
$\mu_{1,3} = 0.0$	0.01	0.06	$\mu_{2,3} = 2.0$	2.01	0.09	$\mu_{3,3} = -2.0$	-2.01	0.06
$\mu_{1,4} = 0.0$	0.01	0.04	$\mu_{2,4} = 2.0$	2.00	0.09	$\mu_{3,4} = -2.0$	-2.00	0.06
$\mu_{1,5} = 0.0$	0.01	0.07	$\mu_{2,5} = 2.0$	2.01	0.08	$\mu_{3,5} = -2.0$	-2.00	0.08
$\mu_{1,6} = 0.0$	-0.00	0.04	$\mu_{2,6} = 2.0$	2.01	0.08	$\mu_{3,6} = -2.0$	-2.00	0.06
$\lambda_{1,21} = 0.4$	0.40	0.05	$\lambda_{2,21} = 0.8$	0.78	0.11	$\lambda_{3,21} = 0.6$	0.60	0.06
$\lambda_{1,42} = 0.4$	0.41	0.06	$\lambda_{2,42} = 0.8$	0.82	0.12	$\lambda_{3,42} = 0.6$	0.62	0.09
$\lambda_{1,63} = 0.4$	0.41	0.05	$\lambda_{2,63} = 0.8$	0.81	0.13	$\lambda_{3,63} = 0.6$	0.60	0.10
$\gamma_{1,1} = 0.6$	0.61	0.08	$\gamma_{2,1} = 0.6$	0.62	0.11	$\gamma_{3,1} = -0.6$	-0.61	0.08
$\gamma_{1,2} = 0.6$	0.62	0.09	$\gamma_{2,2} = -0.6$	-0.61	0.12	$\gamma_{3,2} = 0.6$	0.61	0.11
$\phi_{1,11} = 1.0$	1.00	0.15	$\phi_{2,11} = 1.0$	1.01	0.19	$\phi_{3,11} = 1.0$	0.99	0.15
$\phi_{1,12} = 0.3$	0.31	0.08	$\phi_{2,12} = 0.3$	0.31	0.09	$\phi_{3,12} = 0.3$	0.28	0.08
$\phi_{1,22} = 1.0$	1.01	0.14	$\phi_{2,22} = 1.0$	1.01	0.19	$\phi_{3,22} = 1.0$	1.01	0.19
$\psi_{1,11} = 0.36$	0.34	0.10	$\psi_{2,11} = 0.64$	0.61	0.16	$\psi_{3,11} = 0.36$	0.36	0.08
$\psi_{1,22} = 0.36$	0.36	0.03	$\psi_{2,22} = 0.64$	0.65	0.10	$\psi_{3,22} = 0.36$	0.35	0.04
$\psi_{1,33} = 0.36$	0.37	0.10	$\psi_{2,33} = 0.64$	0.63	0.15	$\psi_{3,33} = 0.36$	0.37	0.12
$\psi_{1,44} = 0.36$	0.35	0.03	$\psi_{2,44} = 0.64$	0.63	0.10	$\psi_{3,44} = 0.64$	0.64	0.07
$\psi_{1,55} = 0.36$	0.37	0.10	$\psi_{2,55} = 0.64$	0.64	0.14	$\psi_{3,55} = 0.64$	0.63	0.13
$\psi_{1,66} = 0.36$	0.36	0.03	$\psi_{2,66} = 0.64$	0.64	0.10	$\psi_{3,66} = 0.64$	0.64	0.08
$\psi_{1,\delta} = 0.36$	0.37	0.11	$\psi_{2,\delta} = 0.64$	0.67	0.15	$\psi_{3,\delta} = 0.36$	0.34	0.10

Table 5: Mean and root mean squares of ML estimates obtained with 'COM Data': Simulation Study 2 with G = 360

