UCLA
Department of Statistics Papers

Title

A Unified Approach to Multi-group Structural Equation Modeling with Nonstandard Samples

Permalink

https://escholarship.org/uc/item/0q8492c0

Authors

Yuan, Ke-Hai
Bentler, Peter M

Publication Date

2000-02-21

A Unified Approach to Multi-group Structural Equation Modeling with Nonstandard Samples[*]

Ke-Hai Yuan
University of North Texas
and
Peter M. Bentler
University of California, Los Angeles

February 21, 2000

1. Introduction

It is well known that structural equation modeling (SEM) has become one of the most popular methods in multivariate analysis, especially in the social and behavioral sciences. In a SEM model with latent variables, the relationships among observed (manifest) variables is formulated through unobserved (latent) constructs. Because measurement errors are explicitly accounted for, coefficients in key parts of a model are uninfluenced by errors of measurement, implying greater theoretical meaningfulness and cross-population stability to the parameters than might be achieved with methods such as regression or analysis of variance that do not correct for unreliability. This stability is a key goal of theory testing with SEM, where a substantive theory or hypothesized causal relationship among the latent constructs, facilitated by path diagrams, can be tested through SEM. With the help of popular software such as LISREL (Jöreskog & Sörbom, 1993) and EQS (Bentler, 2000), applications as well as new technical developments in SEM have increased dramatically in the past decade (e.g., Bollen, 1989; Austin & Calderón, 1996; Austin & Wolfle, 1991; Tremblay & Gardner, 1996). There exists a vast amount of recent introductory (Byrne, 1994; Dunn, Everitt & Pickles, 1993; Kline, 1998; Mueller, 1996; Schumacker & Lomax, 1996) and overview material (Bentler & Dudgeon, 1996; Browne & Arminger, 1995; Hoyle, 1995; Marcoulides & Schumacker, 1996).

A commonly encountered situation is the existence of several samples. These may arise from one or several populations. If the samples are all from one population, their data can be combined for improved inference. On the other hand, if the samples are from several populations, it is important to understand how the populations might differ. For example, it might be interesting to know whether the factor structure of an established instrument, developed for a specific population, is also valid for other populations. In the context of SEM, it is natural to ask whether particular parameters, such as factor loadings, regression coefficients, or variances of factors may be the same or different in various groups such as different ethnic, gender, or age groups. Motivated by such practical problems, Jöreskog (1971) developed a maximum likelihood approach to SEM with multiple groups. Sörbom (1974) studied differences in factor means across groups. Because practical data may not

be normal, Bentler, Lee and Weng (1987) and Muthén (1989) proposed a generalized least squares approach to parameter estimation and model test for multi-group structural models. Recently, Satorra (2000) and Satorra and Bentler (1999) proposed scaled tests in multi-sample analysis of moment structures. The above literature offers important guidance for multi-group modeling in practice. For example, standard software enables users to easily specify simultaneous estimation and evaluation of multi-group models.

With real data obtained under typical testing situations, nonstandard samples that contain missing data, nonnormal data and data with outliers are almost inevitable. As noted above, the literature on multi-group models of sample means and covariance matrices is based on either normal theory maximum likelihood or through generalized least squares. With nonstandard samples, however, there exist various limitations to the current methodologies for using sample moments for multi-group analysis. For example, the typical sample mean vector and sample covariance matrix are not defined when a sample contains missing data. For a complete sample with outliers, the sample mean and covariance matrix will be biased estimates of their population counterparts. Even for a sample from a distribution with heavy tails, the sample moments may not converge at all or at least may not be efficient estimates of the corresponding population moments. These various drawbacks of the sample moments will pass on to an analytical procedure that models these moments. Certain problems with nonstandard samples for single group analysis have been studied and discussed extensively by various authors. Allison (1987), Lee (1986), Muthén, Kaplan and Hollis (1987), Arbuckle (1996), and Jamshidian and Bentler (1999), for example, discussed approaches to dealing with normal missing data. Arminger and Sobel (1990), and Yuan and Bentler (1999), developed approaches for dealing with nonnormal missing data. Techniques for identifying outliers or influential cases can be found in Tanaka, Watadani and Moon (1991), Cadigan (1995), Lee and Wang (1996), Bollen and Arminger (1991), and Berkane and Bentler (1988). Approaches to robust inference for SEM can be found in Yuan and Bentler (1998a,b, 2000). As compared to classical methods which are based on sample means and covariance matrices, these new developments offer various advantages in model estimation and evaluation. It is the aim of this chapter to develop parallel methods for

multi-group analysis with nonstandard samples.

There are various ways to develop multi-group methods for nonstandard samples. Our purpose is to give a unified treatment for multiple groups, aiming to adopt the various developments in the statistical literature in estimating population means and covariance matrices. Suppose we have m groups, and denote the mean vectors and covariance matrices in the population as μ_j and Σ_j, $j = 1, \cdots, m$. Various methods have been developed for estimating μ_j and Σ_j with a nonstandard sample from the jth population. For example, the EM-algorithm based on a normality assumption can be used to estimate μ_j and Σ_j for a normal sample with missing variables. There also exists an EM-algorithm based on a multivariate t-distribution that applies when a missing data sample possesses heavier tails as compared to the normal distribution. And when a sample contains outliers or influential cases, there exist various robust methods for estimating μ_j and Σ_j. Our development will be based on these new advances in estimating the population mean vectors μ_j and covariance matrices Σ_j.

Let $\bar{X}_{nj}$ and S_{nj} be working estimates for μ_j and Σ_j, based on sample size n_j, for $j = 1$, $\cdots$, m. While it is anticipated that $\bar{X}_{nj}$ and S_{nj} might be better estimates than the sample mean vector $\bar{X}_j$ and covariance matrix S_j, we do not exclude the possibility of $\bar{X}_{nj} = \bar{X}_j$ and $S_{nj} = S_j$ in the case of normal sampling with no missing data. Actually, we may just regard $\bar{X}_{nj}$ as a data vector and S_{nj} as a symmetric data matrix which approach μ_j and Σ_j, respectively, as our information about the jth group increases. It is typical that the μ_j in all the groups are of the same dimension, but here we do not need to assume this. Instead, we denote the dimension of μ_j as p_j. Let $\mathrm{vech}(\cdot)$ be an operator which transforms a symmetric matrix into a vector by stacking the columns of the matrix leaving out the elements above the diagonal, $s_{nj} = \mathrm{vech}(S_{nj})$ and $\sigma_j = \mathrm{vech}(\Sigma_j)$. We will use $t_{nj} = (\bar{X}'_{nj}, s'_{nj})'$ and $\delta_j = (\mu'_j, \sigma'_j)'$. We need to assume that each of our data vectors has an appropriate large sample property

$$\sqrt{n_j}(t_{nj} - \delta_j) \xrightarrow{\mathcal{L}} N(0, \Gamma_j), \ j = 1, \cdots, m, \tag{1}$$

where Γ_j is a $p_j^* \times p_j^*$ matrix with $p_j^* = p_j + p_j(p_j + 1)/2$. When $t_{nj} = (\bar{X}'_j, s'_j)'$, the sample moments based on a sample from a normal distribution, then

$$\Gamma_j = \mathrm{diag}[\Sigma_j, 2D_{p_j}^+(\Sigma_j \otimes \Sigma_j)D_{p_j}^{+'}],$$

