Published in final edited form as:

Sociol Methodol. 2010 August ; 40(1): 191–245. doi:10.1111/j.1467-9531.2010.01224.x.

Finite Normal Mixture SEM Analysis by Fitting Multiple Conventional SEM Models*

Ke-Hai Yuan and
University of Notre Dame

Peter M. Bentler
University of California, Los Angeles

Abstract

This paper proposes a two-stage maximum likelihood (ML) approach to normal mixture structural equation modeling (SEM), and develops statistical inference that allows distributional misspecification. Saturated means and covariances are estimated at stage-1 together with a sandwich-type covariance matrix. These are used to evaluate structural models at stage-2. Techniques accumulated in the conventional SEM literature for model diagnosis and evaluation can be used to study the model structure for each component. Examples show that the two-stage ML approach leads to correct or nearly correct models even when the normal mixture assumptions are violated and initial models are misspecified. Compared to single-stage ML, two-stage ML avoids the confounding effect of model specification and the number of components, and is computationally more efficient. Monte-Carlo results indicate that two-stage ML loses only minimal efficiency under the condition where single-stage ML performs best. Monte-Carlo results also indicate that the commonly used model selection criterion BIC is more robust to distribution violations for the saturated model than that for a structural model at moderate sample sizes. The proposed two-stage ML approach is also extremely flexible in modeling different components with different models. Potential new developments in the mixture modeling literature can be easily adapted to study issues with normal mixture SEM.

Keywords

Asymptotics; efficiency; distribution violation; model misspecification; model modification; model evaluation; sandwich-type covariance matrix

1. Introduction

In many disciplines, practical data may come from heterogeneous populations and the membership of each observation is unknown. Ignoring the heterogeneous nature of the sample can lead to false conclusions. As a methodology to account for heterogeneous populations with categorical data, latent class models have been well developed and widely used in social sciences (e.g., Clogg, 1995; Goodman, 1974a,b; Hagenaars & McCutcheon, 2002). Mixture models with continuous variables have also found wide applications in various disciplines (e.g., Everitt & Hand, 1981; McLachlan & Peel, 2000; Titterton, Smith & Makov, 1985).

*This research was supported by NSF grant DMS04-37167, grants DA01070 and DA00017 from the National Institute on Drug Abuse, and a grant from the National Natural Science Foundation of China (30870784).

Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

An early study of mixture confirmatory factor analysis was conducted by Blåfield (1980) using single-stage maximum likelihood (ML). Yung (1994, 1997) proposed three approaches to estimate the parameters in normal mixture factor analysis, two single-stage ML methods by expectation-maximization (EM) and approximate scoring algorithms, and one ad hoc two-stage approach. Single-stage ML for mixture structural equation modeling (SEM) was studied by Jedidi, Jagpal, and DeSarbo (1997) and Dolan and van der Maas (1998). Muthén and Shedden (1999) extended the mixture model to including continuous and categorical variables. Arminger and Stein (1997), and Arminger, Stein and Wittenberg (1999) developed three procedures to model mixtures with conditional means and covariance structures, two single-stage ML and one two-stage procedure. In the first stage of their two-stage procedure, (conditional) saturated means and covariances are estimated together with their asymptotic covariance matrix. The generalized least squares (GLS) or asymptotically distribution free (ADF) approach is used in the second stage. Hoshino (2001) and Zhu and Lee (2001) developed Bayes approaches to mixture SEM. None of these approaches allows standard model evaluation as is done with conventional SEM models. In this paper, we propose a two-stage ML approach to normal mixture SEM. In the first stage, saturated model MLEs and their asymptotic covariance matrices are obtained. Rather than using an approximation to the information matrix to obtain the asymptotic covariance matrix as in Arminger et al. (1999), for greater robustness we propose using a sandwich-type covariance matrix as the asymptotic covariance matrix. In the second stage, we propose fitting conventional SEM models to the means and covariances for each component obtained at stage-1. The paper will develop statistical inference for the proposed two-stage ML, including consistent standard errors (SE) and proper statistics for overall model evaluation.

In the literature on mixture models, one of the criticisms of single-stage ML is that misspecified structural models tend to yield more components than really exist in the population (Bauer & Curran, 2004; Biernacki, Celeux & Govaert, 2000; Lubke & Neale, 2006). Another drawback of single-stage ML seldom discussed in the literature is that, even if the number of components is correctly identified, the problem remains of judging how good the model is when fitting the overall model. It is common knowledge with conventional SEM that any interesting model is only an approximation to the real world, and that substantively and statistically acceptable models are hard to specify. This problem becomes even more pronounced in mixture SEM, where a heterogeneous population will require multiple acceptable models, one for each component. AIC, BIC and other model selection criteria only offer us relative information. Although global fit indices might be constructed when fitting a normal mixture SEM (Bauer & Curran, 2004; Jedidi et al., 1997; Tofighi & Enders, 2008), it is not clear whether we can use established guidelines for the corresponding fit index in a conventional SEM model (Hu & Bentler, 1999) to judge the quality of the mixture model. An unsatisfactory fit index also does not tell us whether one component has a gross misspecification or whether several models are moderately/grossly misspecified. Furthermore, in single-stage ML, a misspecification of one component of the model will affect the estimation of the rest. There might be also a masking effect with multiple misspecifications. Because the number of components is confounded with model misspecification, it is close to impossible to develop a satisfactory procedure for SEM diagnosis.

In contrast to single-stage ML, the two-stage approaches advocated by Yung (1994) and Arminger et al. (1999) avoid the confounding effect of the number of components and misspecified structural models. In their second stage, the ADF or GLS approach still simultaneously evaluates all the structural models, which does not permit us to judge the quality of model-fit for each component. With multiple components, the ADF/GLS approach will have a much higher dimension than their conventional counterpart, and one can expect that neither the test statistic for overall model evaluation nor the SEs of

parameter estimates would allow us to reliably evaluate the quality of the overall model or an individual parameter (Hu, Bentler & Kano, 1992; Yuan & Bentler, 1997).

The idea of the first stage proposed here is the same as in Yung (1994) and Arminger et al. (1999), where the saturated model is estimated by ML using EM algorithm to avoid the confounding of model misspecification and the number of components. In estimating the asymptotic covariance matrix of the obtained MLEs for the saturated means and covariances, we propose to use a sandwich-type covariance matrix that will account for possible violation of distribution conditions. For example, one or all of the component distributions may have heavier tails than that of the normal distribution. In that case, inverting the normal distribution based information matrix will not yield consistent model evaluation. The sandwich-type covariance matrix still permits obtaining consistent SEs and reliable overall model evaluation in the second stage. Details on the sandwich-type covariance matrix will be provided in section 2.

After obtaining the MLEs of means and covariances for all the components and consistent estimates of their asymptotic covariance matrices in the first stage, the problem turns into conventional SEM in the second stage. In particular, we can regard the obtained means and covariances for each component as the sample means and sample covariances from a distribution that may not be normally distributed. Thus, when fitting the means and covariances by the proposed structural models using the normal distribution based discrepancy function, corrections are needed using the obtained asymptotic covariance matrices from stage-1 to obtain consistent SEs and reliable statistics for overall model evaluation at stage-2. The number of models fitted at stage-2 is just the number of components identified at stage-1. Such a practice enjoys several merits. (I) The size of the estimation problem in the second stage is much smaller, thus the procedure should yield more stable estimates than the GLS/ADF approach proposed in Arminger et al. (1999). (II) Existing diagnostic techniques in conventional SEM, such as model modification or Lagrange multiplier (LM) tests, are directly applicable to each of the models fitted at stage-2. (III) Model evaluation procedures (e.g., statistics and fit indices) developed for conventional SEM are ready to use to evaluate the model for each component. Although the commonly used likelihood ratio statistic is not valid anymore, there are still several statistics that have been shown both analytically and empirically to work reliably for overall model evaluation.

Mixture models for continuous and categorical data have been implemented in Mplus (Muthén & Muthén, 2007). The option “ESTIMATOR=MLR;” provides consistent standard errors when distribution is violated. It also provides model modification to guide the specification of a mixture model. Because no statistic exists to judge how good the overall model is in single-stage ML, it is difficult to know whether the modified model is statistically sound. We will illustrate this through an example.

Under a set of idealized conditions the two-stage ML approach is not as desirable as single-stage ML. That is, when (1) the sample is from a truly normal mixture; (2) the number of components is correctly specified; (3) sample size is very large; and (4) we have a correct structural model for each component, then two-stage ML will not generate parameter estimates as efficient as the single-stage ML estimates. But the above conditions are too strong to be satisfied in most real applications of normal mixture SEM. When any of the above conditions is violated, whether or not single-stage ML still maintains an advantage over two-stage ML cannot be established analytically. We will empirically compare the two approaches by examples and Monte Carlo with violation of conditions as well as when the normal mixture assumption perfectly holds.

In the first stage of two-stage ML, we still need to determine the number of components. Following the recommendation of the literature on model selection (e.g., Bauer & Curran, 2004; McLachlan & Peel, 2000), we will use BIC for such a purpose. Because, with distribution violations, BIC tends to choose more components than is necessary with structured models (Bauer & Curran, 2003), we will study its performance with the saturated model. Although the statistics for overall model evaluation at stage-2 allow us to judge whether substantively interesting models can fit the data adequately and it is unlikely for these models to fit the data well when the number of components is wrong, it is always reassuring when BIC for the saturated model performs better. Such a study will be conducted by Monte Carlo.

In contrast to the existing literature of normal mixture SEM, we will emphasize methods for valid statistical inference when conditions are violated. In section 2 we will provide the statistical development for our two-stage ML, including the way to obtain consistent SEs and test statistics for overall model evaluation. In section 3 we will illustrate the proposed procedure using examples with real and simulated data. In section 4, we will use Monte Carlo to compare efficiency and biases of parameter estimates obtained with single-stage and two-stage ML. We will also compare the performance of the model selection criterion BIC with these two approaches when conditions of normal mixture modeling are violated. Section 5 offers some conclusions and discussion. Technical details for obtaining the sandwich-type covariance matrix are given in an appendix.

2. Model Inference

In this section we provide statistics for model inference with two-stage ML. This includes obtaining the MLEs and the sandwich-type covariance matrices for the saturated means and covariances at stage-1, and test statistics for overall model evaluation and consistent SEs for the structural parameter estimates at stage-2.

2.1 The first-stage

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a sample from a p -variate population $\mathbf{x}$. Under the normal mixture model the density function of $\mathbf{x}$ can be written as

$$f(\mathbf{x}) = \sum_{j=1}^m \pi_j f_j(\mathbf{x}), \quad (1)$$

where π_j is the proportion of the j th component that satisfies $\sum_{j=1}^m \pi_j = 1$, and

$$f_j(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_j|^{1/2}} \exp[-(\mathbf{x} - \mu_j)' \Sigma_j^{-1} (\mathbf{x} - \mu_j)/2]$$

is the density function of the p -variate normal distribution $N_p(\mu_j, \Sigma_j)$ for the j th component. Let $\sigma_j = \text{vech}(\Sigma_j)$ be the vector of stacking the columns of the lower-triangular part of Σ_j and $\beta_j = (\mu_j', \sigma_j')'$. Then the vector of parameters for the saturated model is

$$\beta = (\pi_1, \pi_2, \dots, \pi_{m-1}; \beta_1', \beta_2', \dots, \beta_m')'.$$

Let

$$l_i(\beta) = \log\left[\sum_{j=1}^m \pi_j f_j(\mathbf{x}_i)\right] = \log\left[\sum_{j=1}^{m-1} \pi_j [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)] + f_m(\mathbf{x}_i)\right]. \quad (2)$$

The observed log likelihood function is

$$l(\beta) = \sum_{i=1}^n l_i(\beta).$$

Let $\hat{\beta}$ be the proper value that maximizes $l(\beta)$, which can be obtained by the EM algorithm when component membership together with $\mathbf{x}_i$ are the augmented complete data (see chapter 3 of McLachlan & Peel, 2000). To be self-contained, appendix A provides the formulas for obtaining $\hat{\beta}$ using the EM algorithm. Unlike the conventional normal distribution based likelihood function, $l(\beta)$ usually has multiple local maxima. In particular, there exist singular Σ_j at which $l(\beta) = \infty$. In order to get the proper MLE, multiple starting values are needed. At the end of each M-step, it is necessary to check whether the determinant of each Σ_j is close to zero. The iteration process of the EM algorithm needs to restart with a new set of starting values whenever any of the Σ_j ($j = 1, 2, \dots, m$) is close to being singular.

Let $\dot{l}_i(\beta) = \partial l_i(\beta) / \partial \beta$ and $\ddot{l}_i(\beta) = \partial^2 l_i(\beta) / \partial \beta \partial \beta'$. For a proper MLE $\hat{\beta}$, there exists a β_0 satisfying

$$E[\dot{l}_i(\beta_0)] = \mathbf{0}$$

such that $\hat{\beta}$ is consistent for β_0 . Actually, the $\hat{\beta}$ satisfies the estimating equation

$$\sum_{i=1}^n \dot{l}_i(\hat{\beta}) = \mathbf{0}.$$

When n is sufficiently large, $\hat{\beta}$ will be in a neighborhood of β_0 . With a random sample $\mathbf{x}_i$, $i = 1, 2, \dots, n$, the condition for the consistency of $\hat{\beta}$ is $\beta \in \mathcal{B}$, which is a compact subset of the Euclidean space $\mathcal{R}^{m-1+mp(p+3)/2}$. Under the additional regularity condition that β_0 is an interior point of $\mathcal{B}$, we have (Yuan & Jennrich, 1998)

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Gamma_{sw}), \quad (3)$$

where

$$\Gamma_{sw} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$$

with $\mathbf{A} = -E[\ddot{l}_i(\beta_0)]$ and $\mathbf{B} = E[\dot{l}_i(\beta_0)\dot{l}_i'(\beta_0)]$. The result in (3) does not need the $f(\mathbf{x})$ in (1) to truly describe the population distribution. When $f(\mathbf{x})$ only approximately describes the population and m is correctly chosen, the means and covariances in β_0 only approximate the population means and covariances. When the number of components m is not correct, then β_0 still represents the best summary of the underlying population using m mean vectors and covariance matrices, as characterized by the Kullback Leibler discrepancy criterion. When

β_0 represents a good summary of the underlying population, we obtain a better understanding of the heterogeneous nature of the population than treating the sample as from a single homogeneous population. Note that β_0 being an interior point of $\mathcal{B}$ implies that the result in (3) does not hold when either certain Σ_j is singular or certain $\pi_j = 0$. So we need to avoid near singular $\hat{\Sigma}_j$ and/or $\hat{\pi}_j$ that are close to 0 in the MLE $\hat{\beta}$.

When the $f(\mathbf{x})$ in (1) truly describes the population distribution of $\mathbf{x}$, $\mathbf{A}$ is called the information matrix. We have $\mathbf{A} = \mathbf{B}$ and

$$\Gamma_{sw} = \Gamma_{in} = \mathbf{A}^{-1}$$

in such a case. When all the components have heavier tails than that of the normal distribution, $\mathbf{B} - \mathbf{A}$ will be a nonnegative or positive definite matrix. Thus, SEs based on Γ_{sw} tend to be greater than those based on Γ_{in} . Similarly, when all the components have tails lighter than that of the normal distribution, SEs based on Γ_{sw} tend to be smaller than those based on Γ_{in} . Whether $\mathbf{B} > \mathbf{A}$ or not, when $f(\mathbf{x})$ only approximately describes the distribution of $\mathbf{x}$, consistent estimates of $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\hat{\mathbf{A}} = -\frac{1}{n} \sum_{i=1}^n \ddot{l}_i(\hat{\beta}) \quad \text{and} \quad \hat{\mathbf{B}} = \frac{1}{n} \sum_{i=1}^n \dot{l}_i(\hat{\beta}) \dot{l}_i'(\hat{\beta}).$$

The explicit formulas for $\dot{l}_i(\beta)$ and $\ddot{l}_i(\beta)$ are rather complicated and are given in appendix B. These formulas are straightforward to evaluate and are only needed once at the proper MLE $\hat{\beta}$.

Arminger et al. (1999) used the inverse of the information matrix $\mathbf{A}^{-1}$ to describe the asymptotic behavior of $\hat{\beta}$. They used the cross-product of the first derivatives of the so-called “complete data” to approximate the information matrix, neither $\hat{\mathbf{A}}$ nor $\hat{\mathbf{B}}$.