Component 1			Component 2			Component 3		
Par	Mean	Rms	Par	Mean	Rms	Par	Mean	Rms
$\pi_1 = 0.4$	0.39	0.02	$\pi_2 = 0.3$	0.31	0.02	$\pi_3 = 0.3$	0.30	0.01
$\mu_{1,1} = 0.0$	-0.02	0.09	$\mu_{2,1} = 2.0$	2.01	0.11	$\mu_{3,1} = -2.0$	-2.01	0.08
$\mu_{1,2} = 0.0$	-0.01	0.05	$\mu_{2,2} = 2.0$	2.03	0.11	$\mu_{3,2} = -2.0$	-2.02	0.06
$\mu_{1,3} = 0.0$	-0.01	0.08	$\mu_{2,3} = 2.0$	2.01	0.12	$\mu_{3,3} = -2.0$	-2.01	0.07
$\mu_{1,4} = 0.0$	0.00	0.05	$\mu_{2,4} = 2.0$	1.97	0.12	$\mu_{3,4} = -2.0$	-1.99	0.07
$\mu_{1,5} = 0.0$	-0.03	0.09	$\mu_{2,5} = 2.0$	2.04	0.10	$\mu_{3,5} = -2.0$	-2.04	0.10
$\mu_{1,6} = 0.0$	-0.01	0.05	$\mu_{2,6} = 2.0$	1.96	0.10	$\mu_{3,6} = -2.0$	-1.98	0.07
$\lambda_{1,21} = 0.4$	0.36	0.07	$\lambda_{2,21} = 0.8$	0.80	0.14	$\lambda_{3,21} = 0.6$	0.60	0.09
$\lambda_{1,42} = 0.4$	0.41	0.07	$\lambda_{2,42} = 0.8$	0.81	0.14	$\lambda_{3,42} = 0.6$	0.64	0.11
$\lambda_{1,63} = 0.4$	0.43	0.08	$\lambda_{2,63} = 0.8$	0.87	0.20	$\lambda_{3,63} = 0.6$	0.62	0.14
$\gamma_{1,1} = 0.6$	0.64	0.13	$\gamma_{2,1} = 0.6$	0.57	0.12	$\gamma_{3,1} = -0.6$	-0.64	0.12
$\gamma_{1,2} = 0.6$	0.66	0.15	$\gamma_{2,2} = -0.6$	-0.67	0.14	$\gamma_{3,2} = 0.6$	0.61	0.13
$\phi_{1,11} = 1.0$	0.88	0.19	$\phi_{2,11} = 1.0$	1.02	0.22	$\phi_{3,11} = 1.0$	0.94	0.19
$\phi_{1,12} = 0.3$	0.14	0.17	$\phi_{2,12} = 0.3$	0.25	0.11	$\phi_{3,12} = 0.3$	0.26	0.11
$\phi_{1,22} = 1.0$	0.72	0.32	$\phi_{2,22} = 1.0$	0.86	0.25	$\phi_{3,22} = 1.0$	0.91	0.21
$\psi_{1,11} = 0.36$	0.36	0.11	$\psi_{2,11} = 0.64$	0.66	0.17	$\psi_{3,11} = 0.36$	0.39	0.12
$\psi_{1,22} = 0.36$	0.38	0.05	$\psi_{2,22} = 0.64$	0.67	0.13	$\psi_{3,22} = 0.36$	0.38	0.06
$\psi_{1,33} = 0.36$	0.41	0.13	$\psi_{2,33} = 0.64$	0.64	0.17	$\psi_{3,33} = 0.36$	0.41	0.14
$\psi_{1,44} = 0.36$	0.37	0.04	$\psi_{2,44} = 0.64$	0.69	0.14	$\psi_{3,44} = 0.64$	0.65	0.09
$\psi_{1,55} = 0.36$	0.53	0.21	$\psi_{2,55} = 0.64$	0.72	0.21	$\psi_{3,55} = 0.64$	0.67	0.17
$\psi_{1,66} = 0.36$	0.36	0.04	$\psi_{2,66} = 0.64$	0.67	0.14	$\psi_{3,66} = 0.64$	0.67	0.10
$\psi_{1,\delta} = 0.36$	0.38	0.12	$\psi_{2,\delta} = 0.64$	0.62	0.19	$\psi_{3,\delta} = 0.36$	0.34	0.12

Table 6: Mean and root mean squares of ML estimates obtained with 'FO Data':

Simulation Study 2 with G = 360

Component 1			Component 2			Component 3		
Par	Mean	Rms	Par	Mean	Rms	Par	Mean	Rms
$\pi_1 = 0.4$	0.38	0.04	$\pi_2 = 0.3$	0.30	0.03	$\pi_3 = 0.3$	0.32	0.02
$\mu_{1,1} = 0.0$	0.00	0.12	$\mu_{2,1} = 2.0$	1.99	0.14	$\mu_{3,1} = -2.0$	-2.00	0.10
$\mu_{1,2} = 0.0$	0.00	0.07	$\mu_{2,2} = 2.0$	2.01	0.12	$\mu_{3,2} = -2.0$	-2.01	0.08
$\mu_{1,3} = 0.0$	0.01	0.12	$\mu_{2,3} = 2.0$	2.02	0.16	$\mu_{3,3} = -2.0$	-2.02	0.10
$\mu_{1,4} = 0.0$	0.01	0.07	$\mu_{2,4} = 2.0$	2.00	0.12	$\mu_{3,4} = -2.0$	-2.00	0.09
$\mu_{1,5} = 0.0$	0.01	0.12	$\mu_{2,5} = 2.0$	2.02	0.12	$\mu_{3,5} = -2.0$	-2.01	0.12
$\mu_{1,6} = 0.0$	-0.00	0.07	$\mu_{2,6} = 2.0$	2.01	0.12	$\mu_{3,6} = -2.0$	-2.01	0.10
$\lambda_{1,21} = 0.4$	0.40	0.09	$\lambda_{2,21} = 0.8$	0.82	0.19	$\lambda_{3,21} = 0.6$	0.62	0.11
$\lambda_{1,42} = 0.4$	0.41	0.08	$\lambda_{2,42} = 0.8$	0.81	0.18	$\lambda_{3,42} = 0.6$	0.63	0.14
$\lambda_{1,63} = 0.4$	0.42	0.08	$\lambda_{2,63} = 0.8$	0.81	0.19	$\lambda_{3,63} = 0.6$	0.58	0.15
$\gamma_{1,1} = 0.6$	0.60	0.16	$\gamma_{2,1} = 0.6$	0.61	0.17	$\gamma_{3,1} = -0.6$	-0.63	0.17
$\gamma_{1,2} = 0.6$	0.64	0.18	$\gamma_{2,2} = -0.6$	-0.60	0.16	$\gamma_{3,2} = 0.6$	0.61	0.18
$\phi_{1,11} = 1.0$	1.02	0.21	$\phi_{2,11} = 1.0$	1.04	0.28	$\phi_{3,11} = 1.0$	1.01	0.25
$\phi_{1,12} = 0.3$	0.31	0.13	$\phi_{2,12} = 0.3$	0.29	0.13	$\phi_{3,12} = 0.3$	0.31	0.12
$\phi_{1,22} = 1.0$	1.00	0.23	$\phi_{2,22} = 1.0$	1.00	0.28	$\phi_{3,22} = 1.0$	1.05	0.28
$\psi_{1,11} = 0.36$	0.35	0.14	$\psi_{2,11} = 0.64$	0.62	0.23	$\psi_{3,11} = 0.36$	0.36	0.14
$\psi_{1,22} = 0.36$	0.35	0.06	$\psi_{2,22} = 0.64$	0.60	0.17	$\psi_{3,22} = 0.36$	0.35	0.07
$\psi_{1,33} = 0.36$	0.35	0.16	$\psi_{2,33} = 0.64$	0.61	0.23	$\psi_{3,33} = 0.36$	0.36	0.16
$\psi_{1,44} = 0.36$	0.35	0.05	$\psi_{2,44} = 0.64$	0.65	0.17	$\psi_{3,44} = 0.64$	0.63	0.10
$\psi_{1,55} = 0.36$	0.39	0.15	$\psi_{2,55} = 0.64$	0.60	0.23	$\psi_{3,55} = 0.64$	0.57	0.24
$\psi_{1,66} = 0.36$	0.35	0.05	$\psi_{2,66} = 0.64$	0.63	0.15	$\psi_{3,66} = 0.64$	0.63	0.11
$\psi_{1,\delta} = 0.36$	0.32	0.16	$\psi_{2,\delta} = 0.64$	0.64	0.23	$\psi_{3,\delta} = 0.36$	0.33	0.15