3

where $D_{p_j}^+$ is the Moore-Penrose generalized inverse of the duplication matrix D_{p_j} (p. 49 of Magnus & Neudecker, 1988). In such a case, a consistent $\hat{\Gamma}_j$ is easily obtained by replacing Σ_j by S_j. However, we need to obtain a better estimator for Γ_j when dealing with a general nonstandard sample. As we will see in the next section, our proposed inference procedure just depends on (1), and we do not need to have the raw data once $t_{nj} = \hat{\delta}_j$ and a consistent $\hat{\Gamma}_j$ are available. Procedures for obtaining t_{nj} and $\hat{\Gamma}_j$ will be given in section 3, based on our experience with current estimation methodologies in the statistical literature for nonstandard samples.

Suppose we are interested in the mean and covariance structures $\delta_j(\beta_j) = (\mu_j'(\beta), \sigma_j'(\beta_j))'$ for $j = 1, \cdots, m$. There are a variety of ways of using the information in (1) in order to estimate parameter $\theta = (\beta_1', \cdots, \beta_m')'$ and evaluate the structures $\delta_j(\beta_j)$. All involve minimizing some function of the distance between t_{nj} and $\delta_j(\beta_j)$. We choose the distance based on the normal theory likelihood function for the following reasons: (a) When data are normal the estimator based on such a function is most efficient. (b) For data with influential cases or outliers, the robust mean vector and covariance matrix can be regarded as the sample mean vector and sample covariance matrix based on an approximately normal sample (Yuan, Chan & Bentler, 2000). (c) The estimation process of minimizing the maximum likelihood function is quite stable, which is very important when modeling several groups simultaneously.

With $N = n_1 + \cdots + n_m$, the maximum likelihood discrepancy function between t_{nj} and $\delta_j(\beta_j)$ is given by

$$F(\theta) = \frac{1}{N} \sum_{j=1}^m n_j F_j(\beta_j), \qquad (2a)$$

where

$$F_j(\beta_j) = (\bar{X}_{nj} - \mu_j(\beta_j))' \Sigma_j^{-1}(\beta_j)(\bar{X}_{nj} - \mu_j(\beta_j)) + \text{tr}[S_{nj}\Sigma_j^{-1}(\beta_j)] - \log|S_{nj}\Sigma_j^{-1}(\beta_j)| - p_j. \quad (2b)$$

The analysis of multiple groups is interesting only when we put constraints on the separate β_js. In the most restricted case, when it is assumed that all samples come from the same population, parameters from each group may be constrained equal across groups. In a less restricted setting, only certain parameters such as factor means and loadings, or latent

4

variable regression coefficients, may be constrained equal. Let the constraints be represented by a $r \times 1$ vector function

$$h(\theta) = 0. \tag{3}$$

Estimation of θ involves minimizing (2) under constraint (3). We denote such an estimator as $\hat{\theta}$. The classical likelihood ratio test statistic is widely known to be of the form $T_{ML} = NF(\hat{\theta})$. Let $p^* = p_1^* + \cdots + p_m^*$, and q be the number of unknown parameters in θ. We need also to assume $\frac{n_j}{N} \to \gamma_j > 0$ in order to study the statistical properties of $\hat{\theta}$. When $t_{nj} = (\bar{X}_j', s_j')'$ are based on samples from normal distributions, both $\delta_j = \delta_j(\beta_j)$ and constraint (3) hold in the populations, then

$$T_{ML} \xrightarrow{\mathcal{L}} \chi^2_{p^*-q+r}. \tag{4}$$

When data vectors t_{nj} are used in (2), (4) will not hold in general. There also exists a likelihood ratio statistic for testing constraint (3). Let $\hat{\theta}^*$ be the estimate of θ without constraint (3). This $\hat{\theta}^*$ is just a collection of the $\hat{\beta}_j^*$ obtained by minimizing the function $F_j(\beta_j)$ in (2b). The commonly used likelihood ratio statistic in testing constraint (3) is

$$T_{ML}^{(h)} = N[F(\hat{\theta}) - F(\hat{\theta}^*)].$$

which is also commonly referred to as the chi-square difference test. When all the samples follow multivariate normal distributions and $t_{nj} = (\bar{X}_j', s_j')'$, then

$$T_{ML}^{(h)} \xrightarrow{\mathcal{L}} \chi^2_r$$

under the null hypothesis of correct model structures and correct constraint. With a non-standard sample, however, the behavior of $T_{ML}^{(h)}$ will not asymptotically follow a chi-square distribution even when the null hypothesis is correct.

Similarly, when data are normal and $t_{nj} = (\bar{X}_j', s_j')'$, it is easy to obtain standard error estimates for $\hat{\theta}$ based on

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega).$$

The covariance matrix Ω is the inverse of the information or Hessian matrix associated with minimizing (2). For nonstandard samples, however, this matrix is inappropriate for

obtaining standard errors. We need to find another Ω to replace the one based on inverting the information matrix.

The major purpose of this chapter is to give a unified treatment of multi-sample structural equation modeling based on minimizing (2) under constraint (3). The most important results are appropriate standard errors for $\hat{\theta}$ and test statistics for evaluating the overall model structure and the constraint. These inferential procedures will be developed in section 2. Section 3 gives brief guidelines for obtaining $t_{nj} = \hat{\delta}_j$ and $\hat{\Gamma}_j$ for several nonstandard samples. Some concluding remarks and discussions will be offered at the end of the chapter.

2. Model Inference

Under the null hypothesis of correct model structures about $\delta_j(\beta_j)$ and correct constraint $h(\theta) = 0$, we will first study the distribution of $\hat{\theta}$ before studying the properties of T_{ML} and $T_{ML}^{(h)}$. Rescaled statistics T_{RML} and $T_{RML}^{(h)}$ then follow from our study of T_{ML} and $T_{ML}^{(h)}$. Since standard ML theory cannot be applied without the normality assumption for observed data, to obtain the properties of $\hat{\theta}$ we will use a generalized estimating equation approach instead (e.g., Liang & Zeger, 1986; Yuan & Jennrich, 1998). We will use dot on top of a function to imply derivative (e.g., $\dot{h}(\theta) = \partial h(\theta)/\partial \theta'$, $\dot{F}(\theta) = \partial F(\theta)/\partial \theta$). We may omit the argument of a function if evaluated at the population value (e.g., $\delta = \delta(\theta_0)$).