The result in (3) implies that, for the estimates of the means and covariances of the j th component,

$$\sqrt{n}(\hat{\beta}_j - \beta_{j0}) \xrightarrow{\mathcal{L}} N_{p+p^*}(\mathbf{0}, \Gamma_{jj}), \quad (4)$$

where Γ_{jj} is a $(p + p^*) \times (p + p^*)$ submatrix of $\Gamma = \Gamma_{sw}$ or Γ_{in} corresponding to $\hat{\beta}_j$, with $p^* = p(p + 1)/2$. The result in (4) parallels the following result for the sample mean vector $\bar{\mathbf{y}}$ and sample covariance matrix $\mathbf{S}$ based on a sample of size n from a homogeneous population $\mathbf{y}$:

$$\sqrt{n} \begin{pmatrix} \bar{\mathbf{y}} - \boldsymbol{\mu} \\ \text{vech}(\mathbf{S}) - \text{vech}(\boldsymbol{\Sigma}) \end{pmatrix} \xrightarrow{\mathcal{L}} N_{p+p^*}(\mathbf{0}, \Gamma_y), \quad (5)$$

where $\boldsymbol{\mu} = E(\mathbf{y})$, $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{y})$ and $\Gamma_y = \text{Cov}(\mathbf{y}^*)$ with $\mathbf{y}^* = [\mathbf{y}', \text{vech}'\{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\}]'$. Comparing (4) with (5), we have the same amount of information to evaluate any particular structural model $\boldsymbol{\mu}_{j0} = \boldsymbol{\mu}(\theta_{j0})$ and $\boldsymbol{\Sigma}_{j0} = \boldsymbol{\Sigma}_j(\theta_{j0})$ for the j th component in the mixture modeling context as in conventional SEM or SEM with one component.

So we may just treat $\hat{\boldsymbol{\mu}}_j$ as the sample mean vector and $\hat{\boldsymbol{\Sigma}}_j$ as the sample covariance matrix of the j th component for the next stage analysis. Here our interest is in SEM. The second-stage analysis can be principal components, exploratory factor analysis or any multivariate

analysis involving the sample mean vector and sample covariance matrix. Actually, multilevel SEM has been developed using essentially the same idea (Yuan & Bentler, 2007). Compared to single-stage ML, the two-stage approach is more crucial with normal mixture SEM than with multilevel modeling because it allows us to segregate various effects and to be able to use the accumulated knowledge of conventional SEM to study the more challenging problems associated with normal mixture SEM (see Bauer, 2007).

2.2 The second stage

We need to choose a method to fit the structural model $(\mu_j(\theta_j), \Sigma_j(\theta_j))$ to $(\hat{\mu}_j, \hat{\Sigma}_j)$. We will use the normal distribution based discrepancy function¹

$$F_{ML}(\theta_j) = \text{tr}[\hat{\Sigma}_j \Sigma_j^{-1}(\theta_j)] - \log |\hat{\Sigma}_j \Sigma_j^{-1}(\theta_j)| + [\hat{\mu}_j - \mu_j(\theta_j)]' \Sigma_j^{-1}(\theta_j) [\hat{\mu}_j - \mu_j(\theta_j)] - p \quad (6)$$

for mean and covariance structure analysis and

$$F_{MLc}(\theta_{cj}) = \text{tr}[\hat{\Sigma}_j \Sigma_j^{-1}(\theta_{cj})] - \log |\hat{\Sigma}_j \Sigma_j^{-1}(\theta_{cj})| - p \quad (7)$$

for just covariance structure analysis. Let $\hat{\theta}_j$ minimize $F_{ML}(\theta_j)$ and $\hat{\theta}_{cj}$ minimize $F_{MLc}(\theta_{cj})$, we will discuss the properties of these estimators shortly. We choose (6) and (7) because (i) minimizing these functions generates more efficient parameter estimates than the GLS/ADF procedure in the context of conventional SEM even when $\mathbf{y}$ does not follow a normal distribution (Yuan & Bentler, 1997), (ii) there exist several statistics for overall model evaluation with nice properties, (iii) these discrepancy functions are most widely used in practice and are the default procedure in essentially all SEM software. Of course, one may choose another discrepancy function if needed. Actually, with (4), we are back to conventional SEM. Any existing method in conventional SEM can be applied if deemed necessary. For example, it is totally legitimate to be interested in mean and covariance structures for the j th component and only the covariance structure for the k th component. This poses no extra difficulty for the proposed two-stage ML.

Let

$$\beta_j(\theta_j) = \begin{pmatrix} \mu_j(\theta_j) \\ \sigma_j(\theta_j) \end{pmatrix}, \quad \dot{\beta}_j(\theta_j) = \partial \beta_j(\theta_j) / \partial \theta_j', \quad \dot{\sigma}_j(\theta_{cj}) = \partial \sigma_j(\theta_{cj}) / \partial \theta_{cj}';$$

$$\mathbf{W}_{cj}(\theta_{cj}) = \frac{1}{2} \mathbf{D}_p' [\Sigma_j^{-1}(\theta_{cj}) \otimes \Sigma_j^{-1}(\theta_{cj})] \mathbf{D}_p, \quad \text{and } \mathbf{W}_j(\theta_j) = \text{diag}[\Sigma_j^{-1}(\theta_j), \mathbf{W}_{cj}(\theta_{cj})].$$

We will omit the arguments of the above functions when evaluated at θ_{j0} or θ_{cj0} . With the above notation, we have

$$\sqrt{n}(\hat{\theta}_j - \theta_{j0}) \xrightarrow{\mathcal{L}} N_{q_j}(\mathbf{0}, \mathbf{\Omega}_j), \quad (8a)$$

where

¹Although we choose the discrepancy function in (6), we would like to note that $\hat{\Sigma}_j$ cannot be regarded as the sample covariance matrix from a normal distribution. Thus, the second stage is not strictly ML. The first stage or the so-called single-stage ML is not strictly ML either with most practical data. Following the typical use of ML methodology in practice, we call the procedure of minimizing $F_{ML}(\theta_j)$ for parameter estimates “ML” rather than “pseudo ML”.

$$\mathbf{\Omega}_j = (\dot{\beta}_j' \mathbf{W}_j \dot{\beta}_j)^{-1} (\dot{\beta}_j' \mathbf{W}_j \mathbf{\Gamma}_{jj} \mathbf{W}_j \dot{\beta}_j) (\dot{\beta}_j' \mathbf{W}_j \dot{\beta}_j)^{-1}; \quad (8b)$$

and

$$\sqrt{n}(\hat{\theta}_{cj} - \theta_{cj0}) \xrightarrow{\mathcal{L}} N_{q_j}(\mathbf{0}, \mathbf{\Omega}_{cj}), \quad (9a)$$

where

$$\mathbf{\Omega}_{cj} = (\dot{\sigma}_j' \mathbf{W}_{cj} \dot{\sigma}_j)^{-1} (\dot{\sigma}_j' \mathbf{W}_{cj} \mathbf{\Gamma}_{cjj} \mathbf{W}_{cj} \dot{\sigma}_j) (\dot{\sigma}_j' \mathbf{W}_{cj} \dot{\sigma}_j)^{-1}, \quad (9b)$$

with $\mathbf{\Gamma}_{cjj}$ being the submatrix of $\mathbf{\Gamma}_{jj}$ corresponding to $\hat{\sigma}_j$. Consistent estimators of $\mathbf{\Omega}_j$ and $\mathbf{\Omega}_{cj}$ are obtained when replacing θ_{j0} by $\hat{\theta}_j$ and $\mathbf{\Gamma}_{jj}$ by $\hat{\mathbf{\Gamma}}_{jj}$ in (8); and replacing θ_{cj0} by $\hat{\theta}_{cj}$ and $\mathbf{\Gamma}_{cjj}$ by $\hat{\mathbf{\Gamma}}_{cjj}$ in (9). Notice that $\mathbf{\Gamma}_{jj}$ does not equal $\mathbf{W}_j$ even when $\mathbf{x}$ truly follows a normal mixture model. So the SEs of $\hat{\theta}_j$ or Wald statistics regarding elements of θ_j always need to be based on the above sandwich-type covariance matrices $\mathbf{\Omega}_j$ or $\mathbf{\Omega}_{cj}$.

We next turn to overall model evaluation. Let us denote $T_{MLj} = nF_{ML}(\hat{\theta}_j)$. For a sample from a normally distributed population in conventional SEM, and under the null hypothesis, T_{ML} asymptotically follows a chi-square distribution with $p + p^* - q_j$ degrees of freedom, where q_j is the number of parameters in θ_j . In the above two-stage ML approach for the mixture model, T_{MLj} does not follow $\chi_{p+p^*-q_j}^2$ even asymptotically. We will introduce four statistics that have been shown to perform well when (5) holds. They are proposed to test the structural model $\beta_j(\theta_j)$ here because the parallel nature of (4) and (5). Let

$$T_{RADFj} = n \hat{\mathbf{e}}_j' (\hat{\mathbf{\Gamma}}_{jj}^{-1} - \hat{\mathbf{\Gamma}}_{jj}^{-1} \hat{\beta}_j(\hat{\theta}_j) [\hat{\beta}_j'(\hat{\theta}_j) \hat{\mathbf{\Gamma}}_{jj}^{-1} \hat{\beta}_j(\hat{\theta}_j)]^{-1} \hat{\beta}_j'(\hat{\theta}_j) \hat{\mathbf{\Gamma}}_{jj}^{-1}) \hat{\mathbf{e}}_j, \quad (10)$$

be the residual based ADF statistic (Browne, 1984), where $\hat{\mathbf{e}}_j = \hat{\beta}_j - \beta_j(\hat{\theta}_j)$. The first proposed statistic is the corrected residual-based ADF statistic (Yuan & Bentler, 1998)

$$T_{CRADFj} = \frac{T_{RADFj}}{1 + T_{RADFj}/n}. \quad (11)$$

Notice that, according to (4), both T_{RADFj} and T_{CRADFj} asymptotically follow $\chi_{p+p^*-q_j}^2$. But T_{CRADFj} has been shown to possess better finite sample properties. The second proposed statistic is the residual-based F -statistic (Yuan & Bentler, 1998)

$$F_{Rj} = \frac{[n - (p + p^* - q_j)] T_{RADFj}}{(n - 1)(p + p^* - q_j)}, \quad (12)$$

which is also asymptotically distribution free when referred to the F -distribution with degrees of freedom $p + p^* - q_j$ and $n - (p + p^* - q_j)$.

Let

$$\mathbf{U}_j = \mathbf{W}_j - \mathbf{W}_j \dot{\beta}_j (\dot{\beta}_j' \mathbf{W}_j \dot{\beta}_j)^{-1} \dot{\beta}_j' \mathbf{W}_j.$$

The third one is the rescaled statistic (Satorra & Bentler, 1994)

$$T_{RMLj} = \frac{(p + p^* - q_j)}{\text{tr}(\widehat{\mathbf{U}}_j \widehat{\mathbf{\Gamma}}_{jj})} T_{MLj}. \quad (13)$$

The statistic T_{RMLj} is not asymptotically distribution free; it approaches a distribution with mean equal to $p + p^* - q_j$.

Let

$$\widehat{a}_j = \text{tr}[(\widehat{\mathbf{U}}_j \widehat{\mathbf{\Gamma}}_{jj})^2] / \text{tr}(\widehat{\mathbf{U}}_j \widehat{\mathbf{\Gamma}}_{jj}), \text{ and } \widehat{b}_j = [\text{tr}(\widehat{\mathbf{U}}_j \widehat{\mathbf{\Gamma}}_{jj})]^2 / \text{tr}[(\widehat{\mathbf{U}}_j \widehat{\mathbf{\Gamma}}_{jj})^2].$$

The fourth proposed statistic is the adjusted statistic

$$T_{MLaj} = T_{MLj} / \widehat{a}_j, \quad (14)$$

using $T_{MLaj} \sim \chi_{\widehat{b}_j}^2$ for inference. Like T_{RMLj} , T_{MLaj} does not follow $\chi_{b_j}^2$ in general, where b_j is the population counterpart of $\widehat{b}_j$. The first and second moments of T_{MLaj} asymptotically equal those of $\chi_{\widehat{b}_j}^2$. Note that the degrees of freedom in $\chi_{\widehat{b}_j}^2$ is estimated rather than fixed.

Recent simulation results show that $\chi_{\widehat{b}_j}^2$ better describes the behavior of T_{MLaj} than $\chi_{p+p^*-q_j}^2$ for T_{RMLj} when $\mathbf{U}_j$ and $\mathbf{\Gamma}_{jj}$ are controlled (Yuan & Bentler, 2009).

For each of the proposed statistics, we have only given its explicit formulation with mean and covariance structure analysis. Parallel asymptotic distribution free statistics in just covariance structure analysis will be obtained when $\widehat{\beta}_j, \beta_j(\theta_j), \widehat{\theta}_j, \widehat{\mathbf{\Gamma}}_{jj}, p + p^*$ and q_j in (10) to (12) are replaced by $\widehat{\sigma}_j, \sigma_j(\theta_{cj}), \widehat{\theta}_{cj}, \widehat{\mathbf{\Gamma}}_{cjj}, p^*$ and q_{cj} , respectively. Similarly, rescaled and adjusted statistics are obtained when $T_{MLj}, \mathbf{W}_j, \dot{\beta}_j$ and $\mathbf{\Gamma}_{jj}$ in the formulation of (13) and (14) are replaced by $T_{MLcj} = nF_{MLc}(\widehat{\theta}_{cj}), \mathbf{W}_{cj}, \dot{\sigma}_j$ and $\mathbf{\Gamma}_{cjj}$, respectively. The degrees of freedom for the reference chi-square distribution in covariance structure analysis are $p^* - q_{cj}$, and for the reference F -distribution are $p^* - q_{cj}$ and $n - (p^* - q_{cj})$.

In conventional SEM, $T_{CRADFj}, F_{Rj}, T_{RMLj}$ and T_{MLaj} have been shown to perform quite well (see Bentler & Yuan, 1999; Fouladi, 2000; Yuan & Bentler, 1998). We expect that they will perform equally well in the context of mixture normal SEM due to the parallelism of (4) and (5). Because fitting $\beta_j(\theta_j)$ to $\widehat{\beta}_j$ is just conventional SEM, fit indices such as CFI, RMSEA, etc. can be computed in the usual way (see e.g., Hu & Bentler, 1999).

All the four proposed statistics are available in the current version of EQS (Bentler, in press). For the adjusted statistic T_{MLaj} , EQS prints out the integer part of the degrees of freedom $\widehat{b}_j$ and uses it to compute the p -value. T_{RMLj} and T_{MLaj} are available in Mplus (Muthén & Muthén, 2007). The statistic T_{RMLj} also exists in LISREL (Jöreskog et al., 2000, Ch. 4).

3. Illustrations

In this section we use three examples to illustrate the applications of the proposed two-stage ML when the structural model and/or distribution is misspecified. We will use

$$\text{BIC} = -2l(\hat{\beta}) + q \log(n)$$

to determine the proper number of components at stage-1, where q is the number of parameters and n is the sample size. Although we recommend that our methodology primarily be used in a confirmatory context, to recreate the typical practice of empirical model modification that is often used when SEM is applied in a partially exploratory way (Jöreskog, 1993), we also show that the proposed procedures are capable of recovering the generating model in the more difficult context of mixture modeling.

Example 1

Holzinger and Swineford (1939) contains test scores of $n = 145$ students on the following subtests or variables: Visual Perception, Cubes, Lozenges, Paragraph Comprehension, Sentence Completion, Word Meaning, Addition, Counting Dots, Straight-Curved Capitals. The first three variables were designed to measure “spatial ability”, the next three variables were designed to measure “verbal ability”, and the last three variables were administered with a limited time and were designed to measure a “speed” factor in performing the tasks. Thus, a three-factor model will reflect the original design well. One may also assume that all the nine variables measure “general intelligence” and use a one-factor normal mixture model to fit the sample. Of course, we can also fit the sample by a normal mixture with saturated means and covariances. Parallel to the set-up of Yung (1997), who used normal mixture factor model to fit a different set of variables from the same population, at $m = 2$ we set the factor loadings and factor covariances for both the one- and three-factor models equal across the components while the intercepts and error variances are free to vary. Table 1 contains the log likelihood and the BIC for the model with saturated means and covariances, the three-factor model, and the one-factor model at $m = 1$ and 2, respectively.