Table 7: Mean and root mean squares of ML estimates obtained with 'ALL Data':

Simulation Study 3

Component 1			Component 2		
Par	Mean	Rms	Par	Mean	Rms
$\pi_1 = 0.5$	0.50	0.02	$\pi_2 = 0.5$	0.50	0.02
$\mu_{1,1} = -1.0$	-0.99	0.05	$\mu_{2,1} = 1.0$	1.01	0.07
$\mu_{1,2} = -1.0$	-1.00	0.04	$\mu_{2,2} = 1.0$	1.01	0.06
$\mu_{1,3} = -1.0$	-1.00	0.04	$\mu_{2,3} = 1.0$	1.01	0.06
$\mu_{1,4} = -1.0$	-0.99	0.08	$\mu_{2,4} = 1.0$	1.00	0.07
$\mu_{1,5} = -1.0$	-0.99	0.07	$\mu_{2,5} = 1.0$	1.00	0.04
$\mu_{1,6} = -1.0$	-0.99	0.07	$\mu_{2,6} = 1.0$	1.00	0.04
$\mu_{1,7} = -1.0$	-0.99	0.06	$\mu_{2,7} = 1.0$	1.00	0.04
$\lambda_{1,21} = 0.5$	0.50	0.05	$\lambda_{2,21} = 0.7$	0.69	0.06
$\lambda_{1,31} = 0.5$	0.50	0.05	$\lambda_{2,31} = 0.7$	0.69	0.06
$\lambda_{1,52} = 0.7$	0.70	0.04	$\lambda_{2,52} = 0.5$	0.50	0.05
$\lambda_{1,61} = 0.3$	0.31	0.09	$\lambda_{2,61} = 0.3$	0.30	0.09
$\lambda_{1,62} = 0.7$	0.70	0.07	$\lambda_{2,62} = 0.5$	0.51	0.08
$\lambda_{1,72} = 0.7$	0.71	0.04	$\lambda_{2,72} = 0.5$	0.50	0.05
$\gamma_{1,1} = 0.6$	0.59	0.05	$\gamma_{2,1} = -0.6$	-0.60	0.06
$\phi_{1,11} = 1.0$	1.01	0.11	$\phi_{2,11} = 1.0$	1.01	0.11
$\psi_{1,11} = 0.25$	0.25	0.04	$\psi_{2,11} = 0.36$	0.36	0.05
$\psi_{1,22} = 0.25$	0.25	0.03	$\psi_{2,22} = 0.36$	0.37	0.04
$\psi_{1,33} = 0.25$	0.25	0.02	$\psi_{2,33} = 0.36$	0.36	0.04
$\psi_{1,44} = 0.25$	0.25	0.03	$\psi_{2,44} = 0.36$	0.35	0.07
$\psi_{1,55} = 0.25$	0.26	0.02	$\psi_{2,55} = 0.36$	0.36	0.04
$\psi_{1,66} = 0.25$	0.25	0.03	$\psi_{2,66} = 0.36$	0.36	0.04
$\psi_{1,77} = 0.25$	0.25	0.03	$\psi_{2,77} = 0.36$	0.35	0.04
$\psi_{1,\delta} = 0.25$	0.25	0.05	$\psi_{2,\delta} = 0.36$	0.36	0.07

Table 8: Mean and root mean squares of ML estimates obtained with 'COM Data' :

