Title

Mean and Covariance Structure Analysis: Theoretical and Practical Improvements

Authors
K. H. Yuan
P. M. Bentler

Publication Date
2011-10-25

UCLA Statistical Series
Report No. 194

Mean and Covariance Structure Analysis: Theoretical and Practical Improvements*

Ke-Hai Yuan and Peter M. Bentler

June 21, 1995

*Ke-Hai Yuan is Statistician, and Peter M. Bentler is Professor, Department of Psychology and Center for Statistics, University of California, Los Angeles, Box 951563, Los Angeles, CA 90095-1563. This work was supported by National Institute on Drug Abuse Grants DA01070 and DA00017.

Abstract

The most widely used multivariate statistical models in the social and behavioral sciences involve linear structural relations among observed and latent variables. In practice, these variables are generally nonnormally distributed, and hence classical multivariate analysis, based on multinormal error-free variables having no simultaneous interrelations, is not adequate to deal with such data. Since structural relations among variables imply a structure for the multivariate product moments of the variables, general methods for the analysis of mean and covariance structures have been proposed to estimate and test particular model structures. Unfortunately, extant statistical tests, such as the likelihood ratio test (LRT) and a test based on asymptotically distribution free (ADF) covariance structure analysis, have been found to be virtually useless in practical model evaluation at finite sample sizes with nonnormal data. For example, in one condition of a simulation on confirmatory factor analysis, the LRT rejected the true model about 99.5% of the time at sample sizes from $n = 150$ to $n = 5000$, while the ADF test either always rejected the true model or did not converge at $n = 150$, rejected the true model over 90% of the time at $n = 250$, and did not perform nominally until $n = 5000$. Clearly, improved methods are needed.

We take a new look at the basic statistical theory of structural models under arbitrary distributions, using the methodology of nonlinear regression and generalized least squares estimation. For example, we adopt the use of residual weight matrices from regression theory. We develop a series of estimators and tests based on pseudo maximum likelihood and arbitrary distribution theory. We obtain a type of probabilistic Bartlett correction for various test statistics that can be simply applied in practice. A small simulation study replicates the extremely inadequate performance of one of our own, and the classical ADF, model tests. In contrast, our corrected statistics have approximately correct means at all sample sizes, though there is a tendency for their variances to be too low at the smallest sample sizes leading to some "overacceptance" of the true model.

KEY WORDS: Mean and covariance structures; structural relations; structural equations; asymptotic distribution free; test; small sample sizes; corrections on tests.

1 INTRODUCTION

Linear structural equation models can be described as a class of models in which a $p-$variate vector of variables X is presumed to be generated as $X = A\zeta$, where the matrix $A = A(\gamma)$ is a function of a basic vector of parameters, and the underlying k ($k \geq p$) generating variables ζ may represent measured, latent, or residual random or fixed variables (e.g., Anderson, 1989; Bentler, 1983a; Satorra and Neudecker, 1994). Examples of such models are path analysis, confirmatory factor analysis, simultaneous equation, and errors in variables models, and especially the generalized linear structural relations models made popular in the social and behavioral sciences by computer programs such as LISREL (Jöreskog & Sörbom, 1993) and EQS (Bentler & Wu, 1995a,b). These models represent by far the most widely used multivariate models in the social and behavioral sciences (e.g., Bollen & Long, 1993; Byrne, 1994; Hoyle, 1995), to which a new journal *Structural Equation Modeling* is devoted entirely. While there are many approaches towards estimating and testing specialized variants of these models (see e.g., Anderson, 1994; Fuller, 1987), generic classical approaches such as regression often can not be used because the ζ variables may all be hypothetical variables which are in principle not observable. An example is factor analysis, in which the ζ variables are common and unique factors that, unlike principal components, cannot be expressed as linear combinations of the X variables. However, since the models imply a parametric structure for the multivariate moments of the X variables, especially, the means and covariances (but also higher-order moments, Bentler, 1983a), it is possible to estimate and test the models as so-called mean and covariance structure models. That is, the parameters can be estimated, and the model null hypothesis tested, without use of the ζ variables by relying on unstructured sample estimators $\bar{X}$ and S of the population mean vector μ and covariance matrix Σ of the X variables. This can be done because any linear structural model implies a more basic set of parameters θ, so that $\mu = \mu(\theta)$ and $\Sigma = \Sigma(\theta)$. The q parameters in θ represent elements of γ as well as the intercepts, regression coefficients, and variances and covariances of the ζ variables. Of course, moment structure models can be specified without relying on a linear structural model for motivation; a classic example is the intraclass model in which μ is unstructured and $\Sigma = c11^T + (b - c)I$. So, while linear structural models provide the most typical motivation for mean and covariance structural models, such models have a broader general relevance.

Estimation and testing mean and covariance structure models is a straightforward matter when the variables ζ, and hence the X variables, are presumed to be multivariate normally distributed. Then, with a sample $X_1, \ldots, X_n$ from X, classical multivariate analysis can be brought to bear, e.g., via the normal theory maximum likelihood estimator (MLE) and the likelihood ratio test (LRT). Unfortunately, most social and behavioral data are clearly nonnormal (e.g., Micceri, 1989), so classical methods can yield very distorted results. For example, in one condition of a simulation with a confirmatory factor analysis model, Hu, Bentler, and Kano (1992) found that the LRT rejected the true model in 1194 out of 1200 samples at sample sizes that ranged from n=150 to n=5000. Nonetheless, MLE and LRT remain by far the most widely used methodology in practice (e.g., Gierl & Mulvenon, 1995). Some alternatives to LRT in this context have been proposed (Arminger & Schoenberg, 1989; Bentler, 1994; Browne, 1984, eq. 2.20; Kano, 1992; Satorra & Bentler, 1988, 1994), but these methods accept the MLE, which is not fully efficient in the face of violation of distributional assumptions.

In order to solve the fundamental problem of incorrect and misleading LRT statistics, and to obtain an estimator with greater precision, Browne (1982, 1984) and

Chamberlain (1982) used Ferguson's (1958) minimum modified χ^2 principle to develop an "asymptotically distribution free" (ADF) methodology (called "minimum distance" by Chamberlain) for covariance structure analysis (in which μ is unstructured). This approach was extended to asymptotically equivalent linearized estimators by Bentler (1983b) and Bentler and Dijkstra (1985), and to mean and covariance structure analysis by Bentler (1989, Ch. 10) and Muthén (1989). While the ADF methodology is correct asymptotically, and it can perform reasonably well with small models (e.g., Henly, 1993), in larger models with small to medium sized samples it can be extremely misleading (e.g., Hu et al, 1992; Muthén & Kaplan, 1992; West, Finch, & Curran, 1995). For example, in the condition noted above, Hu et al found that the method either always rejected the true model or did not converge at n=150; rejected the true model over 90% of the time at n=250; and did not perform nominally until n=5000. Although a computationally intensive improvement on ADF statistics has been made (Yung & Bentler, 1994), and in spite of technical developments since 1982 as noted below, ADF theory thus also remains clearly inadequate to evaluate linear structural or mean and covariance structure models. See Bentler and Dudgeon (in press) for a review.