In order to obtain $\hat{\theta}$ one generally has to work with the Lagrangian function

$$L(\theta) = F(\theta) + h'(\theta)\lambda,$$

where λ is a $r \times 1$ vector of Lagrangian multipliers (e.g., Aitchison & Silvey, 1958; Bentler & Dijkstra, 1985). Since $\hat{\theta}$ minimizes $L(\theta)$, it satisfies the generalized estimating equation

$$G(\hat{\theta}, \hat{\lambda}) = 0, \tag{5}$$

where

$$G(\theta, \lambda) = \begin{pmatrix} \dot{F}(\theta) + \dot{h}'(\theta)\lambda \\ h(\theta) \end{pmatrix}.$$

Notice that $G(\theta, \lambda)$ is just the derivative of L with respect to $(\theta', \lambda')'$. Since $\lambda_0 = 0$,

$$G(\theta_0, \lambda_0) = \begin{pmatrix} \dot{F}(\theta_0) \\ 0 \end{pmatrix}.$$

Using a first order Taylor expansion on (4) at (θ_0, λ_0), or equivalently, using the estimating equation approach as in Yuan and Jennrich (1998), we obtain

$$\sqrt{N}\left(\begin{array}{c} \hat{\theta} - \theta_0 \\ \hat{\lambda} - \lambda_0 \end{array}\right) = -\dot{G}^{-1}(\theta_0, \lambda_0)\sqrt{N}G(\theta_0, \lambda_0) + o_p(1), \tag{6}$$

where

$$\dot{G}(\theta_0, \lambda_0) = \left(\begin{array}{cc} \ddot{F}(\theta_0) & \dot{h}'(\theta_0) \\ \dot{h}(\theta_0) & 0 \end{array}\right).$$

Denote

$$\dot{G}^{-1}(\theta_0, \lambda_0) = \left(\begin{array}{cc} A^{11} & A^{12} \\ A^{21} & A^{22} \end{array}\right),$$

then it follows from (6) that

$$\sqrt{N}(\hat{\theta} - \theta_0) = -A^{11}\sqrt{N}\dot{F}(\theta_0) + o_p(1). \tag{7}$$

Let $W_j = \mathrm{diag}[\Sigma_j^{-1}, \frac{1}{2}D'_{p_j}(\Sigma_j^{-1} \otimes \Sigma_j^{-1})D_{p_j}]$ and $e_j = t_{nj} - \delta_j$, then with (2b) we have

$$\dot{F}_j(\beta_{0j}) = -2\dot{\delta}'_j W_j e_j + O_p(1/n_j). \tag{8}$$

It follows from (8) that

$$\sqrt{n_j}\dot{F}_j(\beta_{0j}) \xrightarrow{\mathcal{L}} N(0, \Pi_j),$$

where $\Pi_j = 4\dot{\delta}'_j W_j \Gamma_j W_j \dot{\delta}_j$. Since $\dot{F}(\theta_0) = (n_1\dot{F}'_1(\beta_{01})/N, \cdots, n_m\dot{F}'_m(\beta_{0m})/N)'$ and the various $\dot{F}_j(\beta_{0j})$ are independent,

$$\sqrt{N}\dot{F}(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Pi_\gamma), \tag{9}$$

where $\Pi_\gamma = \mathrm{diag}(\gamma_1\Pi_1, \cdots, \gamma_m\Pi_m)$. It follows from (7) and (9) that

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega), \tag{10}$$

where $\Omega = A^{11}\Pi_\gamma A^{11}$. A consistent estimator $\hat{\Omega}$ of Ω can be obtained when θ is replaced by $\hat{\theta}$, γ_j by n_j/N, and Γ_j by $\hat{\Gamma}_j$. Standard errors of $\hat{\theta}$ follow from square roots of the diagonals of $\hat{\Omega}$.

When data are normal, $\Gamma_j = W_j^{-1}$, $\Pi_j = 4\dot{\delta}'_j W_j \dot{\delta}_j$ and

$$\Pi_\gamma = 4\mathrm{diag}(\gamma_1\dot{\delta}'_1 W_1 \dot{\delta}_1, \cdots, \gamma_m\dot{\delta}'_m W_m \dot{\delta}_m).$$

Since

$$\ddot{F}_j(\beta_{0j}) = 2\dot{\delta}_j' W_j \dot{\delta}_j + O_p(1/\sqrt{n_i}) \tag{11}$$

and A^{11} is a generalized inverse of $\ddot{F}(\theta_0)$, we have

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega), \tag{12}$$

where $\Omega = A^{11}\Pi_\gamma A^{11} = A^{11}$. This corresponds to the standard results obtained when using the normality assumption for multi-samples.

Equation (10) characterizes the distribution of $\hat{\theta}$, the parameter estimator obtained by minimizing (2) under constraint (3). Parallel results for $\hat{\theta}^*$ without the constraint are obtained when replacing A^{11} by A^{-1} in (7) to (12), where

$$A = 2\text{diag}(\gamma_1 \dot{\delta}_1' W_1 \dot{\delta}_1, \cdots, \gamma_m \dot{\delta}_m' W_m \dot{\delta}_m).$$

That is

$$\sqrt{N}(\hat{\theta}^* - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega^*),$$

where $\Omega^* = A^{-1}\Pi_\gamma A^{-1}$. Notice that the Π_γ matrix in Ω^* is the same as the one in Ω, which is block diagonal. As A is also block diagonal, the $\hat{\beta}_j^*$ in $\hat{\theta}^*$ are independent. The correlations between various $\hat{\beta}_j$ in (12), due to the constraint (3), are totally characterized by A^{11}.

Parallel to the likelihood ratio test based on the sample covariance matrices under normality, we would like to have statistics that can be used for inference with nonstandard samples. For this purpose, we will first study the statistic $T_{ML} = NF(\hat{\theta})$. Rescaled statistics for testing the structures $\delta_j = \delta_j(\beta_j)$ and constraint (3) will be given next. A parallel version also will be obtained when interest centers on just testing the constraint (3).