Each log likelihood at $m = 2$ is obtained using the proper maximum with 50 starting values. For the saturated model, BIC corresponding to one normal component is smaller. BIC also suggests one component if we use the three-factor model. However, it chooses two components if we fit the data using the one-factor model. Obviously, model specification and the number of components are closely related in mixture modeling (Bauer & Curran, 2004; Biernacki, Celeux & Govaert, 2000). Actually, we do not know the true model for this sample; the three-factor model at $m = 1$ does not fit the sample well, none of the p -values corresponding to the four proposed statistics is above .01. The LM test or model modification index suggests that allowing the 9th variable to load on the 1st factor will greatly improve the model-fit. This parameter is also supported by substantive meaning of the variables (see Sörbom, 1989). After adding this parameter, all the p -values corresponding to the four proposed statistics are above .20, which give us the needed confidence that the model fits the data well.

In the context of mixture model by single-stage ML, existing procedures do not allow us to judge the fit of the model as with the conventional single component SEM model. We will further illustrate this with the two-component one-factor model using the software Mplus (Muthén & Muthén, 2007). The estimation method is specified by “ESTIMATOR=MLR;”. Fifty starting values were used, and each run included up to 30 iterations in the initial maximization stage. Five starting values corresponding to 5 of the maximum log likelihoods from the initial 50 entered into the final stage maximization. The model modification index

is specified by “STANDARDIZED MODINDICES (3.84);”. Let M_1 denote the two-component 1-factor model with equal factor loadings across the components. The output contains $l(\hat{\theta}) = -4558.568$ and $BIC = 9346.066$ together with AIC and sample size adjusted BIC. The model modification index suggested significance of 3 across-component constraints of factor loadings, 13 correlated errors for the first component and 5 correlated errors for the second component. The parameter corresponds to most significant reduction of the likelihood function is ψ_{45} in component 1 and ψ_{78} in component 2. Let M_2 be the model after adding these two parameters, model modification index continues to suggest extra parameters that will significantly improve model-fit. The $l(\hat{\theta})$ and BIC correspond to M_2 are reported in Table 1(d), where the parameters in the parentheses are the extra parameters for the model over the previous one. Because BIC for M_2 still does not tell us how good the model-fit is, we proceed with model modification until model M_4 . Further model modification after M_4 leads to a model identification problem. We may stop at M_4 because it corresponds to the smallest BIC. But we still do not know whether model M_4 fits the data statistically. As discussed in the introduction, such a problem is the nature of mixture modeling where the likelihood ratio statistic does not asymptotically follow a known distribution, not due to a limitation of Mplus software.

We may notice that in Table 1(a) to (c), the two BICs for the saturated model at $m = 1$ and 2 differ most. This may imply that BIC with a correctly specified model is more effective in distinguishing the number of components than that with misspecified models.

We also like to note that this sample is not normally distributed. Its multivariate skewness is 283.54437, having a p -value of 2.672×10^{-8} when referred to χ^2_{165} . Its standardized multivariate kurtosis is 3.037, having a p -value of 0.001 when referred to $N(0, 1)$ (Mardia, 1970). Thus, at $n = 145$, moderate nonnormality of the sample does not make BIC choose a 2-component model. We will study the effect of sample size and distribution violations on BIC in section 4.

For the sample in this example, few researchers would fit it using a one-factor model because, since Jöreskog (1969), the data set has been studied by many researchers using various methods. The example is just to illustrate a typical case of application of mixture modeling when the structure of the population is not so well understood and the number of components is not clear. When both of these are clear, what left in applying a mixture model is just a standard ML procedure for generating parameter estimates and their SEs. Then there is no need for using the proposed two-stage ML.

Example 2

This example contrasts the single-stage and two-stage ML methods when model is misspecified. A sample² of size $n = 400$ is generated from the population

$$f(\mathbf{x}) = \pi_1 N_9(\mu_1, \Sigma_1) + \pi_2 N_9(\mu_2, \Sigma_2),$$

where $\pi_1 = 0.5$, $\pi_2 = 0.5$; $\mu_1 = \mathbf{0}$, $\mu_2 = (3, 3, \dots, 3)'$,

²For readers who want to replicate the study, the sample was generated using SAS IML code (proc IML; seed=1111111111; x=j(400,9,0); do i=1 to 400; ui=uniform(seed); if ui<0.5 then do; yi1=sigh1*normal(j(9,1,seed))+mu1; x[i,]=yi1'; end; else do; yi2=sigh2*normal(j(9,1,seed))+mu2; x[i,]=yi2'; end; end;). In the above notation, mu1 is the mean vector of the first population, $sigh1 = \Sigma_1^{1/2}$ is the 9×9 symmetric matrix that satisfies $\Sigma_1^{1/2} \Sigma_1^{1/2} = \Sigma_1$; mu2 is the mean vector of the second population, $sigh2 = \Sigma_2^{1/2}$ is the 9×9 symmetric matrix that satisfies $\Sigma_2^{1/2} \Sigma_2^{1/2} = \Sigma_2$.

$$\Sigma_1 = \Lambda_1 \Phi_1 \Lambda_1' + \Psi_1 \quad \text{and} \quad \Sigma_2 = \Lambda_2 \Phi_2 \Lambda_2' + \Psi_2,$$

with

$$\Lambda_1 = \begin{pmatrix} 1.0 & 1.0 & 1.0 & 0 & 0 & 0 & 1.0 & 0 & 0 \\ 1.0 & 0 & 0 & 1.0 & 1.0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 & 0 & 1.0 & 1.0 & 1.0 \end{pmatrix}', \quad \Phi_1 = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix},$$

$$\Psi_1 = \mathbf{I}_9;$$

$$\Lambda_2 = \begin{pmatrix} 1.0 & 1.0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0 & 1.0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.0 & 1.0 & 1.0 \end{pmatrix}', \quad \Phi_2 = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix},$$

all the diagonal elements of Ψ_2 are 1.0, nonzero off-diagonal elements of Ψ_2 are $\psi_{15} = \psi_{51} = 0.8$, $\psi_{68} = \psi_{86} = 0.8$. It is easy to see that the nine marginals in each component have the same variance of 2.0, and the marginal means of the two components are $3/\sqrt{2} \approx 2.121$ standard deviation apart.

We fit the sample by $m = 1$ -, 2- and 3-component normal mixtures with a saturated mean vector for each component; the variance-covariance matrix for each component is a three-factor model,

$$\Sigma_j(\theta_{cj}) = \Lambda_j \Phi_j \Lambda_j' + \Psi_j, \quad j=1, 2, 3,$$

where each Λ_j is a 9×3 matrix such that each factor is measured by three unidimensional indicators; each Φ_j is a free 3×3 matrix; and each Ψ_j is a diagonal matrix so that the errors in each component are specified as uncorrelated. The first factor loading for each factor (λ_{11} , λ_{42} and λ_{93}) is fixed at 1.0 for model identification. Thus, the factor loading matrix for the first component is misspecified and the error variance-covariance matrix for the second component is misspecified. Such misspecified structural models better reflect the practice of normal mixture SEM, where it is unlikely to have a model that perfectly follows the unknown population³. With 50 random starting values for the two- and three-component normal mixture models using the EM algorithm, the obtained values of the log likelihood and BIC are given in Table 2(a). Although BIC still chooses two normal components with a misspecified structural model, the estimates $\hat{\pi}_1$ and $\hat{\pi}_2$ contain substantial biases. If we stop at this three-factor model with unidimensional indicators and start to elaborate on the parameter estimates, then we are short of the truth. Even when the model for each component is correctly specified, BIC or other model selection criteria with single-stage ML still does not allow us to endorse the model, as has been illustrated. The two-stage ML

³Although most researchers agree that models are at best only approximations to the real world, there is no agreement on how much difference between the population and initial theoretical model would best represent reality. We may think that the variables in Example 1 are well designed and the sample is well collected. There is still a significant gap between the initial unidimensional three-factor model and the sample/population. Actually, with $\lambda_{11} = \lambda_{42} = \lambda_{93} = 1$ for model identification, the loading estimates are $\hat{\lambda}_{21} = .442$, $\hat{\lambda}_{62} = 2.242$; $\hat{\lambda}_{31}$, $\hat{\lambda}_{52}$, $\hat{\lambda}_{83}$, $\hat{\lambda}_{93}$ are around 1.0. The extra loading identified by the LM test, $\hat{\lambda}_{91} = 3.368$, is the greatest and is almost triple the average of the loadings in the initial model. If letting the LM test search for correlated errors in the initial model, three error covariances are identified as being significant: $\hat{\psi}_{78} = 172.957$, $\hat{\psi}_{23} = 5.451$, and $\hat{\psi}_{17} = -20.624$. All are larger in absolute value than the smallest error variance $\hat{\psi}_{44} = 2.832$.

allows us to evaluate the goodness of model-fit and locate which model causes the problem of lack-of-fit.

Table 2(b) contains the model fitting statistics for $m = 1, 2$ and 3 components with saturated means and covariances; those for $m = 2$ and 3 are obtained with 50 random starting values. As expected, BIC also chooses $m = 2$ as the proper number of components and the estimates $\hat{\pi}_1, \hat{\pi}_2$ are now more appropriate. Let $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ be estimates of the two saturated covariance matrices; $\hat{\Gamma}_{cin11}/n$ and $\hat{\Gamma}_{csw11}/n$ be estimates of the asymptotic covariance matrices of $\hat{\sigma}_1 = \text{vech}(\hat{\Sigma}_1)$ using $\hat{\Gamma}_{in}$ and $\hat{\Gamma}_{sw}$, respectively; and $\hat{\Gamma}_{cin22}/n$ and $\hat{\Gamma}_{csw22}/n$ be estimates of the asymptotic covariance matrices of $\hat{\sigma}_2 = \text{vech}(\hat{\Sigma}_2)$ using $\hat{\Gamma}_{in}$ and $\hat{\Gamma}_{sw}$, respectively. We next fit the two covariance matrices using EQS, starting with the unidimensional three-factor model and default LM test.

Appendix C contains the EQS syntax of fitting $\hat{\Sigma}_1$ and using $\hat{\Gamma}_{cin11}$ for consistent SEs and proposed statistics, where the asymptotic covariance matrix $\hat{\Gamma}_{cin11}$ is within the file "Gamma_cin11.dat". Because the population distribution is truly a mixture normal, we expect $\hat{\Gamma}_{cin11}$ to provide a better estimate of Γ_{c11} than $\hat{\Gamma}_{csw11}$. Notice that the sample size is set at $n = 400$ according to the result in (4). Table 3(a) contains the proposed statistics for overall model evaluation, these are part of the output of the EQS program. We also list the likelihood ratio statistic T_{ML1} in the table to show that it does not work well in the second-stage of the two-stage ML. The results under the LM test in the last column of Table 3(a) are also from the EQS output, only one parameter corresponding to the most significant LM statistic is reported⁴. Model M_1 is the three-factor model with unidimensional indicators; M_2 to M_5 are the modified models by adding the parameter suggested by the LM test based on running the previous model. Below each of the six statistics is the p -value associated with the proposed distribution for the statistic⁵; those below .001 are not reported. According to these statistics, model M_1 does not fit $\hat{\Sigma}_1$ well. The default LM test in EQS suggests that allowing the first variable to load on the second factor would reduce the chi-square value (T_{ML1}) by approximately 68.508. Although M_2 fits $\hat{\Sigma}_1$ better than M_1 , all the statistics are still highly significant. The first model with a good fit is M_4 , which is identical to the population structure that generated the data. At M_4 , the LM test continues to suggest adding parameter λ_{61} , adding this parameter leads to model M_5 with an even better fit, although both the absolute value of the estimate $\hat{\lambda}_{61} = 0.224$ and the corresponding z -score ($= 2.410$) are the smallest among all the parameter estimates. Then the LM test further suggests adding λ_{82} ; following the suggestion leads to an estimate $\hat{\lambda}_{82} = -0.353$ and a corresponding $z = -1.186$ ⁶. So we may regard model M_4 or M_5 as our final model.

Parameter estimates by the single-stage ML for M_1 and the two-stage ML for M_1 to M_5 are reported in Table 3(c). Due to model misspecification, the three estimates of factor variances under the single-stage ML have biases over 1.0; four of the six factor loading estimates have biases over .50. The estimates by the two-stage ML under M_1 have even more biases, mainly because a smaller model has less freedom to buffer the discrepancy between the data and the model. The biases in parameter estimates by the two-stage ML become smaller as the model moves from M_1 to M_4 . Model M_5 is still correctly specified although it is over-parameterized. Parameter biases under M_5 are still a lot smaller than those under M_1 by both single-stage and two-stage ML.

⁴Under the LM test in EQS output, there are other multivariate sequential statistics for improving model-fit. Here we report only the univariate LM test, which is equivalent to the model modification index in LISREL and Mplus.

⁵The p -value under a LM statistic is obtained by comparing the LM statistic to χ^2_1 .

⁶The LM test in EQS is based on the normal distribution assumption, the z -score reported here uses SEs based on the sandwich-type covariance matrix in equation (9).

Parallel to those in Table 3(a), Table 3(b) contains the statistics for overall model evaluation when $\hat{\Gamma}_{csw11}$ is used to obtain the proposed statistics. Because both $\hat{\Gamma}_{cin11}$ and $\hat{\Gamma}_{csw11}$ are consistent, the results in Table 3(a) and (b) are very comparable, though there exist some differences due to more sampling errors in $\hat{\Gamma}_{csw11}$. In particular, the proposed statistics do not support the structural model until M_4 . The LM test continues to suggest M_5 although M_4 is the true model. Notice that the parameter estimates only depend on $\hat{\Sigma}_1$, using $\hat{\Gamma}_{cin11}$ or $\hat{\Gamma}_{csw11}$ generates the same set of $\hat{\theta}_{c1}$, as reported in Table 3(c). The SEs of $\hat{\theta}_{c1}$ (not reported to save space) associated with $\hat{\Gamma}_{cin11}$ and $\hat{\Gamma}_{csw11}$ have only a small difference. For example, the z -score ($= 2.541$) for $\hat{\lambda}_{61} = 0.224$ in M_5 is the smallest among all the parameter estimates; the z -score ($= -1.166$) for $\hat{\lambda}_{82} = -0.353$ following the modification of M_5 is not statistically significant at the .05 level.

Table 4(a) contains the results of fitting $\hat{\Sigma}_2$ using $\hat{\Gamma}_{cin22}$ to obtain the proposed statistics, starting from the three-factor model with unidimensional indicators. As expected, model M_1 does not fit $\hat{\Sigma}_2$ well, none of the statistics has a p -value close to .001. Following the default LM test in EQS, we fit the model M_2 (adding λ_{63}), M_3 (further adding λ_{12}) and M_4 (further adding λ_{82}) in sequence. None of the proposed statistics endorses any of the 4 models. There exists a convergence problem when further including the parameter λ_{51} as evaluated by EQS, due to parameters being linearly dependent. Notice that the default LM test in EQS only identifies needed factor loadings, other needed parameters may have to be requested specifically. Starting from model M_1 , we can tell the LM test to search for correlated errors by adding “SET=PEE;” immediately below the default LM test. The sequence of models identified by the new LM test are M_5 (adding ψ_{86}), M_6 (further adding ψ_{51}), and M_7 (further adding ψ_{98}). It is clear from the lower portion of Table 4(a) that model M_5 is not statistically supported by any of the test statistics. Model M_6 is endorsed by all the statistics and is also the true model. When evaluating the over-parameterized model M_7 , both the estimate $\hat{\psi}_{98}$ ($= 0.297$) and its z -score ($= 2.638$) are the smallest among all the estimates. At M_7 , the LM test continues to suggest adding ψ_{95} . When evaluated, the z -score corresponding to $\hat{\psi}_{95}$ equals 1.721, not statistically significant at the .05 level, so we stop the model modification process.

The parameter estimates for θ_{c2} by single-stage ML for model M_1 and two-stage ML for M_1 to M_7 are reported in Table 4(c). As expected, those under M_6 have the least overall bias. Models M_2 to M_4 all have negative error variance estimates, suggesting possible model misspecification although none of them is statistically significant at level .05. Compared to Table 3(c), we may find that omitting factor loadings leads to more biases in parameter estimates than omitting error covariances.

When $\hat{\Gamma}_{csw22}$ is used to obtain the proposed statistics, the results are reported in Table 4(b). Similar to those for modeling the first component, results in Tables 4(a) and 4(b) are very comparable. In particular, none of the statistics supports models M_1 to M_5 . The estimate $\hat{\psi}_{98}$ has the smallest z -score of 2.150 in model M_7 . Further adding ψ_{95} leads to a none significant z -score of 1.723.

This example shows that all the knowledge accumulated in conventional SEM can be used to analyze the model structure for an individual component in stage-2 of the two-stage ML. In particular, the statistics that have been shown to work well in conventional SEM models allow us to judge whether a covariance matrix (mean vector as well) for an individual component is adequately explained. In the context of mixture analysis, when we have little knowledge about the population structure, these statistics offer us more guidance than is possible with single-stage ML.