As a result of the remarkable failure of ADF theory to be relevant to nonasymptotic samples, it seems time to take another basic look at estimation of structural models under arbitrary distributions. We do this by invoking the methodology of nonlinear regression, which has previously been considered relevant by Browne (1982), Lee and Jennrich (1984), Shapiro (1986), Fuller (1987), and Bentler (1993). Unfortunately, these workers did not provide any methods to improve on ADF. We shall see, however, that theoretical as well as empirical improvements on ADF can be achieved. In particular, based on the standard regression idea of using residual weight matrices in generalized least squares (GLS) estimation, we develop a class of estimators and tests. Among these are Bartlett-type corrected ADF statistics which are asymptotically equivalent to ADF but outperform it in nonasymptotic samples.

Turning now to more technical matters, Satorra (1992) modeled the means and covariances simultaneously. His GLS method is to model the sample covariance of $(X^T, c1)$, where c is a known constant, and use a singular matrix as a weight matrix. Under the assumption that X has finite eighth order moments, Bentler (1989), Muthén (1989), and Browne and Arminger (1995) model the sample mean and covariance matrix $(\bar{X}, S)$ simultaneously, using the inverse of the sample covariance of $(\bar{X}, S)$ as a weight matrix. Our approach is different from Browne and Arminger's (1995) in that we model the raw moment of $(X, vech(XX^T))$, where $vech(.)$ is the function which transforms a symmetric matrix into a vector by picking the nonduplicated elements of the matrix. Model structures on raw moments generated by linear structural models on the variables are well-known (e.g., Bentler, 1983b; Satorra and Neudecker, 1994). Our perspective is different from Satorra's in that we consider the statistical properties of the estimator in a more rigorous way. Following these authors, we can use the inverse of the sample covariance of $(X, vech(XX^T))$ as a weight matrix. However, in the spirit of regression, we show that another asymptotically equivalent weight matrix is the inverse of the cross products of the fitted residuals. Moreover, we do not need any distributional assumption besides that the first four moments of X are finite.

We have the following notation: $Y_i = vech(X_i X_i^T)$, $Z_i = (X_i^T, Y_i^T)^T$, $\sigma(\theta) = vech(\Sigma(\theta))$, $\tau(\theta) = vech(\mu(\theta)\mu^T(\theta))$ and

$$\xi(\theta) = \begin{pmatrix} \mu(\theta) \\ \sigma(\theta) + \tau(\theta) \end{pmatrix}.$$

So we have

$$EZ_i = \xi(\theta_0) + e_i, \tag{1}$$

where e_i are iid with $Ee_i = 0$ and

$$var(e_i) = V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

the true covariance matrix of Z_i. Since we have a correct structure on $var(X)$, $V_{11} = \Sigma(\theta_0)$, but we may not have any structure on V_{12} and V_{22}. Define

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \xi(\theta))^T W_n (Z_i - \xi(\theta)), \tag{2}$$

where W_n is a possibly random weight matrix. Then the estimator $\hat{\theta}_n$ which minimizes $Q_n(\theta)$ will be referred to as a GLS estimator of θ_0. As the stochastic function $Q_n(\theta)$ is the standard quantity to be minimized in a regression model, it is also the objective function we will work with in Section 2. Since

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})^T W_n (Z_i - \bar{Z}) + (\bar{Z} - \xi(\theta))^T W_n (\bar{Z} - \xi(\theta)), \tag{3}$$

and the first term in (3) does not involve θ, the generalized least squares estimator $\hat{\theta}_n$ also minimizes

$$F_n(\theta) = (\bar{Z} - \xi(\theta))^T W_n (\bar{Z} - \xi(\theta)). \tag{4}$$

The equivalence of $F_n(\theta)$ and $Q_n(\theta)$ was formally observed by Shapiro (1986) in the setting of iid Z_i. Note that the equivalence of minimizing $Q_n(\theta)$ and $F_n(\theta)$ holds algebraically even when Z_i are not iid. Consequently, mean and covariance structure analysis can be performed for only independent Z_i with common first four moments. Actually, all the theorems in the next two sections hold if we assume that X_i are independently distributed with common first four moments, and the fifth moments of X_i are uniformly bounded. Since independent random variables with common first four moments are not far from iid, we will only deal with iid X_i because of the theoretical simplicity.

Arminger and Schoenberg (1989) considered modeling the mean and covariance by the pseudo MLE (PMLE) method of Gourieroux, Monfort, and Trognon (1984).

In the setting of iid observations, the assumptions of Gourieroux et al are more than enough. However, Gourieroux et al's assumptions are hard to check. Under a set of much simpler assumptions, we also will consider the statistical properties of the PMLE. Unless specified otherwise, we denote $\dot{h}(\beta) = \partial h/\partial \beta^T$, evaluated at β. In order to get a PMLE estimator, the iteratively reweighted least squares method through a Gauss-Newton algorithm is often used to solve the following equation for $\hat{\theta}_n$,

$$\dot{\xi}^T(\theta)W(\theta)(\bar{Z} - \xi(\theta)) = 0, \tag{5}$$

where

$$W(\theta) = \begin{pmatrix} \Sigma(\theta) & \Delta(\theta) \\ \Delta^T(\theta) & \Psi(\theta) \end{pmatrix}^{-1} \tag{6}$$

and $\Delta(\theta) = cov(X_i, Y_i)$, $\Psi(\theta) = var(Y_i)$ are given by normal theory. When the PML function is not concave, the solution to equation (5) is not unique. Our perspective is different from Gourieroux et al (1984) in that we will show that there is a solution near the true θ_0. Our estimation of standard errors is also different from PMLE. The existence of a root of an equation like (5) was formerly considered by Ferguson (1958). He used the implicit function theorem in proving the existence of the root. The approach we use is different from Ferguson's in that we will use the inverse function theorem to show the existence of a root of (5). Our approach is less involved than that of Ferguson and our assumptions are simpler and easier to check.