Using the Taylor expansion on $F(\hat{\theta})$ at θ_0 we have

$$F(\hat{\theta}) = F(\theta_0) + \dot{F}'(\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)'\ddot{F}(\bar{\theta})(\hat{\theta} - \theta_0), \tag{13}$$

where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}$. Using equation (11) of Yuan and Bentler (1998b) we have

$$F_j(\beta_{0j}) = e_j' W_j e_j + O_p(1/n_j^{3/2}). \tag{14}$$

Let

$$W = \text{diag}(W_1, \cdots, W_m),$$

$$W_\gamma = \text{diag}(\gamma_1 W_1, \cdots, \gamma_m W_m),$$

$$\dot{\delta} = \text{diag}(\dot{\delta}_1, \cdots, \dot{\delta}_m),$$

$$e = (e_1', \cdots, e_m')',$$

$$e_s = (\sqrt{n_1} e_1', \cdots, \sqrt{n_m} e_m')'.$$

From (14) we have

$$NF(\theta_0) = e_s' W e_s + o_p(1). \tag{15}$$

Similarly, from (8) and (7) respectively we obtain

$$\sqrt{N} \dot{F}(\theta_0) = -2\sqrt{N} \dot{\delta}' W_\gamma e + o_p(1) = -2\dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1)$$

and

$$\sqrt{N}(\hat{\theta} - \theta_0) = A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1), \tag{16}$$

which further lead to

$$N\dot{F}'(\theta_0)(\hat{\theta} - \theta_0) = -2e_s' W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1). \tag{17}$$

Equation (11) implies

$$\ddot{F}(\theta_0) = 2\dot{\delta}' W_\gamma \dot{\delta} + O_p(1/\sqrt{N}). \tag{18}$$

It follows from (16) and (18) that

$$\begin{aligned} N(\hat{\theta} - \theta_0)' \ddot{F}(\bar{\theta})(\hat{\theta} - \theta_0) &= 2e_s' W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1) \\ &= 2e_s' W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1). \end{aligned} \tag{19}$$

Combining (13), (15), (17) and (19) gives

$$\begin{aligned} NF(\hat{\theta}) &= e_s' W e_s - 2e_s' W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + e_s' W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} e_s + o_p(1) \\ &= e_s' U e_s + o_p(1), \end{aligned} \tag{20}$$

where

$$U = W - W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}}.$$

Let

$$\Gamma = \text{diag}(\Gamma_1, \cdots, \Gamma_m),$$

then it follows from (1) that $z = \Gamma^{-\frac{1}{2}} e_s \xrightarrow{\mathcal{L}} N_{p^*}(0, I)$. Now we have from (20)

$$NF(\hat{\theta}) = z'(\Gamma^{\frac{1}{2}} W^{\frac{1}{2}})\{I - W_{\gamma}^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_{\gamma}^{\frac{1}{2}}\}(W^{\frac{1}{2}} \Gamma^{\frac{1}{2}})z + o_p(1). \tag{21}$$

The first term on the right hand side of (21) is a quadratic form in z. Consequently, the asymptotic distribution of $T_{ML} = NF(\hat{\theta})$ can be characterized as the distribution of a quadratic form of normal variates (e.g., section 1.4 of Muirhead, 1982). Let τ_j be the nonzero eigenvalues of $U\Gamma$ and χ_{j1}^2 be independent chi-square variates with degree of freedom 1. Then

$$T_{ML} \xrightarrow{\mathcal{L}} \sum_{j=1}^{p^*-q+r} \tau_j \chi_{j1}^2. \tag{22}$$

Unless all the τ_j are equal, there is no simple distribution to describe the randomness of the right hand side of (22). However, a simple rescaling on T_{ML} can result in a statistic that is better approximated by the $\chi_{p^*-q+r}^2$ distribution. Let $c = \text{tr}(U\Gamma)/(p - q + r)$. Then the rescaled statistic

$$T_{RML} = T_{ML}/\hat{c}$$

approaches a distribution with mean equal to that of $\chi_{p^*-q+r}^2$. Similar statistics for inference based on sample covariance matrices have been proposed by Satorra and Bentler (1988) for single sample analysis, and by Satorra (2000) for multi-sample analysis. Simulation work in the single sample case with the sample covariance matrix has shown that this type of correction works remarkably well under a variety of conditions (e.g., Hu, Bentler, & Kano, 1992; Curran, West, & Finch, 1996).

A special case results when data are normal and sample means and covariance matrices are used in (2). Then $\Gamma = W^{-1}$. Since $W_{\gamma}^{\frac{1}{2}} \dot{\delta} A^{11} \dot{\delta}' W_{\gamma}^{\frac{1}{2}}$ is an idempotent matrix with rank $(q - r)$, it follows from (21) that

$$T_{ML} \xrightarrow{\mathcal{L}} \chi_{p^*-q+r}^2,$$

which is the basis for the likelihood ratio statistic.

In order to study the property of the test statistic $T_{ML}^{(h)} = N[F(\hat{\theta}) - F(\hat{\theta}^*)]$, we need also to characterize the distribution of $NF(\hat{\theta}^*)$. This can be obtained when replacing the A^{11} in (21) by A^{-1}. Specifically, let

$$U^* = W - W^{\frac{1}{2}} W_{\gamma}^{\frac{1}{2}} \dot{\delta} A^{-1} \dot{\delta}' W_{\gamma}^{\frac{1}{2}} W^{\frac{1}{2}},$$

10

then

$$NF(\hat{\theta}^*) = e'_s U^* e_s + o_p(1). \tag{23}$$

Since $W_\gamma^{\frac{1}{2}} \dot{\delta} A^{-1} \dot{\delta}' W_\gamma^{\frac{1}{2}}$ is an idempotent matrix with rank q, there are only $p^* - q$ nonzero eigenvalues of $U^* \Gamma$. Denote these as τ_j^*, $j = 1, \cdots, p^* - q$, then

$$NF(\hat{\theta}^*) \xrightarrow{\mathcal{L}} \sum_{j=1}^{p^*-q} \tau_j^* \chi_{j1}^2.$$

Similarly, letting $c^* = \text{tr}(U^* \Gamma)/(p^* - q)$, the rescaled statistic

$$T_{RML}^* = NF(\hat{\theta}^*)/\hat{c}^*$$

approaches a distribution with mean equal to that of $\chi_{p^*-q}^2$.

For testing the constraint $h(\theta) = 0$, based on (20) and (23), the statistic $T_{ML}^{(h)}$ can be expressed as

$$T_{ML}^{(h)} = e'_s(U - U^*)e_s + o_p(1). \tag{24}$$

It can be verified that

$$U - U^* = W^{\frac{1}{2}} W_\gamma^{\frac{1}{2}} \dot{\delta}(A^{-1} - A^{11})\dot{\delta}' W_\gamma^{\frac{1}{2}} W^{\frac{1}{2}} \tag{25}$$

and $W_\gamma^{\frac{1}{2}} \dot{\delta}(A^{-1} - A^{11})\dot{\delta}' W_\gamma^{\frac{1}{2}}$ is an idempotent matrix of rank r. It follows from (24) and (25) that

$$T_{ML}^{(h)} \xrightarrow{\mathcal{L}} \sum_{j=1}^{r} \kappa_j \chi_{j1}^2,$$

where κ_j are the nonzero eigenvalues of $(U - U^*)\Gamma$. Let $c_h = \text{tr}[(U - U^*)\Gamma]/r$, then

$$T_{RML}^{(h)} = T_{ML}^{(h)}/\hat{c}_h$$

converges to a distribution with mean r. Satorra (2000) gave a rescaled version of the Wald type statistic for testing a constraint like (3) when sample moment matrices are used in (2).