Example 3

This example aims to see the robustness of two-stage ML to moderate distributional violations together with model misspecifications. Single-stage ML will not be studied in this example because it does not allow us to judge the goodness-of-fit in each component. Instead of sampling from a mixture normal distribution, a sample⁷ of size $n = 400$ is generated from the population

$$f(\mathbf{x}) = \pi_1 M_{t9}(\mu_1, \Sigma_1; 8) + \pi_2 M_{t9}(\mu_2, \Sigma_2; 8),$$

where $\pi_1 = 0.5$, $\pi_2 = 0.5$, $M_{t9}(\mu_1, \Sigma_1; 8)$ and $M_{t9}(\mu_2, \Sigma_2; 8)$ represent two 9-variate t -distributions, each with 8 degrees of freedom; the population means μ_1 and μ_2 , and covariance matrices Σ_1 and Σ_2 are identical to those in Example 2.

As in Example 2, we first fit the sample by saturated models at stage-1, using BIC to choose the number of components. With 50 random starting values for $m = 2$ and 3 components, the log likelihood $l(\hat{\beta})$ and BIC are reported in Table 5. Obviously, using BIC leads to the correct number of components. The two estimates $\hat{\pi}_1$ and $\hat{\pi}_2$ are also reasonable. The two estimated covariance matrices, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ and the asymptotic covariance matrices $\hat{\Gamma}_{cin11}$, $\hat{\Gamma}_{csw11}$, $\hat{\Gamma}_{cin22}$, $\hat{\Gamma}_{csw22}$ will be used for the second stage analysis below. Note that $\hat{\Gamma}_{cin11}$ is not consistent for Γ_{c11} nor is $\hat{\Gamma}_{cin22}$ for Γ_{c22} . We include them in the study to see the difference between using a consistent estimator and an inconsistent estimator.

Parallel to Table 3(a) for Example 2, the results of fitting $\hat{\Sigma}_1$ starting with the three-factor unidimensional-indicator model (M_1) and using $\widehat{Cov}(\hat{\sigma}_1) = \hat{\Gamma}_{cin11}/n$ are reported in Table 6(a). The model M_4 fits the sample adequately by all standards, which is expected because it is the true model. At M_4 , the LM test continues to suggest adding the parameter λ_{82} , but a further modified model leads to a z -score of -1.261 for the estimate $\hat{\lambda}_{82}$. So we stop at model M_4 . The parameter estimates $\hat{\theta}_{c1}$ for models M_1 to M_4 are reported in Table 6(c). Most of the parameter estimates under M_4 are very close to their population values, but two error variance estimates have biases close to 1.0. This is because the multivariate t -distribution has heavier tails than that of the corresponding normal distribution. The estimates $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are not efficient due to the heavier tails. Any discrepancy between $\hat{\Sigma}_1$ and the population Σ_1 will be inherited by the structural parameter estimates. Estimates by single-stage ML will suffer from the same kind of lack of efficiency.

Table 6(b) contains the results of fitting the same set of models as in Table 6(a) but the matrix $\hat{\Gamma}_{csw11}$ is used to obtain the proposed statistics. The results in Table 6(b) are comparable to those in Table 6(a) except all the models are judged more favorably by the four proposed statistics. This is because the sandwich-type covariance matrix takes the heavy tails in the underlying distribution into account, which means that $\hat{\mathbf{B}} > \hat{\mathbf{A}}$ in the sense of positive definiteness. Thus, part of the discrepancy between $\hat{\Sigma}_1$ and the fitted model is accounted for by the heavy tails. The rest is due to the systematic difference between the data “ $\hat{\Sigma}_1$ ” and the model $\Sigma_1(\theta_{c1})$. Although $\hat{\mathbf{B}} > \hat{\mathbf{A}}$ and model M_3 fits the data pretty well, the z -score corresponding to $\hat{\lambda}_{71} = 0.795$ is 2.793. So the parameter λ_{71} is still needed

⁷For readers who want to replicate the study, the sample was generated using SAS IML code (proc IML; alpha=4; sight1=sigh1*sqrt(3/4); sight2=sigh2*sqrt(3/4); seed=1111111111; x=j(400,9,0); do i=1 to 400; ui=uniform(seed); if ui<0.5 then do; ui=rangam(seed,alpha)/alpha; sighi1=sight1/sqrt(ui); yi1=sighi1*normal(j(9,1,seed))+mu1; x[i,]=yi1; end; else do; ui=rangam(seed,alpha)/alpha; sighi2=sight2/sqrt(ui); yi2=sighi2*normal(j(9,1,seed))+mu2; x[i,]=yi2; end; end;). In the above notation, mu1 is the mean vector of the first population, sigh1= $\Sigma_1^{1/2}$ is the 9×9 symmetric matrix that satisfies $\Sigma_1^{1/2} \Sigma_1^{1/2} = \Sigma_1$; mu2 is the mean vector of the second population, sigh2= $\Sigma_2^{1/2}$ is the 9×9 symmetric matrix that satisfies $\Sigma_2^{1/2} \Sigma_2^{1/2} = \Sigma_2$.

statistically. Due to greater SEs using the sandwich-type covariance matrices, the z -score associated with $\hat{\lambda}_{82}$ is -1.131 , smaller than that based on $\hat{\Gamma}_{cin11}$. Thus, λ_{82} is not needed statistically. The parameter estimates corresponding to Table 6(b) are identical to those for Table 6(a), as reported in Table 6(c).

Both $\hat{\Gamma}_{cin11}$ and $\hat{\Gamma}_{csw11}$ thus allow us to evaluate the model for component 1, and using $\hat{\Gamma}_{csw11}$ gives us more confidence in the results due to larger p -values associated with the four proposed statistics at model M_4 and a smaller z -score for the unnecessary parameter λ_{82} .

Turning to the second component, the results of fitting Σ_2 using $\widehat{Cov}(\hat{\sigma}_2) = \hat{\Gamma}_{cin22}/n$ are reported in Table 7(a). The program has a convergence problem due to linearly related parameters when further adding λ_{63} after M_4 . The LM test for correlated errors, starting from M_1 , suggests a sequence of new models (M_5 to M_8). The model M_6 (containing the covariance parameters λ_{51} and λ_{86}) are evaluated favorably by all the proposed statistics, which is expected because it is the true model. Continuing the model modification after M_8 leads to a z -score of 1.774 for $\hat{\psi}_{98} = 0.318$. Both $\hat{\psi}_{42}$ in M_7 and $\hat{\psi}_{95}$ in M_8 are statistically significant at .05 level, although the estimates and their associated z -scores are the smallest among all the parameter estimates.

The parameter estimates for M_1 to M_8 are reported in Table 7(c). There is a negative estimate of error variance in M_1 ; including λ_{51} (M_2) results in three negative estimates of error variances; $\hat{\psi}_{66}$ continues to be negative until M_4 . Although none of the z -scores corresponding to these negative estimates is statistically significant, literature on conventional SEM suggests that negative estimates of error variances are associated with model misspecification. There are no negative error variance estimates in models M_5 to M_8 . The parameter estimates from M_5 to M_8 are very comparable although M_5 is literally a misspecified model. Such a phenomenon was also observed for Example 2, where omitting error covariances causes less biases than omitting relevant factor loadings.

Table 7(b) contains the results of fitting Σ_2 when $\hat{\Gamma}_{csw22}$ is used to obtain the proposed statistics. Similar to those in Table 7(a), all the p -values are below .001 until model M_6 . Different from those in Table 7(a) is that $\hat{\psi}_{95} = 1.817$ is not statistically significant at the .05 level. Actually, the z -score for $\hat{\psi}_{42} = 0.414$ in M_7 equals 1.970 , only marginally significant at the .05 level. If one thinks that an error covariance with such a significance level should not be included in the model, then we end up with M_6 , which is the model that generates the population. For the second component, both $\hat{\Gamma}_{cin22}$ and $\hat{\Gamma}_{csw22}$ lead us to the true model. But using $\hat{\Gamma}_{csw22}$ allows us to identify the true model easier due to less significant z -scores for unnecessary parameters.

This example shows that, when the component distributions and the structural models are misspecified, two-stage ML with its diagnostic tools still leads us to correct or nearly correct models. It also shows that, when the underlying component distributions have heavier tails than that of the normal distribution, using the sandwich-type covariance matrices to obtain the proposed statistics and SEs will lead to better model evaluation.

4. Efficiency of Parameter Estimates and Robustness of BIC

This section contains two Monte Carlo studies. The first aims to see how much better single-stage ML is over two-stage ML in parameter estimates under the conditions that single-stage ML is designed for. The second study compares the performance of BIC in selecting the correct number of components when distribution conditions are violated. Our purpose is to see whether BIC for the saturated model performs any better than that for a structured model.

4.1 Efficiency of parameter estimates under idealized conditions

Since efficiency of parameter estimates in mixture model is closely related to the separation of components, two normal mixture models are used for the study. Both have two components while they differ in separations.

For the first mixture population, $\pi_1 = \pi_2 = 1/2$; $\mu_1 = \mathbf{0}$, $\mu_2 = (3, 3, \dots, 3)'$; the two covariance matrices are equal, each is specified through a confirmatory factor model with

$$\Sigma = \Lambda \Phi \Lambda' + \Psi,$$

where

$$\Lambda = \begin{pmatrix} 1.0 & 1.0 & 1.0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 1.0 & 1.0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & & 0 & 0 & 1.0 & 1.0 & 1.0 \end{pmatrix}, \Phi = \begin{pmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{pmatrix}, \quad (15)$$

$\Psi = \mathbf{I}_9$. Same as in Examples 2 and 3, the marginal means of the two components are $3/\sqrt{2} \approx 2.121$ standard deviation apart. Parameters in the second mixture population are the same as in the first one except $\mu_2 = (1.5, 1.5, \dots, 1.5)'$. Thus, the marginal means of the two components are approximately 1.061 standard deviation apart. The sample size is $n = 400$; 500 replications are generated from each of the two populations.

The same model is used when fitting the two populations: The means are saturated for both components; the covariance structures for the two components are specified correctly and estimated independently, that is, each nonzero entry of the population matrices Λ , Φ and Ψ is specified as a free parameter in models for both components except λ_{11} , λ_{42} and λ_{73} , which are fixed at 1.0 for model identification. We apply single-stage ML and two-stage ML to each sample. EM algorithms are used for single-stage ML and the first stage of two-stage ML. Fisher-scoring algorithm is used for each component in the second stage of two-stage ML. Because the true population values of each component are known, they are used as the starting values instead of multiple random starting values in both the EM and Fisher-scoring algorithms. At the convergence of each sample we obtain, for each component, the estimates for the six free factor loadings, six factor variances-covariances, and nine error variances. We also obtain 9 saturated means for each component, but we will only focus on the structural parameters of the covariance matrices.

Let $\hat{\theta}_i$ be the estimate of parameter θ at the i th replication and θ_0 be its population value;

$$\text{mean} = \frac{1}{500} \sum_{i=1}^{500} \hat{\theta}_i,$$

$$\text{bias} = \frac{1}{500} \sum_{i=1}^{500} \hat{\theta}_i - \theta_0,$$

and

$$\text{SE} = \left[\frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_i - \text{mean})^2 \right]^{1/2}.$$

Table 8(a) and (b) contains the results of fitting the first and second components of the first population, respectively. At the bottom of each table are the average absolute bias, average variance, and average mean square error (MSE) across the $q = 21$ covariance parameters. These empirical results indicate that, under ideal conditions with component means being about 2 standard deviation apart, single-stage ML does yield more accurate parameter estimates. But the differences in biases, variances, and MSEs between the two procedures are in the third decimal place.

Table 9(a) and (b) contains the parallel results for the second population. With component means approximately one standard deviation apart, both single-stage ML and two-stage ML lead to factor loadings and error variances with little biases. Actually, all the biases for these two sets of parameters following two-stage ML are in the 2nd decimal place while most of those following single-stage ML are also in the 2nd decimal place. But there exist substantial positive biases in the estimates of factor variances and covariances. On average, two-stage ML results in slightly more biases than single-stage ML in Table 9(a) while the two methods have comparable biases in Table 9(b). With respect to efficiency and accuracy, single-stage ML performs better in Table 9(a) and worse in Table 9(b). Overall, two-stage ML leads to slightly more efficient and accurate parameter estimates but with slightly more biases. Although we expect single-stage ML to perform uniformly better, the results in Table 9(b) are not a surprise because statistical theory only says that single-stage ML generates most efficient estimator asymptotically.

We would like to note that the results in Table 9(a) and (b) are based on 491 and 492 converged replications, respectively. All the nonconvergences occurred with the Fisher-scoring algorithm at the second stage of the two-stage ML, which is expected because nonconvergences have been repeatedly reported when fitting conventional SEM models (see e.g., Boomsma, 1985; Anderson & Gerbing, 1984). Neither the single-stage ML nor the first stage of the two-stage ML has any convergence problem due to the fact that EM algorithm always converges to a local/global maxima (Wu, 1983). The results for single-stage ML in Table 9 were based on the same sets of samples that the Fisher-scoring algorithm converged. Although EM algorithm always converges, the rate of convergence with the structured model is slow. On a Pentium(R)4CPU3.40GHz desktop computer, for the second population, the single-stage ML⁸ used 43 minutes to complete the simulation while the two-stage ML only took 9 minutes.

Comparing Table 8 and Table 9, we may notice that model separation has a strong effect on the efficiency of all the parameter estimates by both two-stage ML and single-stage ML. With respect to bias, its effect is mainly on the estimates of factor variances-covariances, not much on those of factor loadings or error variances. This implies that, when the model separation is one standard deviation apart, the estimates of the structural model in finite normal mixture SEM may not be reliable by either of the methods even when the underlying population distribution is truly a mixture normal.

The obtained empirical results on efficiency and MSE are informative but limited. A comprehensive Monte Carlo study might include conditions when the distribution for each component is not normally distributed as well as when the proportions of the components, the separation of the different components, the number of components and the sample size vary.

⁸The convergence for the single-stage ML is set as $l^{(j+1)} - l^{(j)} < .0001$, where $l^{(j)}$ is the log likelihood function evaluated after the j th iteration. The convergence for the first stage of the two-stage ML is also set as $l^{(j+1)} - l^{(j)} < .0001$, and for the second stage it is set as when the sum of squares of the difference between $\theta^{(j+1)}$ and $\theta^{(j)}$ is less than .0001, where $\theta^{(j)}$ is the vector of parameter after the j th iteration.

4.2 Robustness of BIC against distribution violations

The population in this subsection has one component while its distribution varies. We study the performance of the BICs when comparing the one-component saturated model against the two-component saturated model and when comparing a one-component structured model against a two-component structured model. The model corresponding to a smaller BIC is empirically preferred. Ideally, the empirically preferred model is always the one-component model. Because the saturated model is a special case of a structured model, which has been shown to not work well (Bauer & Curran, 2003), we do not expect that BIC with the saturated model performs ideally.

Let

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{e}, \quad (16)$$

where $\boldsymbol{\mu}$ is a 9×1 vector of 1.0's, $\boldsymbol{\Lambda}$ is as given in (15), $\mathbf{f}$ is a 3×1 random vector with $E(\mathbf{f}) = \mathbf{0}$ and $\text{Cov}(\mathbf{f}) = \boldsymbol{\Phi} = (\phi_{jk})$ also as given in (15), and $\mathbf{e} = (e_1, e_2, \dots, e_9)'$ is a 9×1 random vector with $E(\mathbf{e}) = \mathbf{0}$ and $\text{Cov}(\mathbf{e}) = \boldsymbol{\Psi} = \mathbf{I}_9$. Let $\boldsymbol{\Phi}^{1/2}$ be the positive definite symmetric matrix that satisfies $\boldsymbol{\Phi}^{1/2}\boldsymbol{\Phi}^{1/2} = \boldsymbol{\Phi}$ and

$$\mathbf{f} = \boldsymbol{\Phi}^{1/2}\mathbf{z}_3, \quad (17)$$

where $\mathbf{z}_3 = (\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)'$ and $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$ are independent and standardized random variables.