When there is no interest in a structured mean, the unknown parameters will include both the mean parameter μ and the structured covariance parameter. The common practice in covariance structure analysis is to use $\bar{X}$ as the estimator for μ_0 and fit the sample covariance S to $\Sigma(\theta)$ by MLE assuming $X \sim N(\mu_0, \Sigma(\theta_0))$. Since $\bar{X}$ and S are independent when X is normal, modeling S by $\Sigma(\theta)$ is the approach of marginal likelihood or conditional likelihood as defined in Cox and Hinkley (1974, p. 17). When X is not normal, the sample mean $\bar{X}$ and S are not independent any more. Modeling S by $\Sigma(\theta)$ using the ADF method only uses marginal information.

Some information will be lost in general by using a summary statistic S though it is hard to say how much information is lost as discussed by Cox and Hinkley (1974, p. 17-18). When X_i is not normal, $\bar{X}$ may not be the most efficient estimator of μ_0 anymore. Even though μ is a nuisance parameter in covariance structure analysis, the efficiency of an estimator for μ can influence the efficiency of the estimator of the structured covariance parameter (Pierce, 1982). Thus, it is tempting to consider modeling the mean and the covariance simultaneously even when we do not have an interest in a structured mean. One way is to treat the mean μ_0 as unknown and let the parameter θ include both μ and the structured covariance parameter. But not knowing μ_0, it is plausible that we can not extract any information about Σ from $\bar{X}$ as commented by Cox and Hinkley (1974, p. 18) in a similar example. Since for nonsymmetric distributions, it is very hard to clarify the above discussion, we will continue this after some empirical evidence at the end of the paper.

We will investigate the consistency and asymptotic normality of both the GLS estimator and the normal theory MLE. A new estimator of the asymptotically correct covariance matrix will also be given for each estimator. Some rigorous proof will be given whenever necessary. Since for iid samples, the main applications are in structured covariances (e.g., factor analysis model) rather than structured mean models, we will develop the general theory for model (1) but emphasize applications in structured covariances with an unstructured mean. We consider the GLS estimator in Section 2 and the normal theory MLE when the data are not normal in Section 3. In Section 4, we discuss a difference between the different GLS weight matrices and give our corrected test statistics. Some empirical performance of the corrected test statistics will be presented in Section 5. Conclusions and remarks will be given at the end of this paper.

2 RESIDUAL-BASED GENERALIZED LEAST SQUARES

In this section, we consider the consistency, asymptotic normality and tests of the GLS estimator which minimizes (2) or (4). Especially, a consistent estimator of V based on the fitted residuals will be given. The advantage of residual-based test statistics will be discussed in Section 4. We need the following assumptions for our results in this paper.

Assumptions:

A1. $\theta_0 \in \Theta$ which is a compact subset of R^q.

A2. $\xi(\theta) = \xi(\theta_0)$ only when $\theta = \theta_0$.

A3. $\xi(\theta)$ is twice continuously differentiable.

A4. $\dot{\xi}(\theta_0)$ is of full rank.

For consistency and asymptotic normality of $\hat{\theta}_n$, we do not require the covariance of Z to be nonsingular.

The following theorem is about the strong consistency of the GLS estimator.

Theorem 1. *Let W_n be a sequence of weight matrices which converges almost surely to W, a positive definite matrix. Under assumptions A1 and A2,* $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

Proof: Since minimizing $Q_n(\theta)$ is equivalent to minimizing $F_n(\theta)$, we will work on $F_n(\theta)$ here. Since $\bar{Z} \xrightarrow{a.s.} \xi(\theta_0)$ by the strong law of large numbers, we have

$$F_n(\theta) \xrightarrow{a.s.} (\xi(\theta_0) - \xi(\theta))^T W(\xi(\theta_0) - \xi(\theta)). \tag{7}$$

Since all the $\hat{\theta}_n$ lie in Θ which is compact, we can choose a subsequence $\hat{\theta}_{n_i}$ which converges to θ'. Since

$$F_{n_i}(\hat{\theta}_{n_i}) \leq F_{n_i}(\theta_0),$$

letting $n_i \to \infty$, we have

$$(\xi(\theta_0) - \xi(\theta'))^T W(\xi(\theta_0) - \xi(\theta')) \leq 0.$$

Since W is positive definite, we must have $\theta' = \theta_0$. So any convergent subsequence converges a.s. to θ_0 and this proves $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.

Theorem 1 tells us that as long as $\xi(\theta)$ is identified, the estimator that minimizes (1) is strongly consistent. We can choose W_n to be the identity matrix, the inverse of the sample covariance of Z_i, or the inverse of the cross products of the fitted residuals assuming the covariance of Z_i is nonsingular. When the mean is unstructured, an identified $\Sigma(\theta)$ will make $\xi(\theta)$ identified.

In order to get the GLS estimator of θ_0, a common practice is to solve the following equation for $\hat{\theta}_n$ by the Gauss-Newton algorithm,

$$\dot{\xi}^T(\theta) W_n(\bar{Z} - \xi(\theta)) = 0. \tag{8}$$

The following theorem is about the asymptotic normality of the GLS estimator which satisfies (8).

Theorem 2. *Assume A3 and A4, if* $W_n \xrightarrow{P} W$ *and* $\hat{\theta}_n \xrightarrow{P} \theta_0$*, then* $\hat{\theta}_n$ *is asymptotically normal with*

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

where $\Omega = A^{-1} \Pi A^{-1}$ *with*

$$A = \dot{\xi}^T(\theta_0) W \dot{\xi}(\theta_0)$$

and

$$\Pi = \dot{\xi}^T(\theta_0) W V W \dot{\xi}(\theta_0).$$

Proof: Since $\hat{\theta}_n$ satisfies (8), using the Taylor expansion on $\xi(\theta)$, we have

$$\dot{\xi}^T(\hat{\theta}_n) W_n \sqrt{n}\bar{e} = \{\dot{\xi}^T(\hat{\theta}_n) W_n \dot{\xi}(\bar{\theta}_n)\} \sqrt{n}(\hat{\theta}_n - \theta_0), \tag{9}$$

where $\bar{e} = \bar{Z} - \xi(\theta_0)$ and $\bar{\theta}_n$ lies between θ_0 and $\hat{\theta}_n$. Since ξ has continuous derivatives and $\sqrt{n}\bar{e} \xrightarrow{\mathcal{L}} N(0, V)$, the theorem follows from (9) and the Slutsky theorem.

When V is full rank, we have the following corollary which is an asymptotic version of the Gauss-Markov theorem.

Corollary 1. *When $W = V^{-1}$, the Ω in Theorem 2 simplifies to*

$$\Omega^{-1} = \dot{\xi}^T(\theta_0) V^{-1} \dot{\xi}(\theta_0)$$

and we get a minimum variance estimator asymptotically among all estimators which satisfy (8).