A more general version than testing $h(\theta) = 0$ is to test one set of constraints nested within another set of constraints. Let the two sets of constraints be represented by $h(\theta) = 0$ and $g(\theta) = 0$, and

$$\mathcal{R}_h = \{\theta : h(\theta) = 0\} \subset \mathcal{R}_g = \{\theta : g(\theta) = 0\}. \tag{26}$$

11

A rescaled statistic for testing (26) can be derived similarly. Let U_h and U_g represent the U matrices corresponding to the constraints, then the likelihood ratio statistic $T_{ML}^{(h \subset g)} = T_{ML}^{(h)} - T_{ML}^{(g)}$ can be written as

$$T_{ML}^{(h \subset g)} = e_s'(U_h - U_g)e_s + o_p(1).$$

Let r_h and r_g be the numbers of independent constraints in $h(\theta) = 0$ and $g(\theta) = 0$ respectively, then

$$\text{tr}[(U_h - U_g)\Gamma] = \text{tr}(U_h\Gamma) - \text{tr}(U_g\Gamma) = (p^* - q + r_h)c_h - (p^* - q + r_g)c_g,$$

we have

$$c_{(h \subset g)} = [(p^* - q + r_h)c_h - (p^* - q + r_g)c_g]/(r_h - r_g). \tag{27}$$

Suppose a software has already had the rescaling option for nonstandard samples with constraint built in, but rescaling for nested models is still not available. Then we can get $c_{(h \subset g)}$ using (27) in a straightforward way. Let $T_{ML}^{(g)}$ and $T_{RML}^{(g)}$ be the likelihood ratio statistic and the rescaled statistic, respectively. Then $\hat{c}_g = T_{ML}^{(g)}/T_{RML}^{(g)}$, and similarly to obtain $\hat{c}_h$. Since $p^* - q + r_h$ and $p^* - q + r_g$ are just the degrees of freedom in the two models, $\hat{c}_{(h \subset g)}$ immediately follows from (27). A similar procedure for modeling based on sample moments is discussed in Satorra and Bentler (1999).

3. Estimating δ_j and Γ_j for Nonstandard Samples

Estimation of covariance matrices for nonstandard samples can be accomplished by various procedures that have been described in the statistical literature. Since the most commonly encountered nonstandard situations in the social and behavioral sciences are probably nonnormal samples, samples with outliers, and samples with missing data, we will deal with each of these situations in sequence. The following procedures for estimating δ_j and Γ_j are based on our experience with various practical nonstandard samples. A further discussion of these procedures applied to exploratory factor analysis can be found in Yuan, Marshall and Bentler (1999).

3.1. Nonnormal data

When samples come from distributions with heavy tails which are not due to outliers, sample mean vectors and covariance matrices may still be unbiased estimates of their population counterparts. For example, if a sample is from a multivariate t-distribution, the sample does not contain outliers but is still nonnormal. In such a case, using sample mean vectors and covariance matrices in (2) still leads to consistent parameter estimates when all of the population second-order moments exist. In order to obtain consistent standard errors, we need to have the population fourth-order moment matrices to exist. Let $X_{1j}, \cdots, X_{n_j j}$ be the sample from the jth group with sample mean $\bar{X}_j$, let $Y_{ij} = \{X_{ij}', \text{vech}'[(X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j)']\}'$ with sample mean vector $\bar{Y}_j$ and sample covariance matrix S_{Yj}. Then $t_{nj} = \bar{Y}_j$ and

$$\hat{\Gamma}_j = S_{Yj}$$

is a consistent estimator of Γ_j in (1). Using the sample fourth-order moment matrix to estimate its population counterpart was first used by Browne (1982, 1984) in the context of covariance structure analysis. Mooijaart and Bentler (1985) formulated an efficient way to compute S_{Yj}.

3.2. Data with outliers

With nonnormal data, sample covariance matrices are no longer the most efficient estimates of their population counterparts. If the nonnormality is created by outliers, analysis based on sample covariance matrices can be misleading to a greater or lesser degree, depending on the influence of the outliers. There are two ways to deal with outliers. One is to identify the influential cases through some analytical procedure and make a subjective decision whether to keep them or remove them. Another way is to use a robust approach. Whether any cases are outliers or just influential cases, their effect will be automatically downweighted through this approach. Compared with an outlier removal approach, the merit of a downweighting approach was discussed by Rousseeuw and van Zomeren (1990). We will also use the downweighting approach here. We especially recommend the Huber-type weight, because of its explicit control of the percentage of outliers when the majority of a data cloud follows a multivariate normal distribution.

13

For the sample $X_{1j}, \cdots, X_{n_j j}$ from the jth population, let

$$d_{ij} = d(X_{ij}, \mu_j, \Sigma_j) = [(X_{ij} - \mu_j)'\Sigma_j^{-1}(X_{ij} - \mu_j)]^{1/2}$$

be the Mahalanobis distance and $u_1(t)$ and $u_2(t)$ be some nonnegative scalar functions. Maronna (1976) defined robust M-estimators $(\hat{\mu}_j, \hat{\Sigma}_j)$ by solving the following equations

$$\mu_j = \sum_{i=1}^{n_j} u_1(d_{ij})X_{ij} / \sum_{i=1}^{n_j} u_1(d_{ij}), \tag{28a}$$

and

$$\Sigma_j = \sum_{i=1}^{n_j} u_2(d_{ij}^2)(X_{ij} - \mu_j)(X_{ij} - \mu_j)'/n_j. \tag{28b}$$

If $u_1(t)$ and $u_2(t)$ are decreasing functions, cases with larger d_{ij}s will get smaller weights than those with smaller d_{ij}s. If a case lies far away form the majority of the data cloud, its effect will be downweighted. A solution to (28) can be obtained through iteratively reweighted least squares (e.g., Green, 1984). The Huber-type weight is given by

$$u_1(d) = \begin{cases} 1, & \text{if } d \leq r \\ r/d, & \text{if } d > r \end{cases} \tag{29}$$

and $u_2(d^2) = \{u_1(d)\}^2/\beta$ (e.g., Tyler, 1983). Here, r^2 satisfies $P(\chi_{p_j}^2 > r^2) = \alpha$, α is the percentage of outliers one wants to control assuming the massive data cloud follows a multivariate normal distribution, and β is a constant such that $E\{\chi_p^2 u_2(\chi_p^2)\} = p_j$. This approach makes the estimator $\hat{\Sigma}_j$ unbiased for Σ_j if sampling is from a p_j-variate normal distribution. Notice that only the tuning parameter α needs to be decided in applying the Huber-type weight, since r and β are just functions of α.