Eleven distribution conditions are listed in the first column of Table 10, where $\mathbf{f} \sim \chi_m^2(\mathbf{0}, \boldsymbol{\Phi})$ implies that $\mathbf{f}$ is generated according to (17) with z_1 to z_3 each following independent and standardized⁹ χ_m^2 ; the notation $\mathbf{e} \sim \chi_m^2(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \mathbf{I}$ implies that e_1 to e_9 are independent and each follows independent and standardized χ_m^2 . Similarly, the notation $\mathbf{f} \sim \log N(\mathbf{0}, \boldsymbol{\Phi})$ implies that $\mathbf{f}$ is generated according to (17) with z_1 to z_3 each following independent and standardized¹⁰ $\log N(0, 1)$; and $\mathbf{e} \sim \log N(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \mathbf{I}$ implies that e_k are independent and each follows a standardized $\log N(0, 1)$.

Two sample sizes, $n = 500$ and 1000 , are chosen to see the effect of n on BIC. Each sample is fitted by four models: (I) one-component model with saturated means and covariances; (II) two-component model with two saturated mean vectors and covariance matrices; (III) one-component factor model as in (16) with a saturated mean vector, $\boldsymbol{\Lambda}$ containing 6 free factor loadings together with $\lambda_{11} = \lambda_{42} = \lambda_{73} = 1$, corresponding to the nine none zero elements in (15), $\boldsymbol{\Phi}$ being a free covariance matrix and $\boldsymbol{\Psi}$ being a diagonal matrix; (IV) two-component model with two saturated mean vectors, two factor loading matrices each contains 6 free factor loadings, two free factor covariance matrices and two diagonal error covariance matrices. Obviously, both the one-component saturated and structured models are correct while both the two-component models are over-parameterized or incorrect.

The EM algorithm was used in both the saturated and structured two-component models. Fisher-scoring algorithm was used in estimating the one-component structured model. The starting values for the one-component structured models are set at those corresponding to the population values that generate the data. For the starting values in the two-component models, the sample was first split by comparing the simple summation of the 9-observed variables to the summation of the population mean (which is 9 here), cases with $x_{i1} + x_{i2} +$

⁹For a given random variable x , its standardized version is obtained by $z_x = [x - E(x)]/\{\text{Var}(x)\}^{1/2}$.

¹⁰A random variable x following the log-normal distribution $\log N(0, 1)$ is obtained by $x = \exp(z)$ and $z \sim N(0, 1)$.

$\dots + x_{i9} > 9$ are in one group and the rest are in another group. Let $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ be the vectors of sample means of the two groups. Ten starting values for the means of the two components are randomly drawn from $N_9(\bar{\mathbf{x}}_1, \mathbf{S})$ and $N_9(\bar{\mathbf{x}}_2, \mathbf{S})$, respectively, where $\mathbf{S}$ is the sample covariance matrix of the whole sample. The starting values for the sample covariance matrices for the two-component saturated model are both set at $\mathbf{S}$. The starting value of the factor loading matrices, factor covariance matrices and unique variances for the two-component structured model are both set at the population values that generated the data. Because the log likelihood function never decreases in the EM algorithm, we define the convergence as when log likelihood function increases less than 0.0001 from the previous iteration. Let $\tau_1 \geq \tau_2 \geq \dots \geq \tau_p$ be the eigenvalues of the covariance matrix of either of the two components. To avoid converging to near singular covariance matrices we define the current replication as unable to reach a convergence whenever $\tau_p/\tau_1 < 0.001$ for either of the components at the end of an M-step.

With 500 replications, the frequencies for BIC to choose the one-component models for different distribution conditions are reported in Table 10. The number of non-converged replications (after all the 10 starting values) are reported in parentheses, all of which occurred with the two-component models. When the population is normally distributed, BIC always chooses the correct number of components when evaluated at either the saturated or structured models for both $n = 500$ and 1000. At $n = 500$, when $\mathbf{f} \sim \chi_3^2(\mathbf{0}, \Phi)$ and $\mathbf{e} \sim N(\mathbf{0}, \Psi)$, BIC with the saturated model identifies the correct number of components 100% of the time while that with the structured model identifies the correct number of components 92% of the time. At $n = 500$, BIC also performs better for the saturated model than for the structured model under the conditions with moderate distribution violations. In particular, when $\mathbf{x} \sim Mt(\mu, \Sigma, 8)$, BIC with the saturated model chooses the correct number of components 340 times while that with the structured model only chooses the correct model 6 times; when $\mathbf{f} \sim N(\mathbf{0}, \Phi)$ and $\mathbf{e} \sim \chi_3^2(\mathbf{0}, \mathbf{I})$, BIC with the structured model always chooses the wrong number of components while BIC with the saturated model still chooses the correct number of components about 75%; when $\mathbf{f} \sim \chi_1^2(\mathbf{0}, \Phi)$ and $\mathbf{e} \sim N(\mathbf{0}, \Psi)$, BIC with the saturated model selects the correct number of components 46% while BIC with the structured model selects the correct number of components only once. At $n = 1000$, BIC with the saturated model obviously performed better when $\mathbf{f} \sim \chi_3^2(\mathbf{0}, \Phi)$ and $\mathbf{e} \sim N(\mathbf{0}, \Psi)$. For the rest of the nonnormal distribution conditions at $n = 1000$, neither of the BICs could identify the correct number of components.

Notice that the skewness and kurtosis values of the normal distribution, the t -distribution with 8 degrees of freedom, the t -distribution with 5 degrees of freedom, the chi-square distribution with 3 degrees of freedom, the chi-square distribution with 1 degree of freedom, and the lognormal distribution are respectively (0, 0), (0, 1.5), (0, 6), (1.633, 4), (2.828, 12), and (6.185, 110.936). Obviously, the ability of BIC to choose the one-component model decreases when either $\mathbf{f}$ or $\mathbf{e}$ departs from normality as well as when n increases. At $n = 500$, when $\mathbf{e} \sim N(\mathbf{0}, \Psi)$, BIC for both the saturated and structured models can endure mild nonnormality of $\mathbf{f}$. At $n = 1000$ and when $\mathbf{e} \sim N(\mathbf{0}, \Psi)$, BIC for the saturated model also allows mild nonnormality of $\mathbf{f}$ while BIC for the structured model is much less tolerant. The results for the structured model in Table 10 agree well with what Bauer and Curran (2003) have found.

It is nice to see that BIC with the saturated model performs better than that with the structured model, especially for moderate n . But BIC with the saturated model is still strongly affected by distribution violations at a large n . This is because BIC itself is constructed using the normal distribution assumption. As the sample size increases, a slight departure from normal distribution will be magnified at the sample level. When the sample

size is huge, a slight distribution violation might cause BIC to select more components than are needed.

In this section we only studied the performances of the two BICs when the underlying population distribution has one component. A more thorough Monte Carlo study might include conditions when the underlying population has more than one component and the separation of the components varies. Different components may have different distributions as well as different model structures. The proportions of different components might also vary.

5. Conclusion and Discussion

In this paper we developed a two-stage ML approach to normal mixture SEM. The most important feature of two-stage ML is that all the techniques accumulated in the conventional SEM literature can be used to study the model structure for each component. In particular, four statistics can be used to judge whether a particular component is adequately fitted by the structural model. Another feature is that the components are segregated at stage-2, so that model misspecification in one component does not affect the estimation and evaluation of models for other components. The third feature is its flexibility in modeling different components with different models. One component may be fitted by a SEM model suggested by a firm substantive theory while the structure of another component can be explored using principal components or exploratory factor analysis. Also associated with two-stage ML is its readiness to adopt any advances in the mixture model literature. A new statistical development in mixture model is typically directed first towards the model with saturated means and covariances. Once a more reliable method is available, we can immediately apply it to the first stage of the two-stage ML. Compared to single-stage ML, two-stage ML is also computationally more efficient¹¹. Arminger et al. (1999) noted that, with the second stage being a simultaneous GLS procedure, their two-stage approach is a lot faster than single-stage ML. Our limited experience also indicates that the proposed two-stage ML is a lot faster than single-stage ML. This is because in the second-stage of two-stage ML the dimension of the problem is much smaller and the discrepancy function in (6) or (7) is easier to minimize than the simultaneous ADF/GLS function.

Considering that any interesting model is at best only an approximation to the real world, throughout the paper we have emphasized valid statistical inference when conditions are not ideal. When there is a possibility that a population is heterogeneous, we may use a normal mixture to approximate the distribution. If $(\hat{\mu}_j, \hat{\Sigma}_j)$ is fitted well by a theoretical model $(\mu(\theta_j), \Sigma(\theta_j))$, then the mixture model confirms our theory for the j th component. Most likely, the initial model has to be modified. If the modified model can be explained via the substantive meaning of the variables, then the mixture model allows us to better understand the relationship among the variables for the j th component. It is possible that a well fitted model is fundamentally different from an established theory. Then one may need to reexamine the theory, or possibly a mixture model is not a proper statistical technique for extracting information from the data. The proposed statistics facilitate the validation of a theory with the two-stage ML approach to normal mixture SEM. We illustrated the application of the LM test for model modification because of its standard use in statistics (e.g., Buse, 1982) and its appropriate results when carried out in an a priori manner. We do not recommend empirical model modification without a theoretical rationale as a general practice, since it is well known that this can lead to capitalization on chance (e.g., MacCallum, Roznowski, &

¹¹A reviewer conjectured that, with many variables and many components, the single-stage ML may be computationally more efficient than two-stage ML. Considering that with 9 variables and 2 components single-stage ML took 4.7 times longer than two-stage ML, we cannot endorse the conjecture before further research.

Necowitz, 1992). In a purely exploratory context, other search methods also can be consulted (e.g., tabu search; Marcoulides, Drezner, & Schumacker, 1998).

The four proposed statistics for overall model evaluation have been shown to work well in conventional SEM. They also perform consistently in the examples in section 3. In practice, they should agree for most data sets. If they do not agree with a relatively large n , we should trust F_R or T_{CRADF} because their reference distributions are asymptotically correct. If the n is relatively small, we should trust T_{MLa} or T_{RML} more, because they depend less on the asymptotic covariance matrix Γ_{sw} . With a really small n , mixture modeling may not be the best way to do data analysis, since any of the statistics should be considered as a rough description rather than tests that control type I or II errors. Although tempting in current practice, the likelihood ratio statistic T_{ML} should never be used to evaluate the overall model structure at stage-2. With practical data, all the proposed statistics should be evaluated using the sandwich-type covariance matrix $\hat{\Gamma}_{swjj}$ or $\hat{\Gamma}_{cswjj}$; SEs should be estimated using (8) or (9) together with $\hat{\Gamma}_{swjj}$ or $\hat{\Gamma}_{cswjj}$. Of course, when one is 100% confident that the population distribution is mixture normal, using $\hat{\Gamma}_{injj}$ or $\hat{\Gamma}_{cinjj}$ will lead to more accurate inference due to less sampling error. In this paper we did not report any of the popular fit indices.

Obviously, they can be used to evaluate the overall model structure at stage-2. Because T_{ML} is no longer valid for model inference, these fit indices should be defined using T_{RML} , T_{MLa} or T_{CRADF} . We also would like to note that, although these statistics worked well with conventional SEM, further study on their behavior with mixture models is still valuable. In particular, the input for the second stage analysis depends on the goodness of the first stage ML, which has been shown to not work well when different components are not well separated and when sample size is not large enough (see e.g., Hosmer, 1973). Monte Carlo study on the behavior of these statistics would be informative when varying the sample size, the proportions of different components, as well as the separation among the components.

The Monte Carlo results in section 4 indicated that distribution violations affect the ability of BIC to identify the correct number of components, although BIC with the saturated model performs better than that with a structured model. Since such a phenomenon is rooted in the distributional assumption underlying BIC's formulation, it is wise to always cast doubt on the number of components corresponding to the smallest BIC. The four statistics at stage-2 of two-stage ML are especially valuable in providing alternative evaluations on the overall model structure for each component. When an individual component cannot be fitted adequately by a substantively meaningful model, it is likely that the number of components is not correctly identified. Then one may next study the individual component that corresponds to a smaller m with a greater BIC. If the components with smaller m can be adequately fitted by substantively meaningful models, then the mixture model with a smaller m is preferred. If the components with a smaller m still cannot be well fitted by substantively interesting models, it is most likely that finite normal mixture model is not a good description of the underlying population.

Arminger et al. (1999) mainly considered single-stage ML and a two-stage approach for models with covariates. The two-stage ML approach developed here can also be extended to models with covariates. One just needs to change the first-stage ML to conditional ML, using a sandwich-type covariance matrix to estimate the asymptotic covariance matrix of the conditional means and variances-covariances, parallel to (3). At the second stage, the conditional means and covariances are fitted by the model with the same covariates as at stage-1 through minimizing the normal distribution based discrepancy functions.

In the first stage of two-stage ML, we did not consider any constraints across covariances or means. These can also be easily incorporated into two-stage ML. For example, if there is a strong theory to support that the covariances are homogeneous across the mixture

components, then it is necessary and straightforward to incorporate such a constraint into the first stage. The second stage will be the same once the MLEs of the means and covariance matrices of the first stage are obtained together with $\hat{\Gamma}_{sw}$. In this paper, we did not consider any across-component constraints on parameters at stage-2. If such constraints are deemed necessary, then the m discrepancy functions at stage-2 need to be minimized simultaneously under the constraints. Standard errors and test statistics for overall model evaluation can be developed, similar to multiple-group analysis while the mean vectors and covariance matrices in different groups are correlated through the estimate of Γ_{sw} in (3). Because misspecification in such constraints is confounded with misspecification of the overall model in each component, it is wise to evaluate the adequacy of the model for each component first, parallel to the advised configural invariance test in multiple-group analysis.

We have only considered finite normal mixture SEM with continuous variables. The development in Muthén and Shedden (1999) and Muthén et al. (2002) allows the probabilities of categorical outcomes to be accounted for by covariates. Because the procedure is a single-stage ML, the limitation that no test statistics exist for the overall model evaluation also applies. Any future development to effectively evaluate the adequacy of the overall model with categorical outcome variables would be an important contribution to mixture modeling.

A final note is that the two-stage ML developed here will be available in the next version of EQS. The LM test at stage-2 will be based on the sandwich-type covariance matrix $\hat{\Gamma}_{sw}$, not based on the inverse of the information matrix, as is currently formulated when running the program in appendix C.

Acknowledgments

We would like to thank three reviewers and the editor for comments that helped in improving the paper.

Appendix A

This appendix contains the EM algorithm for obtaining the MLE of a mixture normal distribution with saturated means and covariances. For a m -component normal mixture, let the starting values be $\pi_1^{(0)}, \pi_2^{(0)}, \dots, \pi_{m-1}^{(0)}; (\mu_1^{(0)}, \Sigma_1^{(0)}), (\mu_2^{(0)}, \Sigma_2^{(0)}), \dots, (\mu_m^{(0)}, \Sigma_m^{(0)})$. The normal density function evaluated at $\mathbf{x}_i$ and $(\mu_j^{(0)}, \Sigma_j^{(0)})$ is denoted as $f_{ij}^{(0)}$. Let

$$\pi_m^{(0)} = 1 - \sum_{j=1}^{m-1} \pi_j^{(0)},$$

$$f_i^{(0)} = \sum_{j=1}^m \pi_j^{(0)} f_{ij}^{(0)},$$

and

$$w_{ij}^{(0)} = \frac{\pi_j^{(0)} f_{ij}^{(0)}}{f_i^{(0)}}.$$

Then the updated values are

$$\pi_j^{(1)} = \frac{1}{n} \sum_{i=1}^n w_{ij}^{(0)}, j=1, 2, \dots, (m-1);$$

$$\mu_j^{(1)} = \frac{\sum_{i=1}^n w_{ij}^{(0)} \mathbf{x}_i}{\sum_{i=1}^n w_{ij}^{(0)}}, j=1, 2, \dots, m;$$

$$\Sigma_j^{(1)} = \frac{\sum_{i=1}^n w_{ij}^{(0)} \mathbf{x}_i \mathbf{x}_i'}{\sum_{i=1}^n w_{ij}^{(0)}} - \mu_j^{(1)} \mu_j^{(1)'}, j=1, 2, \dots, m.$$

A $\hat{\boldsymbol{\beta}}$ will be obtained when using the updated values as new starting values and reevaluating the above equations until convergence. At the end of each cycle, one needs to check that no $\Sigma_j^{(1)}$ is close to being singular. This can be done by specifying a small number ε and redoing the above procedure with a new set of starting values whenever $|\Sigma_j^{(1)}| < \varepsilon |\mathbf{S}|$, where $\mathbf{S}$ is the sample covariance matrix of the whole sample. With many different (random) starting values, we will get the MLE by choosing $\hat{\boldsymbol{\beta}}$ that maximizes $l(\hat{\boldsymbol{\beta}})$. At the final estimates, any tiny $\hat{\pi}_j$ may indicate that m is overspecified, as in Table 2(b).