When μ_0 is unstructured we denote the estimator as $(\hat{\mu}_n, \hat{\theta}_n)$. If V is nonsingular, we have the following corollary.

Corollary 2. *In Theorem 2, let $W = V^{-1}$. If all the third central moments of X are zero, then $\hat{\mu}_n$ and $\hat{\theta}_n$ are asymptotically independent with*

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{\mathcal{L}} N(0, \Sigma^{-1}(\theta_0))$$

and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Omega_{22}),$$

where

$$\Omega_{22}^{-1} = \dot{\sigma}^T(\theta)B^{-1}\dot{\sigma}(\theta),$$

and $B = V_{22} - V_{21}^T\Sigma^{-1}V_{12}$.
Moreover, when the 4*th central moments of* X *satisfy*

$$\sigma_{ijkl} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk},$$

then Ω_{22}^{-1} *has a simplier form*

$$\Omega_{22}^{-1} = \frac{1}{2}\dot{\sigma}_a^T(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\sigma}_a(\theta), \qquad (10)$$

where $\sigma_a(\theta) = vec(\Sigma(\theta))$ *and all the matrix functions are evaluated at* θ_0.

Since the proof of the above corollary is not so interesting and involves a lot of algebraic operations, we give it in the appendix.

Corollary 2 tells us that if all the third central moments of X are zero, we can model μ and Σ separately and do not lose any information asymptotically. This is the case with the multivariate elliptically symmetric distribution (Fang, Kotz, and Ng, 1990; Shapiro and Browne, 1987).

Since the asymptotic covariance matrix of $\hat{\theta}_n$ in Theorem 2 involves an unknown matrix V, we need a consistent estimator of it in order to do some tests or to compute the standard errors. Further, according to Corollary 1, if we choose a proper weight matrix W_n in Theorem 2, we can get a more efficient estimator. An obvious estimator of V is the sample covariance $S_z = \frac{1}{n}\sum_{i=1}^n(Z_i - \bar{Z})(Z_i - \bar{Z})^T$. An asymptotically equivalent one is the cross product of the fitted residuals. In the context of regression, estimating the variance and covariance matrix through residuals has been used extensively. It can also be used in mean and covariance structure analysis.

Theorem 3. *If $\xi(\theta)$ is a continuous matrix function and $\hat{\theta}_n$ is strongly consistent, then*

$$\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \xi(\hat{\theta}_n))(Z_i - \xi(\hat{\theta}_n))^T \qquad (11)$$

is a strongly consistent estimator of V.

Proof:

$$\begin{aligned} \hat{V}_n &= \frac{1}{n}\sum_{i=1}^{n} e_i e_i^T + (\xi(\theta_0) - \xi(\hat{\theta}_n))\bar{e}^T + \bar{e}(\xi(\theta_0) - \xi(\hat{\theta}_n))^T \\ &+ (\xi(\hat{\theta}_n) - \xi(\theta_0))(\xi(\hat{\theta}_n) - \xi(\theta_0))^T. \qquad (12) \end{aligned}$$

Since ξ is continuous, the last three terms in (12) approach zero. The theorem follows by the strong law of large numbers.

From the above three theorems, we can use a two stage estimating process. First, use least squares, for example, to get a consistent estimator of θ_0. Then, using $W_n = \hat{V}_n^{-1}$ in (11) as the weight matrix in Theorem 2, the corresponding updated estimator will be most efficient asymptotically according to Corollary 1. This is a type of linearized improvement estimator (Bentler, 1983b; Bentler and Dijkstra, 1985). For small to medium sample sizes, it is known that the efficiency of an estimated weight matrix influences the efficiency of the mean parameter (Carroll, Wu, and Ruppert, 1988). For linear regression models, if starting with least squares, Carroll et al recommend repeating this process at least twice. Their recommendation also applies to our estimator in Theorem 2.

Next we propose a test for the overall fit of the model. This test requires either the weight matrix S_z^{-1} or $\hat{V}_n^{-1}$. We will discuss the difference between $\hat{V}_n^{-1}$ and S_z^{-1} in detail in Section 4. The following lemma will simplify the proofs regarding test statistics.

Lemma 1. *For the* $\hat{\theta}_n$ *in Theorem 2, we have*

$$\sqrt{n}(\bar{Z} - \xi(\hat{\theta}_n)) = \{I - \dot{\xi}(\theta_0)[\dot{\xi}^T(\theta_0)W\dot{\xi}(\theta_0)]^{-1}\dot{\xi}^T(\theta_0)W\}\sqrt{n}\bar{e} + o_p(1).$$

Proof: Using the Taylor expansion on $\xi(\theta)$, we have

$$\xi(\hat{\theta}_n) = \xi(\theta_0) + \dot{\xi}(\theta_0)(\hat{\theta}_n - \theta_0) + O_p(\frac{1}{n}). \tag{13}$$

From (9) we have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = [\dot{\xi}^T(\theta_0)W\dot{\xi}(\theta_0)]^{-1}\dot{\xi}^T(\theta_0)W\sqrt{n}\bar{e} + o_p(1). \tag{14}$$

The lemma follows from (13), (14) and

$$\sqrt{n}(\bar{Z} - \xi(\hat{\theta}_n)) = \sqrt{n}(\bar{Z} - \xi(\theta_0)) - \sqrt{n}(\bar{Z} - \xi(\hat{\theta}_n)) + O_p(\frac{1}{\sqrt{n}}).$$

Theorem 4. *Under assumptions A1 to A4, if* $W_n \xrightarrow{P} W$*, then*

$$nF_n(\hat{\theta}_n) = nQ_n(\hat{\theta}_n) - nQ_n(\bar{Z}) \xrightarrow{\mathcal{L}} \sum_k \lambda_k \chi_1^2,$$

where $Q_n(\bar{Z})$ *denotes the fitted index of the unstructed mean and covariance and the* $\lambda_k' s$ *are the nonzero eigenvalues of* $V^{\frac{1}{2}}W^{\frac{1}{2}}MW^{\frac{1}{2}}V^{\frac{1}{2}}$ *with*

$$M = I - W^{\frac{1}{2}}\dot{\xi}(\theta_0)[\dot{\xi}^T(\theta_0)W\dot{\xi}(\theta_0)]^{-1}\dot{\xi}^T(\theta_0)W^{\frac{1}{2}}.$$

Further more, if V *is nonsingular and* $W = V^{-1}$*, then*

$$nF_n(\hat{\theta}_n) = nQ_n(\hat{\theta}_n) - nQ_n(\bar{Z}) \xrightarrow{\mathcal{L}} \chi^2_{p+p^*-q},$$

where $p* = p(p+1)/2$.