Simulation Study 3

Component 1					Component 2				
	G = 200		G = 100			G = 200		G = 100	
Par	Mean	Rms	Mean	Rms	Par	Mean	Rms	Mean	Rms
$\pi_1 = 0.5$	0.05	0.02	0.51	0.02	$\pi_2 = 0.5$	0.50	0.02	0.49	0.02
$\mu_{1,1} = -1.0$	-1.01	0.06	-1.04	0.08	$\mu_{2,1} = 1.0$	1.03	0.08	1.04	0.08
$\mu_{1,2} = -1.0$	-1.00	0.05	-1.01	0.04	$\mu_{2,2} = 1.0$	1.01	0.06	1.02	0.07
$\mu_{1,3} = -1.0$	-1.01	0.05	-1.00	0.05	$\mu_{2,3} = 1.0$	1.02	0.07	1.02	0.07
$\mu_{1,4} = -1.0$	-1.01	0.08	-1.03	0.09	$\mu_{2,4} = 1.0$	1.03	0.08	1.05	0.10
$\mu_{1,5} = -1.0$	-1.00	0.07	-1.01	0.07	$\mu_{2,5} = 1.0$	1.01	0.05	1.02	0.06
$\mu_{1,6} = -1.0$	-1.01	0.08	-1.02	0.09	$\mu_{2,6} = 1.0$	1.02	0.06	1.03	0.06
$\mu_{1,7} = -1.0$	-0.98	0.07	-0.99	0.07	$\mu_{2,7} = 1.0$	1.00	0.04	1.00	0.05
$\lambda_{1,21} = 0.5$	0.53	0.08	0.52	0.09	$\lambda_{2,21} = 0.7$	0.70	0.08	0.70	0.09
$\lambda_{1,31} = 0.5$	0.51	0.07	0.52	0.09	$\lambda_{2,31} = 0.7$	0.69	0.08	0.72	0.10
$\lambda_{1,52} = 0.7$	0.70	0.06	0.71	0.06	$\lambda_{2,52} = 0.5$	0.52	0.06	0.51	0.07
$\lambda_{1,61} = 0.3$	0.31	0.16	0.33	0.17	$\lambda_{2,61} = 0.3$	0.35	0.17	0.37	0.19
$\lambda_{1,62} = 0.7$	0.70	0.11	0.69	0.13	$\lambda_{2,62} = 0.5$	0.55	0.14	0.57	0.16
$\lambda_{1,72} = 0.7$	0.73	0.06	0.76	0.08	$\lambda_{2,72} = 0.5$	0.50	0.06	0.50	0.07
$\gamma_{1,1} = 0.6$	0.56	0.07	0.56	0.08	$\gamma_{2,1} = -0.6$	-0.63	0.08	-0.63	0.09
$\phi_{1,11} = 1.0$	0.96	0.13	0.91	0.17	$\phi_{2,11} = 1.0$	0.98	0.14	0.98	0.17
$\psi_{1,11} = 0.25$	0.28	0.06	0.31	0.09	$\psi_{2,11} = 0.36$	0.39	0.07	0.40	0.09
$\psi_{1,22} = 0.25$	0.25	0.03	0.27	0.04	$\psi_{2,22} = 0.36$	0.38	0.05	0.38	0.06
$\psi_{1,33} = 0.25$	0.25	0.03	0.26	0.03	$\psi_{2,33} = 0.36$	0.37	0.05	0.38	0.07
$\psi_{1,44} = 0.25$	0.27	0.05	0.28	0.06	$\psi_{2,44} = 0.36$	0.39	0.09	0.41	0.11
$\psi_{1,55} = 0.25$	0.26	0.03	0.27	0.04	$\psi_{2,55} = 0.36$	0.36	0.05	0.36	0.05
$\psi_{1,66} = 0.25$	0.27	0.04	0.28	0.05	$\psi_{2,66} = 0.36$	0.36	0.05	0.37	0.05
$\psi_{1,77} = 0.25$	0.26	0.03	0.26	0.03	$\psi_{2,77} = 0.36$	0.36	0.05	0.38	0.05
$\psi_{1,\delta} = 0.25$	0.25	0.06	0.24	0.07	$\psi_{2,\delta} = 0.36$	0.32	0.08	0.29	0.10

Table 9: Mean and root mean squares of ML estimates obtained with 'FO Data' :