Appendix B

This appendix contains the score vectors $l_i(\boldsymbol{\theta})$ and Hessian matrices $\bar{l}_i(\boldsymbol{\theta})$. These are used to obtain the matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ in section 2. Following the notation of section 2, let $f_j(\mathbf{x}_i)$ be the density function of $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}_j, \Sigma_j)$ and

$$l_{Nij}(\beta_j) = \log f_j(\mathbf{x}_i) = -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_j)' \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) / 2,$$

where the subscript N denotes the likelihood of a normal distribution to distinguish it from the $l_i(\boldsymbol{\beta})$ in equation (2). Let $\mathbf{D}_p$ be the duplication matrix such that $\text{vec}(\Sigma_j) = \mathbf{D}_p \text{vech}(\Sigma_j)$,

$$\mathbf{W}_{cj} = \frac{1}{2} \mathbf{D}_p' (\Sigma_j^{-1} \otimes \Sigma_j^{-1}) \mathbf{D}_p,$$

and

$$\mathbf{V}_j(\mathbf{x}_i) = \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) (\mathbf{x}_i - \boldsymbol{\mu}_j)' \Sigma_j^{-1} - \frac{1}{2} \Sigma_j^{-1}.$$

Then standard differential rules lead to

$$\frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j} = \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j),$$

$$\frac{\partial l_{Nij}(\beta_j)}{\partial \sigma_j} = \mathbf{W}_{cj} \text{vech}[(\mathbf{x}_i - \mu_j)(\mathbf{x}_i - \mu_j)' - \Sigma_j],$$

$$\frac{\partial^2 l_{Nij}(\beta_j)}{\partial \mu_j \partial \mu_j'} = -\Sigma_j^{-1},$$

$$\frac{\partial^2 l_{Nij}(\beta_j)}{\partial \mu_j \partial \sigma_j'} = -\{[(\mathbf{x}_i - \mu_j)' \Sigma_j^{-1}] \otimes \Sigma_j^{-1}\} \mathbf{D}_p,$$

$$\frac{\partial^2 l_{Nij}(\beta_j)}{\partial \sigma_j \partial \mu_j'} = \left[\frac{\partial^2 l_{Nij}(\beta_j)}{\partial \mu_j \partial \sigma_j'} \right]',$$

$$\frac{\partial^2 l_{Nij}(\beta_j)}{\partial \sigma_j \partial \sigma_j'} = -\mathbf{D}_p' [\mathbf{V}_j(\mathbf{x}_i) \otimes \Sigma_j^{-1}] \mathbf{D}_p.$$

Applying the above notation on the differential of $l_i(\boldsymbol{\beta})$ and noticing that

$$\frac{\partial f_j(\mathbf{x}_i)}{\partial \theta} = f_j(\mathbf{x}_i) \frac{\partial \log f_j(\mathbf{x}_i)}{\partial \theta} = f_j(\mathbf{x}_i) \frac{\partial l_{Nij}(\beta_j)}{\partial \theta},$$

we have

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \pi_j} = \frac{f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)}{f(\mathbf{x}_i)}, j=1, 2, \dots, (m-1),$$

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \mu_j} = \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j}, j=1, 2, \dots, m,$$

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \sigma_j} = \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma_j}, j=1, 2, \dots, m.$$

The vector $l_i(\hat{\boldsymbol{\beta}})$ for evaluating the $\hat{\mathbf{B}}$ in section 2 is obtained when putting the above three sets of derivatives into a long vector.

Similarly, we will give the elements of $\hat{l}_i(\boldsymbol{\beta})$ using the second derivatives of $l_i(\boldsymbol{\beta})$ with respect to the mean vector and covariance matrix for each component distribution,

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \pi_k} = - \frac{[f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)][f_k(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)}, j, k=1, 2, \dots, (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \mu_j} = \frac{f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 - \frac{\pi_j [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j}, 1 \leq j \leq (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \sigma_j} = \frac{f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 - \frac{\pi_j [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma_j}, 1 \leq j \leq (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \mu_k} = - \frac{\pi_k f_k(\mathbf{x}_i) [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)} \frac{\partial l_{Nik}(\beta_k)}{\partial \mu_k}, 1 \leq j \neq k \leq (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \sigma_k} = - \frac{\pi_k f_k(\mathbf{x}_i) [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)} \frac{\partial l_{Nik}(\beta_k)}{\partial \sigma_k}, 1 \leq j \neq k \leq (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \mu_m} = - \frac{f_m(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 + \frac{\pi_m [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nim}(\beta_m)}{\partial \mu_m}, 1 \leq j \leq (m-1);$$

$$\frac{\partial^2 l_i(\beta)}{\partial \pi_j \partial \sigma_m} = - \frac{f_m(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 + \frac{\pi_m [f_j(\mathbf{x}_i) - f_m(\mathbf{x}_i)]}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nim}(\beta_m)}{\partial \sigma_m}, 1 \leq j \leq (m-1);$$

$$\begin{aligned} \frac{\partial^2 l_i(\beta)}{\partial \mu_j \partial \mu'_j} &= \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 - \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu'_j} \\ &\quad + \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \frac{\partial^2 l_{Nij}(\beta_j)}{\partial \mu_j \partial \mu'_j}, \quad 1 \leq j \leq m; \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l_i(\beta)}{\partial \mu_j \partial \sigma'_j} &= \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 - \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma'_j} \\ &\quad + \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \frac{\partial^2 l_{Nij}(\beta_j)}{\partial \mu_j \partial \sigma'_j}, \quad 1 \leq j \leq m; \end{aligned}$$

$$\frac{\partial^2 l_i(\beta)}{\partial \mu_j \partial \mu'_k} = - \frac{[\pi_j f_j(\mathbf{x}_i)][\pi_k f_k(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j} \frac{\partial l_{Nik}(\beta_k)}{\partial \mu'_k}, 1 \leq j \neq k \leq m;$$

$$\frac{\partial^2 l_i(\beta)}{\partial \mu_j \partial \sigma'_k} = - \frac{[\pi_j f_j(\mathbf{x}_i)][\pi_k f_k(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)} \frac{\partial l_{Nij}(\beta_j)}{\partial \mu_j} \frac{\partial l_{Nik}(\beta_k)}{\partial \sigma'_k}, 1 \leq j \neq k \leq m;$$

$$\begin{aligned} \frac{\partial^2 l_i(\beta)}{\partial \sigma_j \partial \sigma'_j} &= \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \left\{ 1 - \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \right\} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma_j} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma'_j} \\ &\quad + \frac{\pi_j f_j(\mathbf{x}_i)}{f(\mathbf{x}_i)} \frac{\partial^2 l_{Nij}(\beta_j)}{\partial \sigma_j \partial \sigma'_j}, \quad 1 \leq j \leq m; \end{aligned}$$

$$\frac{\partial^2 l_i(\beta)}{\partial \sigma_j \partial \sigma'_k} = - \frac{[\pi_j f_j(\mathbf{x}_i)][\pi_k f_k(\mathbf{x}_i)]}{f^2(\mathbf{x}_i)} \frac{\partial l_{Nij}(\beta_j)}{\partial \sigma_j} \frac{\partial l_{Nik}(\beta_k)}{\partial \sigma'_k}, 1 \leq j \neq k \leq m.$$

Appendix C

```

/TITLE
Fitting the covariance matrix of component 1 using a three-factor model
/SPECIFICATION
weight='d:\mixture\Gamma_cin11.dat';
cases=400; variables=9; matrix=covariance;
analysis=covariance; methods=ML, robust;
/EQUATION
V1=1F1 +E1;
V2=1*F1+E2;
V3=1*F1+E3;
V4=1F2 +E4;
V5=1*F2+E5;
V6=1*F2+E6;
V7=1F3 +E7;
V8=1*F3+E8;
V9=1*F3+E9;
/VARIANCES
E1-E9=*;
F1=1.0*;
F2=1.0*;
F3=1.0*;
/COVARIANCES
F2,F1=0.5*;
F3,F1=0.5*;
F3,F2=0.5*;
/LMtest
/MATRIX
4.25392 1.60997 1.44642 2.62957 1.33318 1.47782 2.58713 1.01342 0.92652
1.60997 1.89629 0.95962 1.12648 0.41540 0.48102 1.45105 0.66431 0.49418
1.44642 0.95962 1.88683 0.99374 0.31901 0.45803 1.51408 0.57168 0.45506
2.62957 1.12648 0.99374 3.94126 1.33554 1.13264 2.41744 1.43989 1.41107

```

1.33318 0.41540 0.31901 1.33554 1.77134 0.78891 0.79523 0.31037 0.36376
 1.47782 0.48102 0.45803 1.13264 0.78891 1.60733 1.02288 0.33154 0.53856
 2.58713 1.45105 1.51408 2.41744 0.79523 1.02288 3.95897 1.46828 1.29994
 1.01342 0.66431 0.57168 1.43989 0.31037 0.33154 1.46828 1.97394 0.85362
 0.92652 0.49418 0.45506 1.41107 0.36376 0.53856 1.29994 0.85362 2.09065
 /END

REFERENCES

- Anderson, James C.; Gerbing, David W. The Effects of Sampling Error on Convergence, Improper Solutions and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis. *Psychometrika*. 1984; 49:155–173.
- Arminger, Gerhard; Wittenberg, Jörg. Finite Mixtures of Covariance Structure Models with Regressors. *Sociological Methods & Research*. 1997; 26:148–182.
- Arminger, Gerhard; Stein, Petra; Wittenberg, Jörg. Mixtures of Conditional Mean- and Covariance Structure Models. *Psychometrika*. 1999; 64:475–494.
- Bauer, Daniel J. Observations on the Use of Growth Mixture Models in Psychological Research. *Multivariate Behavioral Research*. 2007; 42:757–786.
- Bauer, Daniel J.; Curran, Patrick J. Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological Methods*. 2003; 8:338–363. [PubMed: 14596495]
- Bauer, Daniel J.; Curran, Patrick J. The Integration of Continuous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities. *Psychological Methods*. 2004; 9:3–29. [PubMed: 15053717]
- Bentler, Peter M. EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software; in press
- Bentler, Peter M.; Yuan, Ke-Hai. Structural Equation Modeling with Small Samples: Test Statistics. *Multivariate Behavioral Research*. 1999; 34:181–197.
- Biernacki, Christophe; Celeux, Gilles; Govaert, Gérard. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; 22:719–725.
- Eero, Blåfield. *Jyväskylä Studies in Computer Science, Economics, and Statistics*, 2. Finland: Jyväskylä University; 1980. Clustering of Observations from Finite Mixtures with Structural Information.
- Boomsma, Anne. Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation. *Psychometrika*. 1985; 50:229–242.
- Browne, Michael W. Asymptotic Distribution-Free Methods for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*. 1984; 37:62–83. [PubMed: 6733054]
- Buse, Adolf. The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note. *American Statistician*. 1982; 36:153–157.
- Clogg, Clifford C. Latent Class Models. In: Arminger, Gerhard; Clogg, Clifford C.; Sobel, Michael E., editors. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum; 1995. p. 311–359.
- Dolan, Conor V.; van der Maas, Han L J. Fitting Multivariate Normal Finite Mixtures Subject to Structural Equation Modeling. *Psychometrika*. 1998; 63:227–253.
- Everitt, Brian S.; Hand, David J. *Finite Mixture Distributions*. London: Chapman & Hall; 1981.
- Fouladi, Rachel T. Performance of Modified Test Statistics in Covariance and Correlation Structure Analysis under Conditions of Multivariate Nonnormality. *Structural Equation Modeling*. 2000; 7:356–410.
- Goodman, Leo A. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. *Biometrika*. 1974a; 61:215–231.

- Goodman, Leo A. The Analysis of Systems of Qualitative Variables when Some of the Variables Are Unobservable: Part I-A Modified Latent Structure Approach. *American Journal of Sociology*. 1974b; 79:1179–1259.
- Hagenaars, Jacques A.; McCutcheon, Allan L., editors. *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press; 2002.
- Holzinger, Karl J.; Swineford, Frances. *A Study in Factor Analysis: The Stability of a Bi-Factor Solution*. University of Chicago: Supplementary Educational Monographs; 1939. No. 48
- Hoshino, Takahiro. Bayesian Inference for Finite Mixtures in Confirmatory Factor Analysis. *Behaviormetrika*. 2001; 28:37–63.
- Hosmer, David W. On MLE of the Parameters of a Mixture of Two Normal Distributions when the Sample Size is Small. *Communication in Statistics*. 1973; 1:217–227.
- Hu, Li-tze; Bentler, Peter M. Cutoff Criterion for Fit Indices in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Hu, Li-tze; Bentler, Peter M.; Kano, Yutaka. Can Test Statistics in Covariance Structure Analysis be Trusted? *Psychological Bulletin*. 1992; 112:351–362. [PubMed: 1454899]
- Jedidi, Kamel; Jagpal, Harsharanjeet S.; DeSarbo, Wayne S. Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. *Marketing Science*. 1997; 16:39–59.
- Jöreskog, Karl G. A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*. 1969; 34:183–202.
- Jöreskog, Karl G. Testing Structural Equation Models. Bollen, Kenneth A.; Long, Scott J., editors. Newbury Park, CA: Sage; 1993. p. 294–316.
- Jöreskog, Karl G.; Dag Sörbom, D.; Stephen du Toit; Mathilda du Toit. *LISREL 8: New Statistical Features*. Lincolnwood, IL: Scientific Software International; 2000.
- Lubke, Gitta; Neale, Michael C. Distinguishing between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood? *Multivariate Behavioral Research*. 2006; 41:499–532.
- MacCallum, Robert C.; Roznowski, Mary; Necowitz, Lawrence B. Model Modification in Covariance Structure Analysis: The Problem of Capitalization on Chance. *Psychological Bulletin*. 1992; 111:490–504. [PubMed: 16250105]
- Marcoulides, George A.; Drezner, Zvi; Schumacker, Randall E. Model Specification Searches in Structural Equation Modeling using Tabu Search. *Structural Equation Modeling*. 1998; 5:365–376.
- Mardia, Kanti V. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*. 1970; 57:519–530.
- McLachlan, Geoffrey; Peel, David. *Finite Mixture Models*. New York: Wiley; 2000.
- Muthén, Bengt C.; Brown, Hendricks; Masyn, Katherine; Jo, Booil; Khoo, Siek-Toon; Yang, Chih-Chien; Wang, Chen-Pin; Kellam, Sheppard G.; Carlin, John B.; Liao, Jason. General Growth Mixture Modeling for Randomized Preventive Interventions. *Biostatistics*. 2002; 3:459–475. [PubMed: 12933592]
- Muthén, Bengt; Shedden, Kerby. Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics*. 1999; 55:463–469. [PubMed: 11318201]
- Muthén, Linda K.; Muthén, Bengt O. *Mplus User's Guide*. 5th ed.. Los Angeles, CA: Muthén & Muthén; 2007.
- Satorra, Albert; Bentler, Peter M. Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis. In: von Eye, Alexander; Clogg, Clifford C., editors. *Latent Variables Analysis: Applications for Developmental Research*. Thousand Oaks, CA: Sage; 1994. p. 399–419.
- Sörbom, Dag. Model Modification. *Psychometrika*. 1989; 54:371–384.
- Titterton, David M.; Smith, Adrian FM.; Makov, UE. *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley; 1985.
- Tofighi, Davood; Enders, Craig K. Identifying the Correct Number of Classes in Growth Mixture Models. In: Hancock, Gregory R.; Samuelsen, Karen M., editors. *Advances in Latent Variable Mixture Models*. Charlotte, NC: IAP; 2008. p. 317–341.
- Wu, C F Jeff. On the Convergence Properties of the EM Algorithm. *Annals of Statistics*. 1983; 11:95–103.