Proof: From Lemma 1, we have

$$\begin{aligned} nF_n(\hat{\theta}_n) &\xrightarrow{\mathcal{L}} U^T V^{\frac{1}{2}} W^{\frac{1}{2}} M W^{\frac{1}{2}} V^{\frac{1}{2}} U \\ &= \sum_k \lambda_k \chi_1^2, \end{aligned} \tag{15}$$

where $U \sim N(0, I)$. Furthermore, when $W = V^{-1}$, $V^{\frac{1}{2}} W^{\frac{1}{2}} M W^{\frac{1}{2}} V^{\frac{1}{2}} = M$ which is a projection matrix of rank $p + p^* - q$, the theorem follows.

Theorem 4 gives us a way to test the general fit of the hypothetical structure. Note that when W does not equal V^{-1}, $\hat{\theta}_n$ will not be asymptotically efficient. Then the distribution of $nQ_n(\hat{\mu}_n, \hat{\theta}_n) - nQ_n(\bar{X}, s)$ can be approximated by $\alpha\chi_r^2$, where r is the rank of $V^{\frac{1}{2}} W^{\frac{1}{2}} M W^{\frac{1}{2}} V^{\frac{1}{2}}$ and $r\alpha = tr(V^{\frac{1}{2}} W^{\frac{1}{2}} M W^{\frac{1}{2}} V^{\frac{1}{2}})$. Details of such an approximation can be found in Satorra and Bentler (1988, 1994). It works well in covariance structure practice (Chou, Bentler, and Satorra, 1991; Hu et al, 1992). A second alternative is to use an approximation to the distribution of a mixture of χ_1^2 variates, as proposed by Bentler (1994). The success of such a testing procedure will depend on the quality of the approximation. A third test alternative is to extend Browne's (1984, Proposition 4) residual covariance test to mean and covariance structure analysis. To implement it, we need a consistent estimator $\tilde{V}_n$ of V.

Corollary 3. *Under assumptions A1 to A4, if* $\tilde{V}_n \xrightarrow{P} V$, *then*

$$n(\bar{Z} - \xi(\hat{\theta}_n))^T \dot{\xi}_c(\hat{\theta}_n) \{\dot{\xi}_c^T(\hat{\theta}_n) \tilde{V}_n \dot{\xi}_c(\hat{\theta}_n)\}^{-1} \dot{\xi}_c^T(\hat{\theta}_n)(\bar{Z} - \xi(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} \chi^2_{p+p^*-q},$$

where $\dot{\xi}_c(\hat{\theta}_n)$ *is a* $(p + p^*) \times (p + p^* - q)$ *matrix of full column rank with columns that are orthogonal to* $\dot{\xi}(\hat{\theta}_n)$.

Proof: From Lemma 1, we have

$$\sqrt{n}(\bar{Z} - \xi(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} (0, \Xi),$$

where $\Xi = (I - H)V(I - H)^T$ with

$$H = I - \dot{\xi}(\theta_0)[\dot{\xi}^T(\theta_0)W\dot{\xi}(\theta_0)]^{-1}\dot{\xi}^T(\theta_0)W,$$

so

$$\sqrt{n}\dot{\xi}_c^T(\hat{\theta}_n)(\bar{Z} - \xi(\hat{\theta}_n)) \xrightarrow{\mathcal{L}} N(0, \dot{\xi}_c^T(\theta_0)V\dot{\xi}_c(\theta_0)). \tag{16}$$

The corollary follows.

Even though $\hat{\theta}_n$ does not need to be most efficient in Corollary 3, it still requires a consistent estimator of V. As in Theorem 4, both S_z and $\hat{V}_n$ can be used in place of $\tilde{V}_n$ under arbitrary distributions. Browne's test and Satorra's (1992) extension to mean and covariance structures were based on S_z. The relation between the resulting tests will be discussed in detail in Section 4. As noted by Bentler (1989) and Satorra (1992), if the distribution is known to be normal, a normal theory estimator $\tilde{V}_n$ can be used instead. The χ^2 test in Theorem 4 can be invoked by a two step procedure if $W \neq V^{-1}$. For example, if we start with a least squares fit, we can use $\hat{V}_n^{-1}$ as a weight matrix and update $\hat{\theta}_n$, then $nF(\hat{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2_{p+p^*-q}$.

3 NORMAL THEORY MLE WHEN DATA ARE NOT NORMAL

In this section, we will consider the behavior of normal theory MLE when the data are not normal. As in the last section, we suppose that the mean and covariance structure $\xi_0 = \xi(\theta_0)$ is correct and we model Z_i in (1) as a nonlinear regression model. We also need to assume that $\Sigma(\theta_0)$ is positive definite in this whole section. We need some preparations first.

Let $B_r(x_0)$ be a ball of radius r with center at x_0. The following lemma is a modified version of the fundamental inverse function theorem (Rudin, 1976, p. 221).

Lemma 2. *Let $f(x)$ be a continuously differentiable mapping from R^p to R^p. Let A be a nonsingular $p \times p$ matrix and $\kappa = \frac{1}{2}||A^{-1}||^{-1}$. If $B_r(x_0)$ is a ball on which*

$$||\dot{f}(x) - A|| < \kappa,$$

then $f(B_r(x_0))$ contains the ball $B_{\kappa r}(f(x_0))$.

Since both $\Delta(\theta)$ and $\Psi(\theta)$ in (6) are functions of $\mu(\theta)$ and $\Sigma(\theta)$, when $\Sigma(\theta_0)$ is nonsingular, we can check by a tedious verification that $W(\theta_0)$ defined in (6) exists and is positive definite. Now let

$$g_n(\theta) = \dot{\xi}^T(\theta)W(\theta)(\bar{Z} - \xi(\theta)), \tag{17}$$

we have

$$\dot{g}_n(\theta) = -\dot{\xi}^T(\theta)W(\theta)\dot{\xi}(\theta) + \{\frac{\partial}{\partial\theta^T}[\dot{\xi}^T(\theta)W(\theta)]\}(\bar{Z} - \xi(\theta)).$$

Let

$$g(\theta) = \dot{\xi}^T(\theta)W(\theta)(\xi(\theta_0) - \xi(\theta)), \tag{18}$$

then we have

$$\dot{g}(\theta) = -\dot{\xi}^T(\theta)W(\theta)\dot{\xi}(\theta) + \{\frac{\partial}{\partial\theta^T}[\dot{\xi}^T(\theta)W(\theta)]\}(\xi(\theta_0) - \xi(\theta))$$

and both

$$g_n(\theta) \xrightarrow{a.s.} g(\theta),$$

and

$$\dot{g}_n(\theta) \xrightarrow{a.s.} \dot{g}(\theta)$$

uniformly on Θ.