Simulation Study 3

Component 1					Component 2				
	G = 200		G = 100			G = 200		G = 100	
Par	Mean	Rms	Mean	Rms	Par	Mean	Rms	Mean	Rms
$\pi_1 = 0.5$	0.50	0.01	0.50	0.03	$\pi_2 = 0.5$	0.50	0.01	0.50	0.03
$\mu_{1,1} = -1.0$	-0.98	0.09	-1.04	0.17	$\mu_{2,1} = 1.0$	0.99	0.11	1.00	0.15
$\mu_{1,2} = -1.0$	-0.98	0.07	-1.01	0.11	$\mu_{2,2} = 1.0$	0.99	0.10	0.99	0.14
$\mu_{1,3} = -1.0$	-0.99	0.07	-1.00	0.10	$\mu_{2,3} = 1.0$	0.99	0.10	1.00	0.13
$\mu_{1,4} = -1.0$	-1.00	0.12	-1.03	0.20	$\mu_{2,4} = 1.0$	1.00	0.10	1.02	0.18
$\mu_{1,5} = -1.0$	-0.99	0.09	-1.01	0.15	$\mu_{2,5} = 1.0$	1.00	0.07	0.99	0.12
$\mu_{1,6} = -1.0$	-0.99	0.11	-1.02	0.20	$\mu_{2,6} = 1.0$	1.00	0.07	1.00	0.10
$\mu_{1,7} = -1.0$	-0.99	0.10	-1.01	0.17	$\mu_{2,7} = 1.0$	1.01	0.07	1.01	0.12
$\lambda_{1,21} = 0.5$	0.52	0.10	0.51	0.17	$\lambda_{2,21} = 0.7$	0.70	0.10	0.72	0.15
$\lambda_{1,31} = 0.5$	0.51	0.10	0.51	0.16	$\lambda_{2,31} = 0.7$	0.68	0.11	0.74	0.18
$\lambda_{1,52} = 0.7$	0.70	0.07	0.70	0.11	$\lambda_{2,52} = 0.5$	0.51	0.08	0.52	0.13
$\lambda_{1,61} = 0.3$	0.31	0.22	0.36	0.38	$\lambda_{2,61} = 0.3$	0.33	0.21	0.34	0.30
$\lambda_{1,62} = 0.7$	0.71	0.15	0.67	0.27	$\lambda_{2,62} = 0.5$	0.54	0.19	0.57	0.27
$\lambda_{1,72} = 0.7$	0.70	0.07	0.72	0.11	$\lambda_{2,72} = 0.5$	0.51	0.08	0.52	0.14
$\gamma_{1,1} = 0.6$	0.59	0.09	0.60	0.14	$\gamma_{2,1} = -0.6$	-0.63	0.11	-0.60	0.18
$\phi_{1,11} = 1.0$	1.05	0.22	1.03	0.32	$\phi_{2,11} = 1.0$	0.97	0.21	1.02	0.29
$\psi_{1,11} = 0.25$	0.25	0.07	0.25	0.12	$\psi_{2,11} = 0.36$	0.34	0.09	0.31	0.14
$\psi_{1,22} = 0.25$	0.25	0.04	0.25	0.06	$\psi_{2,22} = 0.36$	0.36	0.06	0.35	0.10
$\psi_{1,33} = 0.25$	0.24	0.04	0.24	0.06	$\psi_{2,33} = 0.36$	0.36	0.08	0.35	0.11
$\psi_{1,44} = 0.25$	0.24	0.06	0.24	0.08	$\psi_{2,44} = 0.36$	0.37	0.12	0.35	0.17
$\psi_{1,55} = 0.25$	0.25	0.04	0.25	0.06	$\psi_{2,55} = 0.36$	0.35	0.06	0.34	0.09
$\psi_{1,66} = 0.25$	0.24	0.06	0.23	0.07	$\psi_{2,66} = 0.36$	0.33	0.07	0.32	0.09
$\psi_{1,77} = 0.25$	0.24	0.04	0.24	0.06	$\psi_{2,77} = 0.36$	0.35	0.07	0.36	0.08
$\psi_{1,\delta} = 0.25$	0.25	0.08	0.25	0.13	$\psi_{2,\delta} = 0.36$	0.36	0.12	0.37	0.17

original data set (variables 116, 117, 180, 132, 96, 255, 254 and 252) that are related with respondents' job and homelife are taken as manifest variables in $\mathbf{y} = (y_1, \cdots, y_8)^T$. For completeness, questions corresponding to these variables are presented in the Appendix. These variables are measured via a 10-point scale and hence are treated as continuous in this illustration. There are 2009 random observations, many of them are with missing entries and some of the sample sizes within the missing patterns are very small. The complicated missing patterns are presented in Table 10. We observe that there are 1054 fully observed data, and forty-five missing patterns. From the questions associated with the manifest variables, it is natural to consider a measurement model (2) with three latent variables: η, ξ_1 and ξ_2, such that the first two manifest variables are indicators for η, the 3rd, 4th and 5th manifest variables are indicators for ξ_1 and the remaining variables are indicators for ξ_2. Although other structures for the loading matrix can be considered, we choose the structure that gives nonoverlapping latent variables for clear interpretation and for identifying the model. Hence, the following specifications on the parameter matrices of each component are used: $\mathbf{B} = \mathbf{0}$, $\mathbf{\Gamma} = (\gamma_{11}, \gamma_{12})$, $\mathbf{\Psi}_\delta = \psi_\delta$,

$$\begin{aligned} \mathbf{\Lambda}^T &= \begin{pmatrix} 1.0^* & \lambda_{21} & 0^* & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 1.0^* & \lambda_{42} & \lambda_{52} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0^* & 1.0^* & \lambda_{73} & \lambda_{83} \end{pmatrix}, \qquad (26) \\ \mathbf{\Phi} &= \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}, \end{aligned}$$

and $\mathbf{\Psi} = \mathrm{diag}(\psi_1, \cdots, \psi_8)$. The number of parameters in the model is pretty large, for example, for a three-component model, the total number of parameters is 84. The latent variables can be roughly interpreted as 'job satisfaction, η', 'homelife, ξ_1', and 'job attitude, ξ_2'.

This data set is analyzed by our proposed approach on the basis of a mixture LISREL models with K ($K = 1, 2, 3$) components. The formulation of the model in every component is taken to be same as in (26). The proposed MCEM procedure is used to obtain the ML estimates and their standard errors. Here we set $T = 100$ to control the variance of the Monte Carlo error is less than 0.01, see Shi and Copas (2002). For our problem, as the average 'batch mean' based on 50 iterations seems acceptable, we take $J = 50$. Finally, following the suggestion of Shi and Copas (2002), we take the stopping rule with $\gamma_0 = 5$, $\delta_0 = 0.001$ and $\delta = 0.001$. Again, the algorithm is terminated after the stopping rule is satisfied for five successive iterations. Using this stopping rule to the 'COM Data', the MCEM algorithm stopped at the 101st and 294th iteration for one-component and two-component models, respectively , and $\bar{\boldsymbol{\theta}}^{(101)}$ and $\bar{\boldsymbol{\theta}}^{(294)}$ are be taken as ML estimates in corresponding models. To give some

Table 10: Missing patterns and their sample sizes: ICPSR data set, '×' and 'o' indicate missing and observed entries, respectively.