- Yuan, Ke-Hai; Bentler, Peter M. Improving Parameter Tests in Covariance Structure Analysis. *Computational Statistics and Data Analysis*. 1997; 26:177–198.
- Yuan, Ke-Hai; Bentler, Peter M. Normal Theory Based Test Statistics in Structural Equation Modeling. *British Journal of Mathematical and Statistical Psychology*. 1998; 51:289–309. [PubMed: 9854947]
- Yuan, Ke-Hai; Bentler, Peter M. Multilevel Covariance Structure Analysis by Fitting Multiple Single-Level Models. *Sociological Methodology*. 2007; 37:53–82.
- Yuan, Ke-Hai; Bentler, Peter M. Two Simple Approximations to the Distributions of Quadratic Forms. *British Journal of Mathematical and Statistical Psychology*. 2009
- Yuan, Ke-Hai; Jennrich, Robert I. Asymptotics of Estimating Equations under Natural Conditions. *Journal of Multivariate Analysis*. 1998; 65:245–260.
- Yung, Yiu-Fai. Finite Mixtures in Confirmatory Factor-Analytic Models. PhD Dissertation: UCLA; 1994.
- Yung, Yiu-Fai. Finite Mixtures in Confirmatory Factor-Analytic Models. *Psychometrika*. 1997; 62:297–330.
- Zhu, Hong-Tu; Lee, Sik-Yum. A Bayesian Analysis of Finite Mixtures in the LISREL Model. *Psychometrika*. 2001; 66:133–152.

Table 1

Fitting statistics at stage-1 for Example 1 (Holzinger and Swineford, 1939), with $m = 1$ and 2 components.

(a) saturated model			
m	$l(\hat{\beta})$	BIC	$\hat{\pi}$
1	-4522.561	9313.866	$\hat{\pi}_1 = 1.000$
2	-4461.783	9466.030	$\hat{\pi}_1 = .816, \hat{\pi}_2 = .184$
(b) three-factor model			
m	$l(\hat{\theta})$	BIC	$\hat{\pi}$
1	-4548.332	9245.967	$\hat{\pi}_1 = 1.000$
2	-4507.782	9259.424	$\hat{\pi}_1 = 0.483, \hat{\pi}_2 = 0.517$
(c) one-factor model			
m	$l(\hat{\theta})$	BIC	$\hat{\pi}$
1	-4613.768	9361.908	$\hat{\pi}_1 = 1.000$
2	-4558.568	9346.066	$\hat{\pi}_1 = 0.673, \hat{\pi}_2 = 0.327$
(d) model modification with two-component one-factor model			
model	$l(\hat{\theta})$	BIC	
M_1	-4558.568	9346.066	
$M_2 (\psi_{45}^{(1)}, \psi_{78}^{(2)})$	-4539.393	9317.669	
$M_3 (\psi_{78}^{(1)}, \psi_{89}^{(2)})$	-4523.860	9296.556	
$M_4 (\psi_{56}^{(1)}, \lambda_6^{(1)} \neq \lambda_6^{(2)})$	-4515.359	9289.509	

Table 2

Fitting statistics at stage-1 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), with $m = 1, 2$ and 3 components.

(a) structural model			
m	$l(\hat{\theta})$	BIC	$\hat{\pi}$
1	-6299.607	12778.958	$\hat{\pi}_1 = 1.000$
2	-6201.834	12769.148	$\hat{\pi}_1 = .356, \hat{\pi}_2 = .644$
3	-6137.616	12826.447	$\hat{\pi}_1 = .268, \hat{\pi}_2 = .475, \hat{\pi}_3 = .257$
(b) saturated model			
m	$l(\hat{\beta})$	BIC	$\hat{\pi}$
1	-6225.866	12775.272	$\hat{\pi}_1 = 1.000$
2	-6045.523	12744.116	$\hat{\pi}_1 = .517, \hat{\pi}_2 = .483$
3	-5966.523	12915.646	$\hat{\pi}_1 = .479, \hat{\pi}_2 = .471, \hat{\pi}_3 = .053$

Table 3

Table 3(a). Test statistics for overall model evaluation of the first component at stage-2 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), $\hat{\Gamma}_{c11} = \hat{\Gamma}_{c111}$, p -values below 0.001 are not reported.

	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	205.164	90.085	73.833(20)	52.780	2.389 (24,376)	68.508 (λ_{12})
p				0.001		
M_2	151.213	60.081	57.463 (20)	44.150	2.040 (23,377)	65.341 (λ_{43})
p				0.005	0.004	
M_3	93.362	40.857	37.366 (20)	30.929	1.444 (22,378)	47.390 (λ_{71})
p		0.009	0.011	0.098	0.090	
M_4	42.347	18.590	18.090 (20)	17.193	0.813 (21,379)	7.714 (λ_{61})
p	0.004	0.611	0.581	0.699	0.704	0.005
M_5	33.980	14.930	14.566 (20)	13.507	0.666 (20,380)	4.657 (λ_{82})
p	0.026	0.780	0.801	0.855	0.860	0.031

Table 3(b). Test statistics for overall model evaluation of the first component at stage-2 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), $\hat{\Gamma}_{c11} = \hat{\Gamma}_{c111}$, p -values below 0.001 are not reported.

	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	205.164	87.740	61.698 (17)	62.109	2.890 (24,376)	68.508 (λ_{12})
p						
M_2	151.213	63.430	49.945 (17)	48.194	2.253 (23,377)	65.341 (λ_{43})
p				0.002	0.001	
M_3	93.362	38.465	29.921 (17)	36.224	1.716 (22,378)	47.390 (λ_{71})
p		0.016	0.027	0.029	0.024	
M_4	42.347	17.321	14.175 (17)	16.979	0.802 (21,379)	7.714 (λ_{61})
p	0.004	0.691	0.655	0.712	0.718	0.005
M_5	33.980	13.689	11.203 (16)	14.442	0.714 (20,380)	4.657 (λ_{82})
p	0.026	0.846	0.797	0.807	0.813	0.031

Table 3(c). Structural parameter estimates by the single-stage and two-stage ML for the first component of Example 2 (a simulated sample from a mixture of two multivariate normal distributions), the models M_2 to M_5 under the two-stage ML follow from the results of the univariate LM test.

θ_{c1}	single-stage ML		2-stage ML				
	M_1	M_1	M_2	M_3	M_4	M_5	
λ_{21}	0.489	0.506	1.342	0.986	0.974	0.824	
λ_{31}	0.410	0.470	1.274	0.929	0.946	0.804	
λ_{52}	0.811	0.463	0.495	1.130	0.998	1.057	
λ_{62}	0.617	0.458	0.491	1.107	0.962	0.765	
λ_{83}	0.356	0.486	0.469	0.508	1.130	1.135	
λ_{93}	0.317	0.446	0.426	0.458	1.039	1.040	
ϕ_{11}	2.471	3.224	0.587	1.072	1.033	1.413	
ϕ_{21}	1.926	2.557	0.776	0.379	0.479	0.447	
ϕ_{31}	1.844	2.600	1.064	1.405	0.510	0.618	
ϕ_{22}	2.017	2.775	2.657	0.639	0.805	0.806	
ϕ_{32}	1.225	2.367	2.217	0.778	0.364	0.306	
ϕ_{33}	2.240	3.042	3.161	2.982	0.753	0.738	
ψ_{11}	1.021	1.030	1.195	0.973	0.916	0.932	
ψ_{22}	0.904	1.070	0.840	0.854	0.917	0.937	
ψ_{33}	1.179	1.176	0.935	0.962	0.961	0.973	
ψ_{44}	1.229	1.166	1.284	1.383	1.051	0.945	
ψ_{55}	0.755	1.177	1.120	0.955	0.969	0.872	
ψ_{66}	0.956	1.026	0.967	0.824	0.863	0.911	
ψ_{77}	0.874	.917	0.798	0.977	1.144	1.133	
ψ_{88}	1.436	1.256	1.278	1.205	1.012	1.022	
ψ_{99}	1.586	1.485	1.518	1.465	1.277	1.292	
λ_{12}			0.716	1.359	1.199	1.081	
λ_{43}				0.583	1.249	1.358	
λ_{71}					1.003	0.854	
λ_{61}						0.224	

Table 4

Table 4(a). Test statistics for overall model evaluation of the second component at stage-2 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), $\hat{\Gamma}_{c22} = \hat{\Gamma}_{c1w22}$, p -values below 0.001 are not reported.

	T_{ML2}	T_{RM2}	T_{MLa2}	T_{CRADF2}	F_{R2}	LM
M_1	529.987	244.070	178.544 (18)	99.290	5.194 (24,376)	50.431 (λ_{63})
p						
M_2	409.758	185.175	127.495 (16)	83.885	4.366 (23,377)	102.959 (λ_{12})
p						
M_3	300.651	137.581	110.341 (18)	73.889	3.907 (22,378)	54.816 (λ_{82})
p						
M_4	248.304	115.962	95.531 (17)	68.655	3.753 (21, 379)	48.848 (λ_{51})
p						
M_1	529.987	244.070	178.544 (18)	99.290	5.194 (24,376)	199.101 (ψ_{86})
p						
M_5	299.251	138.697	114.132 (19)	62.400	3.040 (23,377)	215.712 (ψ_{51})
p						
M_6	44.879	21.004	20.226 (21)	18.200	0.821 (22,378)	12.900 (ψ_{98})
p	0.003	0.520	0.507	0.694	0.699	
M_7	31.712	14.903	14.491 (20)	14.578	0.684 (21,379)	6.677 (ψ_{95})
p	0.063	0.828	0.805	0.844	0.849	0.010

Table 4(b). Test statistics for overall model evaluation of the second component at stage-2 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), $\hat{\Gamma}_{c22} = \hat{\Gamma}_{c1w22}$, p -values below 0.001 are not reported.

	T_{ML2}	T_{RM2}	T_{MLa2}	T_{CRADF2}	F_{R2}	LM
M_1	529.987	217.602	139.175 (15)	105.997	5.673 (24,376)	50.431 (λ_{63})
p						
M_2	409.758	165.943	86.979 (12)	99.612	5.458 (23,377)	102.959 (λ_{12})
p						
M_3	300.651	121.131	76.612 (14)	84.312	4.606 (22,378)	54.816 (λ_{82})
p						

Table 4(b). Test statistics for overall model evaluation of the second component at stage-2 for Example 2 (a simulated sample from a mixture of two multivariate normal distributions), $\hat{\Gamma}_{c22} = \hat{\Gamma}_{c1w22}$, p -values below 0.001 are not reported.

	T_{ML2}	T_{RM2}	T_{MLa2}	T_{CRADF2}	F_{R2}	LM
M_4	248.304	101.004	67.355 (14)	74.164	4.123 (21,379)	48.848 (λ_{51})
p						
M_1	529.987	217.602	139.175 (15)	105.997	5.673 (24,376)	199.101 (ψ_{86})
p						
M_5	299.251	123.979	85.974 (16)	68.418	3.394 (23, 377)	215.712 (ψ_{51})
p						
M_6	44.879	18.498	14.152 (17)	19.557	0.886 (22, 378)	12.900 (ψ_{98})
p	0.003	0.676	0.656	0.611	0.614	
M_7	31.712	13.586	11.034 (17)	16.570	0.782 (21,379)	6.677 (ψ_{95})
p	0.063	0.887	0.855	0.737	0.742	0.010

Table 4(c). Structural parameter estimates by the single-stage and two-stage ML for the second component of Example 2 (a simulated sample from a mixture of two multivariate normal distributions), the models M_2 to M_7 under the two-stage ML follow from the results of the univariate LM test.

	single-stage ML							2-stage ML						
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_1	M_2	M_3	M_4	M_5	M_6	M_7
θ_{c2}														
λ_{21}	1.119	1.075	1.393	1.355	1.013	1.147	1.163							
λ_{31}	1.122	1.074	1.472	1.427	0.996	1.162	1.176							
λ_{52}	1.701	1.255	4.310	5.089	1.375	1.274	1.273							
λ_{62}	1.511	1.137	.714	0.281	0.985	1.062	1.041							
λ_{83}	1.253	0.996	1.715	1.830	0.865	0.892	0.668							
λ_{93}	1.323	1.001	1.023	1.094	1.090	1.106	0.822							
ϕ_{11}	1.335	1.142	0.682	0.726	1.243	1.049	1.022							
ϕ_{21}	0.737	0.594	0.096	0.073	0.657	0.507	0.491							
ϕ_{31}	0.747	0.587	0.244	0.274	0.647	0.553	0.699							
ϕ_{22}	0.711	0.672	0.193	0.146	0.743	0.767	0.757							
ϕ_{32}	0.756	0.615	0.066	0.123	0.501	0.504	0.602							
ϕ_{33}	1.024	1.078	0.608	0.615	1.070	1.075	1.409							
ψ_{11}	1.400	1.228	1.007	1.014	1.127	1.359	1.373							

Table 4(c). Structural parameter estimates by the single-stage and two-stage ML for the second component of Example 2 (a simulated sample from a mixture of two multivariate normal distributions), the models M_2 to M_7 under the two-stage ML follow from the results of the univariate LM test.

θ_{c2}	single-stage ML		2-stage ML						
	M_1		M_1	M_2	M_3	M_4	M_5	M_6	M_7
ψ_{22}	1.146		1.086	1.741	1.079	1.073	1.129	1.025	1.022
ψ_{33}	1.124		1.198	1.859	1.036	1.038	1.281	1.099	1.102
ψ_{44}	1.437		1.288	1.656	1.767	1.813	1.217	1.193	1.202
ψ_{55}	1.313		1.226	-0.432	-1.298	-1.500	0.879	1.120	1.134
ψ_{66}	0.690		0.693	0.676	0.660	0.642	0.901	0.782	.784
ψ_{77}	1.168		1.157	1.741	1.627	1.619	1.164	1.160	.826
ψ_{88}	0.787		0.870	-0.071	0.151	0.180	1.085	1.071	1.282
ψ_{99}	1.140		1.122	1.631	1.565	1.466	0.931	0.888	1.250
λ_{63}				1.116	1.079	1.160			
λ_{12}					1.448	1.638			
λ_{82}						-0.980			
ψ_{86}							0.801	0.783	0.795
ψ_{51}								1.022	1.034
ψ_{98}									0.297

Table 5

Fitting statistics at stage-1 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), saturated model with $m = 1, 2$ and 3 components.

m	$l(\hat{\beta})$	BIC	$\hat{\pi}$
1	-6674.715	13672.969	$\hat{\pi}_1 = 1.000$
2	-6501.935	13656.939	$\hat{\pi}_1 = .524, \hat{\pi}_2 = .476$
3	-6383.058	13748.716	$\hat{\pi}_1 = .174, \hat{\pi}_2 = .394, \hat{\pi}_3 = .432$

Table 6

Table 6(a). Test statistics for overall model evaluation of the first component at stage-2 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), $\Gamma_{e11} = \Gamma_{uin11}$, p -values below 0.001 are not reported.

	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	139.734	67.511	58.317 (21)	45.527	2.018 (24,376)	58.300 (λ_{12})
p				0.005	0.003	
M_2	93.548	45.104	41.191 (21)	37.042	1.678 (23,377)	35.479 (λ_{43})
p		0.004	0.005	0.032	0.027	
M_3	60.282	29.112	27.613 (21)	27.498	1.272 (22,378)	8.840 (λ_{71})
p		0.142	0.151	0.193	0.186	
M_4	47.721	23.018	22.097 (20)	21.615	1.034 (21,379)	9.386 (λ_{82})
p	0.001	0.343	0.335	0.422	0.421	0.002

Table 6(b). Test statistics for overall model evaluation of the first component at stage-2 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), $\Gamma_{e11} = \Gamma_{uin11}$, p -values below 0.001 are not reported.

	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	139.734	49.741	28.799 (14)	44.990	1.992 (24,376)	58.300 (λ_{12})
p		0.002	0.011	0.006	0.004	
M_2	93.548	33.173	20.644 (14)	36.777	1.665 (23,377)	35.479 (λ_{43})
p		0.078	0.111	0.034	0.029	
M_3	60.282	20.419	12.772 (14)	25.893	1.193 (22,378)	8.840 (λ_{71})
p		0.557	0.545	0.256	0.250	
M_4	47.721	15.649	9.608 (13)	17.249	0.816 (21,379)	9.386 (λ_{82})
p	0.001	0.789	0.726	0.696	0.701	0.002

Table 6(c). Structural parameter estimates by the two-stage ML for the first component of Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), the models M_2 to M_4 follow from the results of the univariate LM test.

θ_{c1}	M_1	M_2	M_3	M_4
λ_{21}	0.517	1.426	1.039	1.050
λ_{31}	0.515	1.436	1.059	1.064

Table 6(c). Structural parameter estimates by the two-stage ML for the first component of Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), the models M_2 to M_4 follow from the results of the univariate LM test.