Let X_{ij}, $i = 1, \cdots, n_j$ follow an elliptical distribution (e.g., Fang, Kotz, & Ng, 1990) and $S_{nj} = \hat{\Sigma}_j$ be a robust covariance matrix estimate. S_{nj} generally does not converge to the population covariance matrix. Instead, it converges to a constant times the population covariance matrix: $\kappa_j \Sigma_j$. The positive scalar κ_j depends on the weight function used in the estimation procedure, as well as the unknown underlying distribution of the data. Because of this issue, we recommend using the Huber-type weight with the same α for every sample of the m groups. Since multiple samples are commonly obtained by administering the same questionnaire to m groups, the massive data cloud in each sample should resemble the massive

14

data clouds of other samples, even though one may contain fewer or more influential cases than the others. Actually, robust covariance matrices from separate samples are much more similar than traditional sample counterparts when data have heavy tails (Yuan, Marshall & Weston, 1999).

We will resort to the estimating equation approach for getting a consistent estimator of Γ_j. Rewrite (28) as

$$\frac{1}{n_j} \sum_{i=1}^{n_j} G_j(X_{ij}, \delta_j) = 0, \tag{30a}$$

where

$$G_j(x, \delta_j) = \left(\begin{array}{c} u_1[d(x, \mu_j, \Sigma_j)](x - \mu_j) \\ u_2[d^2(x, \mu_j, \Sigma_j)]\text{vech}[(x - \mu_j)(x - \mu_j)'] - \sigma_j \end{array} \right). \tag{30b}$$

Then

$$\sqrt{n_j}(\hat{\delta}_j - \delta_{j0}) \xrightarrow{\mathcal{L}} N(0, \Gamma_j), \tag{31}$$

where $\Gamma_j = H_j^{-1} B_j H_j'^{-1}$ with

$$H_j = E[\dot{G}_j(X_{ij}, \delta_{j0})] \ \text{ and } \ B_j = E[G_j(X_{ij}, \delta_{j0})G'(X_{ij}, \delta_{j0})].$$

A consistent estimator of Γ_j can be obtained by using consistent estimates for H_j and B_j; these are given by

$$\hat{H}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \dot{G}_j(X_{ij}, \hat{\delta}_j) \ \text{ and } \ \hat{B}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} G_j(X_{ij}, \hat{\delta}_j)G'_j(X_{ij}, \hat{\delta}_j).$$

3.3. Normal missing data

Data are said to be missing completely at random (MCAR) if their absence does not depend on the missing values themselves nor on the observed values of the other variables. Data are said to be missing at random (MAR) if the missing data do not depend on the missing values themselves, but may depend on the observed values of other variables. For the jth sample with missing data, denote X_{ij} as the vector of observed variables for the ith case with dimension p_{ij}. Then $E(X_{ij}) = \mu_{ij}$ and $\text{Cov}(X_{ij}) = \Sigma_{ij}$ are respectively subvector of μ_j and submatrix of Σ_j. Under the assumption of normality, the log likelihood function based on X_{ij} is

$$l_{ij}(\delta_j) = \frac{p_{ij}}{2} \log(2\pi) - \frac{1}{2}[\log |\Sigma_{ij}| + (X_{ij} - \mu_{ij})'\Sigma_{ij}^{-1}(X_{ij} - \mu_{ij})]. \tag{32a}$$

15

The MLE of δ_j is actually obtained by maximizing

$$l_j(\delta_j) = \sum_{i=1}^{n_j} l_{ij}(\delta_j). \tag{32b}$$

Consequently, $\hat{\delta}_j$ satisfies the following generalized estimating equation

$$G_j(\hat{\delta}_j) = 0, \tag{33a}$$

where

$$G_j(\delta_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} \dot{l}_{ij}(\delta_j). \tag{33b}$$

A solution to (33) is straightforward using the EM-algorithm developed in Dempster, Laird and Rubin (1977). Specific steps are also discussed in detail in Little and Rubin (1987). Assuming the missing data mechanism is MAR, using the result for generalized estimating equations (e.g., Liang & Zeger, 1986; Yuan & Jennrich, 1998), we have

$$\sqrt{n_j}(\hat{\delta}_j - \delta_{j0}) \xrightarrow{\mathcal{L}} N(0, \Gamma_j), \tag{34a}$$

where $\Gamma_j = A_j^{-1} B_j A_j^{-1}$ with

$$A_j = -E[\dot{G}_j(\delta_{j0})], \quad B_j = E[\frac{1}{n_j} \sum_{i=1}^{n_j} \dot{l}_{ij}(\delta_j) \dot{l}'_{ij}(\delta_j)]. \tag{34b}$$

A consistent estimate of Γ_j is given by

$$\hat{\Gamma}_j = \hat{A}_j^{-1} \hat{B}_j \hat{A}_j^{-1}$$

with

$$\hat{A}_j = -\dot{G}_j(\hat{\delta}_j), \quad \hat{B}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \dot{l}_{ij}(\hat{\delta}_j) \dot{l}'_{ij}(\hat{\delta}_j).$$

When $X_{ij} \sim N(\mu_{ij}, \Sigma_{ij})$, the corresponding observed information matrix is given by $\hat{A}_j$ (Little & Rubin, 1987; Kenward & Molenberghs, 1998) and $\hat{\Gamma}_j = \hat{A}_j^{-1}$ is consistent for Γ_j in (34). For a general nonnormal distribution, the result in (34) is also correct as long as the missing data mechanism is MCAR. However, as discussed in Laird (1988), some bias may exist in using $\hat{\delta}_j$ to estimate δ_{j0} when data are not normal and missing data are MAR. Ideally, it would be desirable to model a data set through ML to avoid bias. However, because of complexity of the real world, there will be always a discrepancy between the

16

underlying distribution of the data and a carefully specified modeling distribution. So we would consider the normal distribution assumption for missing data to offer only a working assumption in multivariate analysis. Fortunately, for estimating population mean vectors and covariance matrices, a recent simulation study by Yuan and Bentler (1999) indicates that the bias is minimal for a variety of nonnormal distributions. It is important to realize that once t_{nj} is used in (2), the parameter estimate $\hat{\theta}$ will be the same whatever missing data mechanism is assumed. The important question is, which procedure leads to a more accurate evaluation of model structures. According to the results for single group analysis in Yuan and Bentler (1999), inference based on (34) is much more accurate than that based on the observed information matrix. We recommend using (34) for estimating Γ_j.