		Manifest Variables							
	size	1	2	3	4	5	6	7	8
1	1054	o	o	o	o	o	o	o	o
2	792	×	×	o	o	o	o	o	o
3	41	×	×	o	o	o	o	×	o
4	33	o	o	o	o	o	×	o	o
5	26	×	×	o	o	o	×	o	o
6	18	o	o	o	o	o	o	o	×
7	12	o	o	o	o	o	×	×	×
8	11	o	o	o	o	o	o	o	×
9	11	o	o	o	o	o	×	×	o
10	11	×	×	o	o	o	o	o	×
11	9	o	o	o	×	o	o	o	o
12	8	×	×	o	o	o	×	×	o
13	8	×	×	o	o	o	×	×	×
14	6	o	o	×	o	o	o	o	o
15	6	×	×	×	o	o	o	o	o
16	5	×	o	o	o	o	o	o	o
17	5	×	×	o	×	o	o	o	o
18	4	o	×	o	o	o	o	o	o
19	4	×	×	o	o	o	o	×	×
20	4	×	×	o	o	o	×	o	×
21	4	×	×	o	o	×	o	o	o
22	2	o	o	o	o	o	×	o	×
23	2	o	o	o	o	×	o	o	o
24	2	o	o	o	×	o	×	o	o
25	2	×	o	o	o	o	o	×	×
26	2	×	×	o	o	×	o	×	o
27	1	o	o	o	o	o	o	×	×
28	1	o	o	o	o	×	o	o	×
29	1	o	o	o	×	o	o	×	×
30	1	o	o	×	o	×	o	o	o
31	1	o	×	o	o	o	×	o	o
32	1	o	×	o	o	o	×	×	×
33	1	o	×	o	×	o	o	×	o
34	1	o	×	o	×	o	×	×	×
35	1	×	o	o	×	o	o	o	o
36	1	×	×	o	×	o	×	o	o
37	1	×	×	o	×	o	×	o	×
38	1	×	×	o	×	o	×	×	o
39	1	×	×	o	×	o	×	×	×
40	1	×	×	o	×	×	×	o	×
41	1	×	×	×	o	o	×	o	o
42	1	×	×	×	o	×	o	o	o
43	1	×	×	×	o	×	×	×	o
44	1	×	×	×	o	×	×	×	×
45	1	×	×	×	×	×	o	×	o

idea on convergence, plots of sequences of some randomly selected parameters are displayed in Figure 1.

We use the BIC to select a model with the most appropriate number of components. In the path sampling, for each $t_{(s)}$, a total of $J_1 = 500$ additional observations are simulated by the Gibbs sampler after a burn-in phase of 200 iterations for computing $\bar{U}_{(s)}$ in (20), and then the observed-data log-likelihood ratio is estimated via (20), using 20 fixed grids in [0,1]. To calculate $z(0)$, $J_2 = 4000$ additional observations are simulated by the Gibbs sampler after a burn-in phase of 200 iterations. Then the observed-data log-likelihood for the model with K components is estimated via (21). Let M_K denotes the mixture model with K $(K = 1, 2, 3)$ components, the estimated log-likelihood functions are equal to: -29685.62, -29341.78 and -29287.60, the BIC values for comparing M_1 to M_3 are $\text{BIC}_{21} = -474.84$ and $\text{BIC}_{32} = 104.48$. Thus, a one-component model is significantly worse than a two-component model which is significantly better than a three-component model. Hence, it can be concluded that a mixture model with two components should be chosen. For completeness, the observed-data likelihood function for M_1 to M_3 based on the fully observed data are also computed, they are equal to -17153.05, -16802.91 and -16707.98, the corresponding BIC values are $\text{BIC}_{21} = -505.39$ and $\text{BIC}_{32} = 5.03$. Although the same conclusion can be obtained on the basis of the fully observed data, the supporting evidence of a two-component model based on BIC_{32}(=5.03) is not as strong as before. Indeed, the BIC values obtained by using only the fully observed data are quite different from the BIC value obtained by using all the data. ML estimates of the selected mixture model with two components (C1, C2) based on 'COM Data' and 'FO Data' are presented in Table 11 together with the standard errors estimates. Note that standard errors estimates obtained from 'COM Data' are smaller. Moreover, the two sets of ML estimates are different. In particular, substantial differences are found in estimates of $\{\gamma_{12}, \lambda_{73}, \lambda_{83}, \psi_7, \phi_{11}, \phi_{12}, \phi_{22}\}$ in the first component, and $\{\phi_{11}, \phi_{12}, \phi_{22}, \psi_\delta, \psi_5\}$ in the second component. Hence, the inferior results obtained by just using the fully observed data may lead to misleading conclusions.