θ	M_1	M_2	M_3	M_4
λ_{52}	0.498	0.520	1.181	1.068
λ_{62}	0.455	0.470	1.042	0.935
λ_{83}	0.518	0.511	.540	1.108
λ_{93}	0.557	0.547	.571	1.156
ϕ_{11}	4.473	0.729	1.351	1.304
ϕ_{21}	3.573	1.122	0.568	0.673
ϕ_{31}	3.233	1.259	1.687	0.756
ϕ_{22}	3.644	3.600	0.863	1.029
ϕ_{32}	2.930	2.811	3.083	0.553
ϕ_{33}	3.180	3.231	1.066	0.860
ψ_{11}	1.332	1.578	1.344	1.281
ψ_{22}	1.387	1.101	1.124	1.147
ψ_{33}	1.487	1.172	1.159	1.199
ψ_{44}	1.353	1.397	1.580	1.347
ψ_{55}	1.815	1.745	1.513	1.543
ψ_{66}	1.344	1.303	1.160	1.197
ψ_{77}	1.713	1.662	1.810	2.006
ψ_{88}	1.906	1.915	1.860	1.703
ψ_{99}	2.204	2.222	2.185	2.039
λ_{12}		0.722	1.351	1.232
λ_{43}			0.628	1.217
λ_{71}				0.795

Table 7

Table 7(a). Test statistics for overall model evaluation of the second component at stage-2 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), $\Gamma_{c22} = \Gamma_{c22s}$, p -values below 0.001 are not reported.						
	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	590.086	237.789	151.992 (15)	94.356	4.856 (24,376)	118.992 (λ_{51})
p						
M_2	416.523	167.754	114.830 (16)	78.919	4.044 (23,377)	92.486 (λ_{82})
p						
M_3	311.259	126.846	96.711 (17)	65.531	3.378 (22,378)	36.537 (λ_{12})
p						
M_4	277.596	116.447	92.005 (17)	65.258	3.531 (22,378)	32.904 (λ_{63})
p						
M_1	590.086	237.789	151.992 (15)	94.356	4.856 (24,376)	328.927 (ψ_{51})
p						
M_5	324.094	132.261	106.502 (19)	61.017	2.961 (23,377)	220.416 (ψ_{86})
p						
M_6	67.981	28.710	27.492 (21)	25.061	1.152 (22,378)	18.708 (ψ_{42})
p						
		0.153	0.155	0.294	0.289	
M_7	49.247	20.737	20.082 (20)	18.254	0.865 (21,379)	13.577 (ψ_{95})
p						
		0.475	0.453	0.633	0.637	
M_8	35.747	15.109	14.747 (20)	13.799	0.681 (20,3380)	6.610 (ψ_{98})
p						
	0.016	0.770	0.791	0.841	0.846	0.010

Table 7(b). Test statistics for overall model evaluation of the second component at stage-2 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), $\Gamma_{c22} = \Gamma_{c22s}$, p -values below 0.001 are not reported.						
	T_{ML1}	T_{RML1}	T_{MLd1}	T_{CRADF1}	F_{R1}	LM
M_1	590.086	174.560	93.277 (13)	84.495	4.212 (24,376)	118.992 (λ_{51})
p						
M_2	416.523	122.576	67.067 (13)	75.632	3.836 (23,377)	92.486 (λ_{82})
p						

Table 7(b). Test statistics for overall model evaluation of the second component at stage-2 for Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), $\Gamma_{c22} = \Gamma_{c22|22}$, p -values below 0.001 are not reported.

	T_{ML1}	T_{RML1}	$T_{ML\alpha1}$	T_{CRADF1}	F_{R1}	LM
M_3	311.259	94.870	56.546 (13)	65.884	3.400 (22,378)	36.537 (χ^2_{12})
p						
M_4	277.596	85.447	51.877 (13)	67.752	3.693 (21,379)	32.904 (χ^2_{63})
p						
M_1	590.086	174.560	93.277 (13)	84.495	4.212 (24,376)	328.927 (ψ_{51})
p						
M_5	324.094	92.484	52.516 (13)	56.300	2.694 (23,377)	220.416 (ψ_{86})
p						
M_6	67.981	20.135	12.292 (13)	25.247	1.161 (22,378)	18.708 (ψ_{42})
p		0.575	0.504	0.285	0.280	
M_7	49.247	14.806	9.954 (14)	17.095	0.808 (21,379)	13.577 (ψ_{95})
p		0.833	0.766	0.705	0.710	

Table 7(c). Structural parameter estimates by the two-stage ML for the second component of Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), the models M_2 to M_8 follow from the results of the univariate LM test.

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
θ_{c2}								
λ_{21}	0.507	.398	0.473	.462	1.136	1.185	1.205	1.223
λ_{31}	0.488	.381	0.451	.483	1.157	1.157	1.246	1.257
λ_{52}	2.480	.847	0.838	.342	1.101	1.117	1.114	1.060
λ_{62}	0.994	2.655	4.426	4.992	1.131	.976	1.056	1.040
λ_{83}	1.056	1.885	0.989	.987	1.166	1.078	1.096	1.048
λ_{93}	1.192	1.191	1.265	1.256	1.122	1.260	1.259	1.178
ϕ_{11}	2.618	3.199	2.780	2.970	1.066	1.070	0.978	.945
ϕ_{21}	0.794	0.176	0.110	0.240	0.549	0.568	0.505	.521
ϕ_{31}	0.742	0.389	0.679	0.789	0.670	0.583	0.560	.607
ϕ_{22}	0.434	0.346	0.186	0.147	0.933	1.081	0.994	1.009
ϕ_{32}	0.317	0.295	0.105	0.087	0.704	0.599	0.577	0.592
ϕ_{33}	1.350	0.753	1.262	1.276	1.303	1.269	1.274	1.350

Table 7(c). Structural parameter estimates by the two-stage ML for the second component of Example 3 (a simulated sample from a mixture of two multivariate t -distributions, each with 8 degrees of freedom), the models M_2 to M_8 follow from the results of the univariate LM test.

θ_{c2}	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
ψ_{11}	0.251	-.329	0.089	0.227	1.755	1.765	1.785	1.797
ψ_{22}	2.169	2.336	2.221	2.208	1.469	1.341	1.443	1.463
ψ_{33}	2.520	2.679	2.578	2.452	1.718	1.713	1.626	1.652
ψ_{44}	2.229	2.317	2.477	2.517	1.730	1.582	1.674	1.661
ψ_{55}	-0.005	1.179	1.115	1.041	1.601	1.407	1.467	1.561
ψ_{66}	1.393	-.619	-1.832	-1.832	0.628	0.891	0.817	.759
ψ_{77}	1.221	1.817	1.308	1.294	1.267	1.302	1.297	1.221
ψ_{88}	1.025	-.146	0.807	.805	0.758	1.055	1.043	1.010
ψ_{99}	0.915	1.765	0.812	.820	1.193	0.818	0.814	.953
λ_{51}		0.577	0.682	0.708				
λ_{82}			1.158	1.317				
λ_{12}				-0.972				
ψ_{51}					1.400	1.396	1.393	1.400
ψ_{86}						0.850	0.835	0.773
ψ_{42}							0.414	0.416
ψ_{95}								0.212

Efficiency and accuracy of parameter estimates, $\boldsymbol{\mu}_1 = (0, \dots, 0)'$ and $\boldsymbol{\mu}_2 = (3, \dots, 3)'$, three-factor model, unidimensional measurement in the population, the model is correctly specified in number of components, number of population and covariance structures.

Table 8

(a) Component 1		2-stage ML			single-stage ML		
		Mean	Bias	SE	Mean	Bias	SE
θ_{c1}		1.014	0.014	0.174	1.009	0.009	0.160
λ_{21}		1.028	0.028	0.170	1.025	0.025	0.161
λ_{31}		1.041	0.041	0.180	1.030	0.030	0.171
λ_{42}		1.033	0.033	0.190	1.022	0.022	0.158
λ_{43}		1.025	0.025	0.179	1.024	0.024	0.169
λ_{43}		1.027	0.027	0.179	1.027	0.027	0.172
ϕ_{11}		1.037	0.037	0.300	1.020	0.020	0.268
ϕ_{21}		0.525	0.025	0.200	0.509	0.009	0.168
ϕ_{31}		0.535	0.035	0.208	0.513	0.013	0.175
ϕ_{22}		1.013	0.013	0.299	1.007	0.007	0.268
ϕ_{32}		0.523	0.023	0.205	0.506	0.006	0.173
ϕ_{33}		1.037	0.037	0.297	1.010	0.010	0.260
ψ_{11}		0.988	-0.012	0.167	0.986	-0.014	0.165
ψ_{22}		0.988	-0.012	0.154	0.991	-0.009	0.151
ψ_{33}		0.979	-0.021	0.152	0.982	-0.018	0.149
ψ_{44}		0.994	-0.006	0.156	0.990	-0.010	0.154
ψ_{55}		0.974	-0.026	0.163	0.977	-0.023	0.156
ψ_{66}		0.984	-0.016	0.168	0.987	-0.013	0.161
ψ_{77}		0.995	-0.005	0.165	0.995	-0.005	0.163
ψ_{88}		0.970	-0.030	0.176	0.973	-0.027	0.169
ψ_{99}		0.976	-0.024	0.157	0.976	-0.024	0.154
Average		bias	Var	MSE	bias	Var	MSE
		0.023	0.039	0.040	0.016	0.033	0.033

Table 8(b) Component 2

	2-stage ML			single-stage ML		
	Mean	Bias	SE	Mean	Bias	SE
θ_{22}	1.003	0.003	0.172	1.002	0.002	0.160
λ_{21}	1.016	0.016	0.185	1.014	0.014	0.170
λ_{31}	1.024	0.024	0.174	1.015	0.015	0.162
λ_{52}	1.023	0.023	0.175	1.018	0.018	0.170
λ_{83}	1.023	0.023	0.178	1.022	0.022	0.166
λ_{93}	1.019	0.019	0.176	1.021	0.021	0.171
ϕ_{11}	1.041	0.041	0.301	1.029	0.029	0.269
ϕ_{21}	0.512	0.012	0.201	0.504	0.004	0.172
ϕ_{31}	0.518	0.018	0.210	0.504	0.004	0.177
ϕ_{22}	1.009	0.009	0.282	1.009	0.009	0.259
ϕ_{32}	0.507	0.007	0.195	0.496	-0.004	0.169
ϕ_{33}	1.012	0.012	0.281	0.996	-0.004	0.253
ψ_{11}	0.976	-0.024	0.165	0.979	-0.021	0.161
ψ_{22}	1.003	0.003	0.161	1.004	0.004	0.159
ψ_{33}	0.981	-0.019	0.182	0.982	-0.018	0.170
ψ_{44}	0.985	-0.015	0.154	0.985	-0.015	0.153
ψ_{55}	0.975	-0.025	0.155	0.980	-0.020	0.149
ψ_{66}	0.977	-0.023	0.171	0.977	-0.023	0.164
ψ_{77}	0.981	-0.019	0.165	0.983	-0.017	0.159
ψ_{88}	0.976	-0.024	0.171	0.982	-0.018	0.163
ψ_{99}	0.980	-0.020	0.162	0.981	-0.019	0.160
Average	bias	Var	MSE	bias	Var	MSE
	0.018	0.038	0.039	0.014	0.033	0.033

Table 9

Efficiency and accuracy of parameter estimates, $\mu_1 = (0, \dots, 0)'$ and $\mu_2 = (1.5, \dots, 1.5)'$, three-factor model, unidimensional measurement in the population, the model is correctly specified in number of components, number of population and covariance structures.

(a) Component 1 (based on 491 converged replications)									
	2-stage ML			single-stage ML					
	Mean	Bias	SE	Mean	Bias	SE			
θ_{c1}	1.069	0.069	0.746	1.040	0.040	0.447			
λ_{21}	1.076	0.076	0.800	1.089	0.089	0.587			
λ_{31}	1.069	0.069	0.556	1.074	0.074	0.437			
λ_{52}	1.065	0.065	0.644	1.081	0.081	0.766			
λ_{62}	1.041	0.041	0.363	1.042	0.042	0.409			
λ_{83}	1.040	0.040	0.346	1.067	0.067	0.486			
λ_{93}	1.316	0.316	0.565	1.217	0.217	0.577			
ϕ_{11}	0.784	0.284	0.336	0.699	0.199	0.400			
ϕ_{21}	0.784	0.284	0.332	0.713	0.213	0.411			
ϕ_{31}	1.319	0.319	0.513	1.228	0.228	0.545			
ϕ_{22}	0.796	0.296	0.344	0.695	0.195	0.380			
ϕ_{32}	1.331	0.331	0.515	1.248	0.248	0.531			
ϕ_{33}	0.931	-0.069	0.473	0.924	-0.076	0.345			
ψ_{11}	0.936	-0.064	0.412	0.895	-0.105	0.331			
ψ_{22}	0.967	-0.033	0.334	0.901	-0.099	0.348			
ψ_{33}	0.944	-0.056	0.346	0.932	-0.068	0.338			
ψ_{44}	0.956	-0.044	0.354	0.916	-0.084	0.350			
ψ_{55}	0.939	-0.061	0.382	0.907	-0.093	0.323			
ψ_{66}	0.958	-0.042	0.367	0.957	-0.043	0.381			
ψ_{77}	0.918	-0.082	0.406	0.902	-0.098	0.341			
ψ_{88}	0.935	-0.065	0.413	0.892	-0.108	0.363			
Average	bias	Var	MSE	bias	Var	MSE			
	0.129	0.225	0.254	0.118	0.200	0.218			

Table 9(b) Component 2 (based on 492 converged replications)

	2-stage ML			single-stage ML		
	Mean	Bias	SE	Mean	Bias	SE
θ_{c1}						
λ_{21}	1.024	0.024	0.279	1.014	0.014	0.453
λ_{31}	1.056	0.056	0.361	1.031	0.031	0.405
λ_{52}	1.045	0.045	0.406	1.020	0.020	0.334
λ_{62}	1.049	0.049	0.406	1.058	0.058	0.498
λ_{83}	1.044	0.044	0.480	1.040	0.040	0.585
λ_{93}	1.052	0.052	0.396	1.108	0.108	1.633
ϕ_{11}	1.333	0.333	0.568	1.274	0.274	0.535
ϕ_{21}	0.800	0.300	0.344	0.732	0.232	0.371
ϕ_{31}	0.802	0.302	0.361	0.737	0.237	0.386
ϕ_{22}	1.323	0.323	0.508	1.278	0.278	0.501
ϕ_{32}	0.786	0.286	0.350	0.735	0.235	0.374
ϕ_{33}	1.319	0.319	0.578	1.248	0.248	0.518
ψ_{11}	0.946	-0.054	0.417	0.884	-0.116	0.322
ψ_{22}	0.967	-0.033	0.317	0.927	-0.073	0.304
ψ_{33}	0.933	-0.067	0.397	0.901	-0.099	0.306
ψ_{44}	0.968	-0.032	0.380	0.915	-0.085	0.303
ψ_{55}	0.937	-0.063	0.363	0.901	-0.099	0.290
ψ_{66}	0.931	-0.069	0.376	0.926	-0.074	0.307
ψ_{77}	0.934	-0.066	0.430	0.915	-0.085	0.322
ψ_{88}	0.951	-0.049	0.378	0.923	-0.077	0.316
ψ_{99}	0.948	-0.052	0.344	0.921	-0.079	0.326
Average	bias	Var	MSE	bias	Var	MSE
	0.125	0.167	0.196	0.122	0.278	0.300

Table 10

Model selections using BIC: Saturated model vs. structured model.

distribution condition	<i>n</i> = 500		<i>n</i> = 1000	
	saturated model	structured model	saturated model	structured model
$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	496(4)	499(1)	491(9)	496(4)
$\mathbf{x} \sim Mt(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 8)$	340	6	0	0
$\mathbf{x} \sim Mt(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 5)$	0	0(1)	0	0(2)
$\mathbf{f} \sim \chi^2_3(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Psi})$	500	463	487	27
$\mathbf{f} \sim N(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim \chi^2_3(\mathbf{0}, \boldsymbol{\Psi})$	376	0	1	0
$\mathbf{f} \sim \chi^2_3(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim \chi^2_3(\mathbf{0}, \boldsymbol{\Psi})$	83	0	0	0
$\mathbf{f} \sim \chi^2_1(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Psi})$	230	1	0	0
$\mathbf{f} \sim N(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim \chi^2_1(\mathbf{0}, \boldsymbol{\Psi})$	0	0	0	0
$\mathbf{f} \sim \chi^2_1(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim \chi^2_1(\mathbf{0}, \boldsymbol{\Psi})$	0	0	0	0
$\mathbf{f} \sim \log N(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Psi})$	30(4)	0(2)	0(1)	0(2)
$\mathbf{f} \sim N(\mathbf{0}, \boldsymbol{\Phi}), \mathbf{e} \sim \log N(\mathbf{0}, \boldsymbol{\Psi})$	0	0(1)	0	0(14)