3.4. Nonnormal missing data

When a sample contains both missing data and outliers, normal theory based missing data procedures will lead to inaccurate conclusions. As in the situation with complete data, appropriate downweighting procedures are needed for better inference. Little and Smith (1987) proposed several methods for such a purpose. Little (1988) further proposed the EM-algorithm for modeling missing data by a multivariate t-distribution as well as a multivariate contaminated normal distribution. Here, we will outline a procedure for using the multivariate t-distribution to get $t_{nj} = \hat{\delta}_j$ and $\hat{\Gamma}_j$.

The density of the p-variate t-distribution with degrees of freedom k is given by

$$f(x|\mu, \Sigma, k) = \frac{\Gamma[(p+k)/2]}{(k\pi)^{p/2}\Gamma(k/2)}|\Sigma|^{-1/2}(1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{k})^{-(p+k)/2}. \tag{35}$$

If X follows (35) with $k > 2$, then $E(X) = \mu$ and $\text{Cov}(X) = k\Sigma/(k-2)$. So the MLE of Σ will converge to $\kappa\text{Cov}(X)$ with $\kappa = (k-2)/k$. As discussed in subsection 3.2, we recommend using t-distributions with the same degrees of freedom for each of the m samples.

Denote (35) as $Mt_p(\mu, \Sigma, k)$. Since a marginal distribution of (35) is also a t-distribution with the same degrees of freedom (e.g., Fang, Kotz, & Ng, 1990; Kano, 1994), if $X_{ij} \sim Mt_{p_{ij}}(\mu_{ij}, \Sigma_{ij}, k)$, its log likelihood function is

$$l_{ij}(\delta_j) = c_{ij} - \frac{1}{2}\log|\Sigma_{ij}| - \frac{(p_{ij}+k)}{2}\log[1 + \frac{(X_{ij}-\mu_{ij})'\Sigma_{ij}^{-1}(X_{ij}-\mu_{ij})}{k}]. \tag{36a}$$

where $\delta_j = (\mu'_j, \delta'_j)'$. The MLE of δ_j can be obtained by maximizing

$$l_j(\delta_j) = \sum_{i=1}^{n_j} l_{ij}(\delta_j). \tag{36b}$$

Similarly, as in the last subsection, the $\hat{\delta}_j$ satisfy the following generalized estimating equation

$$G_j(\hat{\delta}_j) = 0, \tag{37a}$$

where

$$G_j(\hat{\delta}_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} \dot{l}_{ij}(\delta_j). \tag{37b}$$

We can maximize (36) for δ_j and k simultaneously. However, a data set may not exactly follow a t-distribution, and the simultaneous ML procedure may not lead to the most efficient estimator of δ_j. In addition to requiring much more complicated computations, a nonadmissible MLE of k may occur with some practical data as discussed in Lange, Little and Taylor (1989). Little (1988) recommended using several prefixed ks, and then using the $\hat{\delta}_j$ corresponding to the largest $l_j(\hat{\delta}_j)$ as the final parameter estimator. Real data examples in Yuan and Bentler (1998a,b) indicate that most of the smaller ks ($1 \leq k \leq 5$) can effectively control the influence of outliers in SEM. In practice, we suggest following Little's recommendation to try several prefixed k (e.g., $1 \leq k \leq 5$). With a fixed k, the solution to (37) is straightforward using the EM-algorithm developed in Little (1988).

As discussed for the normal theory based likelihood function, the t-distribution in (36) is only a working assumption for downweighting outliers. Real data may not exactly follow such an assumption. Consequently, computations to obtain good standard error estimators need to be modified. We will use a sandwich-type covariance matrix to describe the distribution of $\hat{\delta}_j$. With a MAR assumption for the missing data mechanism, this is given by

$$\sqrt{n_j}(\hat{\delta}_j - \hat{\delta}_{j0}) \xrightarrow{\mathcal{L}} N(0, \Gamma_j), \tag{38a}$$

where $\Gamma_j = A_j^{-1} B_j A_j^{-1}$ with

$$A_j = -E[\dot{G}_j(\delta_{j0})], \quad B_j = E[\frac{1}{n_j} \sum_{i=1}^{n_j} \dot{l}_{ij}(\delta_{j0}) \dot{l}'_{ij}(\delta_{j0})]. \tag{38b}$$

A consistent estimate of Γ_j is obtained from

$$\hat{\Gamma}_j = \hat{A}_j^{-1} \hat{B}_j \hat{A}_j^{-1} \tag{38c}$$

18

with

$$\hat{A}_j = -\dot{G}_j(\hat{\delta}_j), \quad \hat{B}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} \dot{l}_{ij}(\hat{\delta}_j)\dot{l}'_{ij}(\hat{\delta}_j).$$

When evidence suggests that a data set does closely follow the t-distribution used in estimating $\hat{\delta}_j$, we may use the inverse of the observed information matrix $\hat{A}_j^{-1}$ instead of (38c) to describe the behavior of $\hat{\delta}_j$. However, the result in (38) is more accurate under violation of distributional assumptions. When the missing data mechanism is MAR, and the data set does not follow a multivariate t-distribution, there may exist a bias for using $\hat{\delta}_j$ to estimate δ_{j0} (Laird, 1988). That is, the $\hat{\delta}_j$ may not approach δ_j as the sample size increases. Based on results in Yuan and Bentler (1999), we suspect that the bias would be minimal for most of the commonly encountered continuous distributions. Further studies on bias associated with the MLE from a misspecified t-distribution and different missing data mechanisms would provide a valuable guide for future application of the method. For the same reason as discussed for the normality working assumption in the previous subsection, our interest is to obtain a better description of the variability in $t_{nj} = \hat{\delta}_j$.

4. Discussion and Conclusion

Motivated by the typical nonstandard samples for survey data in practice, that is, samples with nonnormal distributions, missing data, and outliers, we proposed replacing the sample mean vectors and sample covariance matrices by more appropriate quantities t_{nj} in the normal theory based likelihood function for multi-group SEM. Because the parameter estimator $\hat{\theta}$ depends on t_{nj}, possible merits of t_{nj} such as efficiency and robustness are inherited by $\hat{\theta}$. Standard errors of $\hat{\theta}$ are obtained through a generalized estimating equation approach. Two rescaled test statistics, one for the overall structural model with constraints, and one just for the constraints, are provided. Procedures for obtaining appropriate t_{nj} for each situation, and their large sample covariance matrices, are given for each of several nonstandard sampling setups. Our approach is so general that it can be applied to any types of nonstandard samples once a new method for estimating the population mean vector and covariance matrix together with the associated Γ matrix are available for such samples.