Theorem 5. *Under assumptions A3 and A4, with probability 1 there is a $r > 0$ such that $g_n(\theta)$ has a zero point in $B_r(\theta_0)$ for all n sufficiently large.*

Proof: Let $A = \dot{g}(\theta_0)$ which is nonsingular and $\kappa = \frac{1}{2}||A^{-1}||^{-1}$. Since $\dot{g}_n(\theta)$ converges to $\dot{g}(\theta)$ uniformly on Θ, there exist positive numbers N_1 and r such that for all $n > N_1$

$$\begin{aligned} ||\dot{g}_n(\theta) - A|| &\leq ||\dot{g}_n(\theta) - \dot{g}(\theta)|| + ||\dot{g}(\theta) - \dot{g}(\theta_0)|| \\ &< \kappa, \quad \theta \in B_r(\theta_0). \end{aligned}$$

Applying Lemma 2 to $g_n(\theta)$, it follows that $g_n(B_r(\theta_0))$ contains a ball $B_{\kappa r}(g_n(\theta_0))$ for all $n > N_1$. Since $g_n(\theta_0) \xrightarrow{a.s.} 0$, there exists a number N_2 such that $||g_n(\theta_0)|| < \kappa r$ for all $n > N_2$. Let $N = \max(N_1, N_2)$, we have $0 \in B_{\kappa r}(g_n(\theta_0))$ for all $n > N$. Thus $0 \in B_{\kappa r}(g_n(\theta_0))$ and there is a zero point of g_n in $B_r(\theta_0)$ for all $n > N$.

When the mean is unstructured, then

$$\dot{\xi}(\mu, \theta) = \begin{pmatrix} I & 0 \\ \dot{\tau}(\mu) & \dot{\sigma}(\theta) \end{pmatrix}.$$

Since

$$\begin{pmatrix} I & \dot{\tau}^T(\mu) \\ 0 & \dot{\sigma}^T(\theta) \end{pmatrix} \begin{pmatrix} \Sigma(\theta) & \Delta(\theta) \\ \Delta^T(\theta) & \Psi(\theta) \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma^{-1} & 0 \\ -\dot{\sigma}^T(\theta)B^{-1}\Delta^T\Sigma^{-1} & \dot{\sigma}^T(\theta)B^{-1} \end{pmatrix},$$

(5) is equivalent to

$$\begin{cases} \hat{\mu}_n = \bar{X} \\ \dot{\sigma}_a^T(\theta)(\Sigma^{-1}(\theta) \otimes \Sigma^{-1}(\theta))(vec(S) - \sigma_a(\theta)) = 0. \end{cases} \tag{19}$$

From Theorem 5 and (19), we have the following corollary.

Corollary 4. *Under assumptions A3 and A4, with probability 1 there is a $r > 0$ such that the second equation in (19) has a solution in $B_r(\theta_0)$ for all n sufficiently* *large.*

The following theorem is about the strong consistency of $\hat{\theta}_n$ which satisfies (5).

Theorem 6. *Under assumptions A2, A3 and A4, if $\hat{\theta}_n$ satisfies $g_n(\hat{\theta}_n) = 0$ and is in $B_r(\theta_0)$, then $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.*

Proof: Since $\dot{g}(\theta)$ is nonsingular in $B_r(\theta_0)$, the function $g(\theta)$ has a unique zero point of θ_0 in $B_r(\theta_0)$. Now let $\hat{\theta}_{n_i}$ be any converged subsequence of $\hat{\theta}_n$ with limit $\theta' \in B_r(\theta_0)$, we have $g_{n_i}(\hat{\theta}_{n_i}) = 0$. Let $n_i \to \infty$ we obtain $g(\theta') = 0$. Since $g(\theta)$ has a unique zero point of θ_0, we have $\theta' = \theta_0$ and the theorem follows.

When $\mu(\theta) = \mu$, the unstructured mean, we have the following corollary.

Corollary 5. *Under assumptions A3 and A4, if $\hat{\theta}_n$ satisfies the second equation in (19) and is in $B_r(\theta_0)$, then $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$.*

The following theorem is about the asymptotic distribution of $\hat{\theta}_n$.

Theorem 7. *Under assumptions A3 and A4, if $\hat{\theta}_n \xrightarrow{P} \theta_0$, then*

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

where $\Omega = D^{-1}GD^{-1}$ with

$$D = \dot{\mu}^T(\theta)\Sigma^{-1}\dot{\mu}(\theta) + \frac{1}{2}\dot{\sigma}_a^T(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\sigma}_a(\theta),$$

$$\begin{aligned} G &= \dot{\mu}^T(\theta)\Sigma^{-1}\dot{\mu}(\theta) + \dot{\mu}^T(\theta)(\Sigma^{-1}V_{12} - \dot{\tau}^T(\mu))B^{-1}\dot{\sigma}(\theta) \\ &+ \dot{\sigma}^T(\theta)B^{-1}(V_{21}\Sigma^{-1} - \dot{\tau}(\mu))\dot{\mu}(\theta) + \dot{\sigma}^T(\theta)B^{-1}\dot{\sigma}(\theta), \end{aligned}$$

$B = (b_{kl,st})$ with $b_{kl,st} = \sigma_{ks}\sigma_{lt} + \sigma_{kt}\sigma_{ls}$ and all the matrix functions are evaluated at θ_0.

Proof: Since $\hat{\theta}_n$ satisfy $g_n(\hat{\theta}_n) = 0$, using the Taylor expansion we have

$$g_n(\theta_0) + \dot{g}_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0) = 0,$$

where $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0. Since

$$\sqrt{n} g_n(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Pi)$$

with

$$\Pi = \dot{\xi}^T(\theta_0) W(\theta_0) V W(\theta_0) \dot{\xi}(\theta_0)$$

and $\dot{\xi}(\theta)$ is continuous, we have

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

with $\Omega^{-1} = A^{-1} \Pi A^{-1}$, where

$$A = \dot{\xi}^T(\theta_0) W(\theta_0) \dot{\xi}(\theta_0).$$

Since $W(\theta)$ is given by the normal theory, from Corollary 2, we have

$$A = (\dot{\mu}^T(\theta), I) \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \frac{1}{2}\dot{\sigma}_a^T(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\sigma}_a(\theta) \end{pmatrix} \begin{pmatrix} \dot{\mu}(\theta) \\ I \end{pmatrix} = D.$$

So we only need to show that $\Pi = G$, and this is just a simplification with some algebra.

When there is no structure on the mean μ, we have $\hat{\mu}_n = \bar{X}$. From Theorem 7, we have the following corollary.