Based on the better ML estimates obtained from 'COM Data', we observe that the estimates of μ_6, μ_7, μ_8 in the first component and the second component are very different. As the mean difference has stronger influence in distinguishing the components, the heterogeneity of the two components is mainly due to the mean differences between the last three manifest variables, which are the indicators of ξ_2. We also observe substantial differences between the estimates of $\lambda_{83}, \phi_{11}, \phi_{22}, \psi_\delta, \psi_1, \psi_3, \psi_4, \psi_6, \psi_7$ and ψ_8 in components 1 and 2. Hence, the covariance matrices are quite different. However, the behaviors of the latent variables in these components are similar; for example

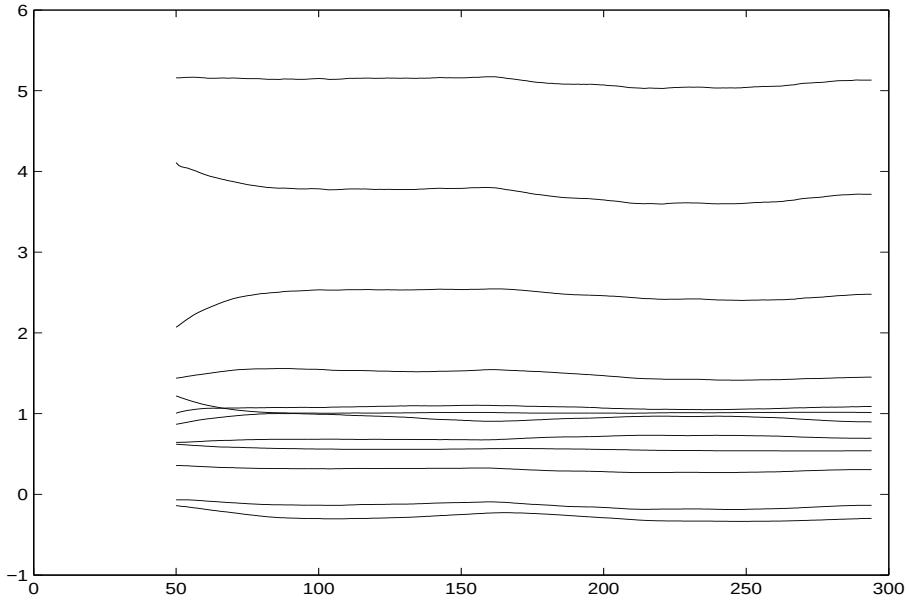

Figure 1: Estimates of π_1, π_2, λ_{52}, γ_1, γ_2, ϕ_{11} and ψ_4 in a two-component SEM for ICPSR data with $T = 100$.

the causal effects of ξ_1 to η are both positive (see $\hat{\gamma}_{11}$ under C1 and C2 in Table 11), the causal effects ξ_2 to η are both negative (see $\hat{\gamma}_{12}$), and the correlations of ξ_1 and ξ_2 in the first and second components are equal to -0.62 and -0.31, respectively. The estimates of the variances of the error measurements indicate that the model associated with the second component is more stable. From the estimates of π, we know that most observations are from the first component.

7. Summary

In this paper, a ML approach is developed for analyzing a finite mixture of SEMs with missing data that are missing at random with an ignorable mechanism. Inspired by the idea of data augmentation and the power of the EM algorithm, the observed data are augmented with the latent vectors in the model, the real missing data, and allocation variables in the ML analysis. A MCEM type algorithm is developed, in which the E-step is completed by approximating the conditional expectations with the observations generated from the conditional distributions via the Gibbs sampler, the M-step is completed by conditional maximization (Meng & Rubin, 1994), and convergence is monitored by an average methods and a stopping rule proposed in Shi and Copas (2002). According to our experience in the empirical studies, the algorithm proposed is efficient.

Table 11: ML estimates of parameters and their standard errors in the selected model with 2 components in analyzing the ICPSR data set.

	'COM Data'				'FO Data'			
	C1		C2		C1		C2	
Par	Est	Std	Est	Std	Est	Std	Est	Std
π	0.69	0.01	0.31	0.01	0.68	0.01	0.32	0.01
μ_1	7.10	0.03	7.07	0.03	6.92	0.04	7.78	0.04
μ_2	6.11	0.05	6.03	0.07	5.96	0.07	6.80	0.09
μ_3	7.40	0.03	7.43	0.04	7.23	0.05	8.05	0.04
μ_4	6.55	0.05	6.97	0.04	6.42	0.07	7.60	0.05
μ_5	7.24	0.03	7.06	0.04	7.21	0.04	7.75	0.04
μ_6	5.13	0.06	2.44	0.04	5.08	0.09	2.89	0.05
μ_7	3.80	0.05	2.05	0.02	3.69	0.06	2.30	0.04
μ_8	5.08	0.07	2.24	0.03	4.88	0.10	2.64	0.06
λ_{21}	1.03	0.03	1.17	0.05	1.07	0.05	1.14	0.08
λ_{42}	0.81	0.03	1.11	0.03	0.82	0.05	1.15	0.06
λ_{52}	1.01	0.02	1.05	0.03	1.08	0.03	1.19	0.05
λ_{73}	0.93	0.05	1.03	0.03	1.55	0.06	0.91	0.04
λ_{83}	0.35	0.06	0.87	0.04	0.72	0.10	0.90	0.06
γ_{11}	0.54	0.02	0.90	0.02	0.58	0.03	1.04	0.03
γ_{12}	-0.30	0.03	-0.14	0.04	-0.19	0.05	-0.14	0.03
ϕ_{11}	2.48	0.09	1.79	0.10	1.70	0.09	0.77	0.05
ϕ_{12}	-0.62	0.05	-0.31	0.04	-0.20	0.05	-0.23	0.04
ϕ_{22}	1.25	0.05	0.59	0.03	0.97	0.05	1.00	0.07
ψ_δ	1.28	0.05	0.57	0.03	1.40	0.07	0.23	0.02
ψ_1	1.36	0.05	0.61	0.03	1.40	0.07	0.44	0.03
ψ_2	3.52	0.13	2.66	0.13	3.53	0.19	2.52	0.19
ψ_3	1.73	0.06	0.83	0.05	2.10	0.11	0.58	0.04
ψ_4	3.72	0.14	1.09	0.06	3.22	0.17	0.84	0.06
ψ_5	1.45	0.05	1.13	0.06	1.35	0.07	0.55	0.04
ψ_6	5.25	0.19	0.89	0.05	5.34	0.28	0.95	0.07
ψ_7	3.65	0.13	0.33	0.02	2.52	0.12	0.47	0.04
ψ_8	7.00	0.26	0.76	0.04	6.64	0.34	1.09	0.08