We have chosen to use the normal theory based likelihood function as the discrepancy function to measure the distance between t_{nj} and δ_j because of its relative advantage in

reaching convergence. A generalized least squares approach using $\hat{\Gamma}_j^{-1}$ as weights is equally general, and development along this line is straightforward.

It would be ideal to demonstrate the procedures in section 2 with a practical example for each of the various types of nonstandard samples considered. Due to the unavailability of multiple samples that contain the various features, such a demonstration will not be done at present. Future research clearly should be directed to evaluating our proposals. Based on our experience with the inference procedures in section 2 for various one group nonstandard samples, we would expect the proposed procedures to generally give much more reliable model and parameter evaluation than classical procedures based on sample covariance matrices. Our recommendation is to use the proper methods given in section 3 to estimate (μ_j, Σ_j) and Γ_j when nonstandard samples occur in practice, and follow the procedure in section 2 for model evaluation.

References

Aitchison, J. & Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, *29*, 813–828.

Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (ed.) *Sociological methodology 1987* (pp. 71–103). San Francisco: Jossey-Bass.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (eds.) *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). New Jersey: Lawrence Erlbaum Associates.

Arminger, G. & Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, *85*, 195–203.

Austin, J. T. & Calderón, R. F. (1996). Theoretical and technical contributions to structural equation modeling: An updated annotated bibliography. *Structural Equation Modeling*, *3*, 105–175.

Austin, J. T. & Wolfle, D. (1991). Annotated bibliography of structural equation modeling: Technical work. *British Journal of Mathematical and Statistical Psychology, 44*, 93–152.

Bentler, P. M. (2000). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software. (www.mvsoft.com)

Bentler, P. M. & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (ed.) *Multivariate analysis VI* (pp. 9–42). Amsterdam: North-Holland.

Bentler, P. M. & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47*, 541–570.

Bentler, P. M., Lee, S.-Y. & Weng, L.-J. (1987). Multiple population covariance structure analysis under arbitrary distribution theory. *Communication in Statistics-Theory and Method, 16*, 1951–1964.

Berkane, M. & Bentler, P. M. (1988). Estimation of contamination parameters and identification of outliers in multivariate data. *Sociological Methods and Research, 17*, 55–64.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A. & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. In P. V. Marsden (Ed.), *Sociological Methodology 1991,* pp. 235–262. Oxford: Basil Blackwell.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.) *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Browne, M. W. & Arminger, G. (1995). Specification and estimation of mean and covariance structure models. In G. Arminger, C.C. Clogg, & M.E. Sobel (eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows,* Thousand Oaks, CA: Sage.

Cadigan, N. G. (1995). Local influence in structural equation models. *Structural Equation Modeling, 2*, 13–30.

Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B, 39*, 1–38.

Dunn, G., Everitt, B. & Pickles, A. (1993). *Modeling covariances and latent variables using EQS.* London: Chapman & Hall.

Fang, K.-T., Kotz, S. & Ng., K. W. (1990). *Symmetric multivariate and related distributions.* London: Chapman & Hall.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistent alternatives (with discussion). *Journal of the Royal Statistical Society, Series B, 46*, 149–192.

Hoyle, R. (ed.) (1995). *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage.

Hu, L.T., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.

Jamshidian, M. & Bentler, P. M. (1999). Using complete data routines for ML estimation of mean and covariance structures with missing data. *Journal Educational and Behavioral Statistics, 23*, 21–41.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8 user's reference guide.* Chicago: Scientific Software International.

Kano, Y. (1994). Consistency property of elliptical probability density functions. *Journal of Multivariate Analysis, 51*, 139–147.

Kenward, M. G. & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science, 13*, 236–247.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine, 7*, 305–315.

Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association, 84*, 881–896.

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

Lee, S. Y. & Wang, S. J. (1996). Sensitivity analysis of structural equation models. *Psychometrika, 61*, 93–108.

Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Little, R. J. A. (1988). Robust estimation of the mean and covariance matrix from data

with missing values. *Applied Statistics, 37*, 23–38.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Little, R. J. A. & Smith, P. J. (1987). Editing and imputation for quantitative survey data. *Journal of the American Statistical Association, 82*, 58–68.

Magnus, J. R. & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* New York: Wiley.

Marcoulides, G. A. & Schumacker, R. E. (eds.) (1996). *Advanced structural equation modeling: Issues and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates.

Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics, 4*, 51–67.

Mooijaart, A. & Bentler, P. M. (1985). The weight matrix in asymptotic distribution-free methods. *British Journal of Mathematical and Statistical Psychology, 38*, 190–196.

Mueller, R. O. (1996). *Basic principles of structural equation modeling.* New York: Springer Verlag.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* New York: Wiley.

Muthén, B. (1989). Multiple group structural modelling with nonnormal continuous variables. *British Journal of Mathematical and Statistical Psychology, 42*, 55–62.

Muthén, B., Kaplan & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52*, 431–462.

Rousseeuw, P. J. & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association, 85,* 633–639.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In D.D.H. Heijmans, D.S.G. Pollock, & A. Satorra (eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233-247). Dordrecht: Kluwer Academic.

Satorra, A. & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 proceedings of Business and Economics Sections* (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A. & Bentler, P. M. (1999). A scaled difference chi-square test statistic for moment structure analysis. Under review.

Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation mod-

eling. Mahwah, NJ: Lawrence Erlbaum Associates.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.

Tanaka, Y., Watadani, S., & Moon, S. H. (1991). Influence in covariance structure analysis: With an application to confirmatory factor analysis. *Communication in Statistics-Theory and Method, 20*, 3805–3821.

Tremblay, P. F. & Gardner, R. C. (1996). On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling, 3*, 93–104.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika, 70*, 411–420.

Yuan, K.-H. & Bentler, P. M. (1998a). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology, 51*, 63–88.

Yuan, K.-H. & Bentler, P. M. (1998b). Structural equation modeling with robust covariances. In A. E. Raftery (ed.) *Sociological methodology 1998* (pp. 363–396). Boston: Blackwell Publishers.

Yuan, K.-H. & Bentler, P. M. (1999). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Under review.

Yuan, K.-H. & Bentler, P. M. (2000). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika, 65*, 43–58.

Yuan, K.-H., Chan, W. & Bentler, P. M. (2000). Robust transformation with applications to structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 52*.

Yuan, K.-H. & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis, 65*, 245–260.

Yuan, K.-H., Marshall, L., & Bentler, P. M. (1999). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. Under review.

Yuan, K.-H., Marshall, L. & Weston, R. (1999). Cross-validation through downweighting influential cases in structural equation modeling. Under review.