Corollary 6. *Under assumptions A3 and A4, if* $\hat{\theta}_n \xrightarrow{P} \theta_0$, *then*

$$\sqrt{n} \begin{pmatrix} \bar{X} - \mu_0 \\ \hat{\theta}_n - \theta_0 \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Omega),$$

where $\Omega = D^{-1} G D^{-1}$ *with*

$$D = \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \frac{1}{2}\dot{\sigma}_a^T(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\sigma}_a(\theta) \end{pmatrix}$$

and

$$G = \begin{pmatrix} \Sigma^{-1} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where

$$G_{12} = G_{21}^T = (\Sigma^{-1}V_{12} - \dot{\tau}^T(\mu))B^{-1}\dot{\sigma}(\theta),$$

$$G_{22} = \dot{\sigma}^T(\theta)B^{-1}(V_{22} - \Delta^T\Sigma^{-1}V_{12} - V_{21}\Sigma^{-1}\Delta + \Delta^T\Sigma^{-1}\Delta)B^{-1}\dot{\sigma}(\theta)$$

$B = (b_{kl,st})$ *with* $b_{kl,st} = \sigma_{ks}\sigma_{lt} + \sigma_{kt}\sigma_{ls}$, *and all the matrix functions are evaluated at* θ_0.

Note that when

$$(V_{22} - \Delta^T\Sigma^{-1}V_{12} - V_{21}\Sigma^{-1}\Delta + \Delta^T\Sigma^{-1}\Delta) < B,$$

the normal theory MLE overestimates the standard errors of $\hat{\theta}_n$, otherwise, MLE underestimates the standard errors. A consistent estimator of Ω is to put $\hat{\mu}_n = \bar{X}$ and $\hat{\theta}_n$ into the structured parameters and use the result in Theorem 3 to estimate V. Arminger and Schoenberg (1989, p. 414) stated that "Without loss of generality we assume that $\mu(\theta_0) = 0$", then they proceeded in the context of $\mu(\theta_0) = 0$ to get a simplification of the standard errors of $\hat{\theta}_n$ in covariance structure analysis. Their result is not true in general. In practice, even if the mean is zero, the finite sample efficiency will be different when treating $\mu(\theta_0) = 0$. From Corollary 6, we can see that $\bar{X}$ and $\hat{\theta}_n$ are generally dependent in normal theory covariance structure analysis. If all the third central moment of X are zero, then $\bar{X}$ and $\hat{\theta}_n$ are independent asymptotically, as with elliptically contoured distributions with an unstructured mean. In this case, we may model the mean and covariances separately for computational covenience.

When the underlying distribution is not normal, the normal theory LRT will give incorrect inference generally. However, the test statistic in Corollary 3 will behave correctly asymptotically and can be used to evaluate the model. Finally, we note that there are specialized independence and model conditions under which, asymptotically, some parameters in a linear structure can be estimated efficiently, some standard errors obtained correctly, and the model null hypothesis tested using normal theory statistics even if the data are not normal. See Satorra and Neudecker (1994) for a recent contribution and further references to asymptotic robustness theory.

4 SEVERAL CORRECTED TEST STATISTICS

Our intention in this paper is to present a unified approach to mean and covariance structure analysis through regression. This point of view implies that the inverse of the cross products of the residuals, which has been used extensively in the regression literature, should also be considered for use in covariance structure analysis. Here we discuss further implications of this viewpoint. In the ADF test statistic, the inverse of the sample covariance S_z of Z_i is used to get $\hat{\theta}_n$ and the corresponding test statistic is $nF_n(\hat{\theta}_n)$. Since $\hat{V}_n$ is also a consistent estimator of V from Theorem 3, we can also use the inverse of $\hat{V}_n$ in estimating $\hat{\theta}_n$ and get a corresponding test statistic through a two stage estimation process. By an ANOVA decomposition, we have

$$\hat{V}_n = S_z + (\bar{Z} - \xi(\hat{\theta}_n))(\bar{Z} - \xi(\hat{\theta}_n))^T. \tag{20}$$

Even though S_z and $\hat{V}_n$ are asymptotically equivalent, from (20) it follows that $\hat{V}_n \geq S_z$. Consequently, we expect that S_z and $\hat{V}_n$ should have different effects on test statistics. From (20) we also have

$$\hat{V}_n^{-1} = S_z^{-1} - \frac{S_z^{-1}(\bar{Z} - \xi(\hat{\theta}_n))(\bar{Z} - \xi(\hat{\theta}_n))^T S_z^{-1}}{1 + (\bar{Z} - \xi(\hat{\theta}_n))^T S_z^{-1}(\bar{Z} - \xi(\hat{\theta}_n))}. \tag{21}$$

So the estimator $\hat{\theta}_n$ which satisfies

$$\dot{\xi}^T(\hat{\theta}_n) S_z^{-1}(\bar{Z} - \xi(\hat{\theta}_n)) = 0 \tag{22}$$

also satisfies

$$\dot{\xi}^T(\hat{\theta}_n)\hat{V}_n^{-1}(\bar{Z} - \xi(\hat{\theta}_n)) = 0. \tag{23}$$

As a result, the two stage estimation process is uncessary if we start with S_z^{-1} as a weight matrix. Let $F_n(\hat{\theta}_n)$ denote the minimized function using S_z^{-1} as the weight matrix and $T_1 = nF_n(\hat{\theta}_n)$ be the corresponding test statistic. In the context of covariance structures with an unstructured mean, T_1 would be the ADF test statistic, but in our regression context it is not the same. Then, from (21), a corrected test statistic corresponding to the weight matrix $\hat{V}_n^{-1}$ is

$$T_2 = \frac{T_1}{1 + F_n(\hat{\theta}_n)}. \tag{24}$$

As we mentioned earlier, the ADF statistic in covariance structure analysis has been found to reject the null hypothesis exceptionally frequently in small to medium sample sizes (Hu et al, 1992). Possibly the same result can occur with T_1. However, since $T_2 < T_1$ generally, we expect T_2 to behave better in small to medium sample sizes. From

$$nF_n(\hat{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2_d,$$

we have

$$F_n(\hat{\theta}_n) = O_p(\frac{1}{n}).$$

So we can write

$$T_2 = (1 - F_n)T_1 + O_p(\frac{1}{n^2}). \tag{25}$$

Comparing (25) with Bartlett-type corrections for LRT statistics, the correction term $\frac{1}{1+F_n}$ represents a correction to the ADF type test statistic T_1. The difference between a standard Bartlett-type correction for LRT statistics and $\frac{1}{1+F_n}$ for T_1 is that the Bartlett-type correction shifts the LRT statistics towards zero by a positive factor of order $O(\frac{1}{n})$ while $\frac{1}{1+F_n}$ shifts T_1 towards zero by a positive factor of order $O_p(\frac{1}{n})$. If we use the inverse of the cross products of the fitted residuals from an ADF fitting as a weight matrix, the correction is automatic.