As expected, ML estimates obtained by utilizing all the informations with those given by data point with missing entries are better than estimates obtained by just using the fully observed data. We expect that ignoring incomplete data in the analysis may occasionally select the wrong model with an inappropriate number of components. Because of the computational burden, our simulation is admittedly limited. Extensive and detailed simulation studies are necessary in order to get better understanding of the impact of ignoring the incomplete data in various settings. Moreover, further research is necessary to develop methods for handling non-ignorable missing data, for achieving local and global influence measures for identifying influential observations, and for more complicated SEMs.

Appendix: Manifest variables in the ICPSR example

The number of the variable corresponding to the original data set is given in parenthesis at the end of each statement.

- y_1: Overall, how satisfied or dissatisfied are you with your job? (V116)
- y_2: How free are you to make decisions in your job? (V117)
- y_3: Overall, how satisfied are you with your home life? (V180)
- y_4: How satisfied are you with the financial situation of your household? (V132)
- y_5: All things considered, how satisfied are you with your life as a whole in these day? (V96)
- y_6: In the long run, hard work usually bring a better life. (V255)
- y_7: Competition is good. It simulates people to work hard and develop new ideas. (V254)
- y_8: Individual should take more responsibility for providing for themselves. (V252)

References

ALLISON, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed), *Sociological Methodology*, Washington, D. C: American sociological Association, 71-103.

ARMINGER, G. & STEIN, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research, 26*, 148-182.

ARMINGER, G., STEIN, P. & WITTERBERG (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika, 64*, 475-494.

BOOTH, J. G. & HOBERT, J. P. (1999). Maximum generalized linear model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B, 61*, 265-285.

COOK, R. D. (1977). Detection of influential observations in linear regression. *Technometrics, 19*, 15-18.

COOK, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B, 48*, 133-169.

CRAWFORD, S. L. (1994). An application of the Laplace method to finite mixture distributions. *Journal of the American Statistical Association, 89*, 259-267.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

DIEBOLT, J. & ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B, 56*, 363-375.

DOLAN, C. V. & VAN DER MAAS, J. J. L. (1998). Fitting multivariate normal mixtures subject to structural equation modeling. *Psychometrika, 63*, 227-253.

FINKBEINER, C. (1979). Estimation for the multiple factor models when data are missing. *Psychometrika, 44*, 409-420.

GELMAN, A. & MENG, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science, 13*, 163-185.

GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721-741.

JAMSHIDIAM, M. & BENTLER, P. M. (1999). Maximum likelihood of mean and covariance structure with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24*, 21-41.

JEDIDI, K., JAGPAL, H. S. & DESARBO, W. S. (1997a). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*, 39-59.

JEDIDI, K., JAGPAL, H. S. & DESARBO, W. S. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification, 14*, 23-50

JÖRESKOG, K. G. & SÖRBOM , D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International: Hove and London.

KASS, R. E. & RAFTERY, A. E. (1995) Bayes Factors. *Journal of the American Statistical Association, 90*, 773-795.

LEE, S. Y. (1986). Estimation for structural equation models with missing data. *Psychometrika, 51*, 93-99.

LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

LOUIS, T. A. (1982). Finding the observed information matrix when using EM algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

MCCULLOCH, C. E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association, 89*, 330-335.

MENG, X. L. & RUBIN, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika, 80*, 267-278.

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

OGATA, Y. (1989). A Monte Carlo method for high dimensional integration. *Numer. Math. , 55*, 137-157.

PETTIT, L. I. & SMITH, A. F. M. (1985). Outliers and influential observations in linear models. In *Bayesian Statistics, 2* (eds J. M. Bernardo *et al.*), 473-494. Amsterdam: North Holland,

ROEDER, K. & WASSERMAN, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association, 92*, 894-902.

SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statist., 6*, 461-464.

SHI, J. Q. & COPAS, J. (2002). Publication bias and meta-analysis for 2×2 tables: an average Markov chain Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B, 64*, 221-236.

SONG, X. Y. & LEE, S. Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika, 67*, 261-288.

TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American statistical Association, 82*, 528-550.

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

WEI, G. C. G. & TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association, 85*, 699-704.

WORLD VALUES SURVEY, 1981-1984 AND 1990-1993. (1994). ICPSR version. Ann Arbor, MI: Institute for Social Research [producer], 1994. Ann Arbon, MI: Interuniversity Consortium for Political and Social Research [distribution], 1994.

YUNG, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297-330.

ZHU, H. T. & LEE S. Y. (2001) A Bayesian analysis of finite mixtures in the LISREL model. *Psychometrika, 66*, 133-152.