Since $T_2 < T_1$ generally, it is possible that T_2 may lose some power as a test statistic. We will show that asymptotically T_2 has the same power as that of T_1 against alternatives. When the null hypothesis is not true, then $\xi_0 = E\bar{Z} \neq \xi(\theta)$ for any $\theta \in \Theta$. Let

$$\delta(\theta) = (\xi_0 - \xi(\theta))^T W (\xi_0 - \xi(\theta))$$

and θ^* minizes $\delta(\theta)$ on Θ. We can rewrite (7) as

$$F_n(\theta) \xrightarrow{a.s.} \delta(\theta).$$

Exactly the same argument as in Theorem 1 shows that $\hat{\theta}_n \xrightarrow{a.s.} \theta^*$. So

$$F_n(\hat{\theta}_n) = \delta(\theta^*) + o_p(1),$$

and

$$T_2 = \frac{nF_n(\hat{\theta}_n)}{1 + F_n(\hat{\theta}_n)} \xrightarrow{P} \infty.$$

If we assume that

$$\xi(\theta^*) = \xi_0 + \frac{\delta_0}{\sqrt{n}}, \tag{26}$$

which is a standard condition for considering the power of a test statistic, then $\xi(\hat{\theta}_n) \xrightarrow{a.s.} \xi_0$ and $F_n(\hat{\theta}_n) = O_p(\frac{1}{n})$. Since T_1 is a noncentral chi-square variate under (26), from (25), T_2 is also a noncentral chi-square variate with the same noncentrality parameter and degrees of freedom. So T_2 has exactly the same behavior as T_1 asymptotically.

With a similar empirical behavior to that of T_1, the test statistic in Corollary 3 based on S_z also rejects the null hypothesis too often (e.g., Chan, 1995). Let $T(\tilde{V}_n)$ denote the test statistic in Corollary 3. From (20), we have

$$T(\hat{V}_n) = \frac{T(S_z)}{1 + T(S_z)/n}. \tag{27}$$

So $T(\hat{V}_n)$ represents a correction to $T(S_z)$ and the merit of $T(\hat{V}_n)$ comparing with $T(S_z)$ is exactly the same as that of T_2 comparing with T_1.

We have discussed the correction for ADF type statistics by modeling the mean and covariance simultaneously as a regression model. The correction factor in (24) and (27) also can be applied to only covariance structure analysis. Notice that since the covariance structure analysis of modeling S by $\Sigma(\theta)$ is equivalent to modeling $(X_i - \bar{X})(X_i - \bar{X})^T$ by $\Sigma(\theta)$, we can use the fitted residuals $vech\{(X_i - \bar{X})(X_i - \bar{X})^T\} - vech(\Sigma(\hat{\theta}_n))$ and get a corresponding residual weight matrix. Exactly the same algebra shows that using the weight from the residuals corresponds to the correction (24) and (27) of the ADF statistic of Browne (1982, 1984) and Chamberlain (1982) and the statistic in Proposition 4 of Browne (1984).

5 EMPIRICAL PERFORMANCE OF THE CORRECTED STATISTICS

We have presented our corrected statistics in last section and discussed their merits from a theoretical point of view. In order to see the practical effect of our correction, a small scale simulation was performed. The model is the same as the one used by Hu et al (1992), i.e. a 3-factor model with each factor having its own 5-indicators. Since we do not put any structure on the mean vector, we have $p + p^* = 135$ and $q = 48$. So the degrees of freedom of the chi-square statistic in Theorem 4 is 87. The observed variables X_i were generated under two conditions. In the first condition, both the common factors and the unique factors are normal, so $X_i \sim N(\mu, \Sigma)$. In the second condition, the common factors are still normal, but the unique factors are independent lognormal variates so the skewnesses of the observed X_i are not zero anymore. We chose sample sizes 150 to 1000 for each condition. 500 simulation replications were performed for a given sample size. We estimated the model in two ways: first, a covariance structure model only with ADF estimation, and second, as a regression model with an unstructured mean and a structured covariance with an optimal weight matrix. Thus we study asymptotically efficient estimators only. The results are summarized in Table 1 and Table 2, where ADF is the covariance structure test statistic of Browne (1984); CADF represent the corrected ADF test statistic (24) which is equivalent to using the inverse of the cross products of the fitted residuals as a weight matrix. With the same notation as in Hu et al (1992), M and SD represent the means and the standard deviations, respectively, of the empirical test statistics across the 500 replications. The Freq represents the rejection frequency of the empirical test statistics using the 95% percentile of the χ^2_{87}.

Table 1

Emprical Behavior of the Correction on Test Statistics

$F \sim Normal(0, \Phi)$, $E \sim Normal(0, \Psi)$

		Sample Size				
Method		150	200	300	500	1000
ADF:						
	M	217.43	165.80	130.01	110.09	97.15
	SD	46.16	33.06	22.83	18.99	15.84
	Freq	444/445	482/495	415	241	101
CADF:						
	M	87.81	89.78	90.13	89.83	88.36
	SD	7.72	9.76	10.93	12.60	13.08
	Freq	0/445	11/495	20	32	35
T_1:						
	M	203.82	162.18	129.04	109.87	97.10
	SD	40.53	31.20	22.34	18.91	15.83
	Freq	422/424	479/497	411	237	100
T_2:						
	M	85.58	88.75	89.67	89.69	88.32
	SD	7.29	9.43	10.76	12.55	13.07
	Freq	0/424	6/497	19	30	35

Table 2

Emprical Behavior of the Correction on Test Statistics

$F \sim Normal(0, \Phi)$, $E \sim Lognormal(0, \Psi)$

		Sample Size				
Method		150	200	300	500	1000
ADF:						
	M	211.74	158.06	125.29	107.54	97.97
	SD	40.17	28.46	19.81	16.76	14.21
	Freq	471/472	484/498	391	203	92
CADF:						
	M	87.04	87.60	87.93	88.19	89.07
	SD	6.92	8.75	9.79	11.28	11.75
	Freq	0/472	1/498	7	14	26
T_1:						
	M	199.99	153.94	124.52	107.37	97.94
	SD	35.28	26.23	19.45	16.63	14.22
	Freq	453/455	481/495	388	203	91
T_2:						
	M	85.06	86.38	87.56	88.08	89.05
	SD	6.54	8.25	9.66	11.20	11.77
	Freq	0/455	1/495	6	15	26

From the results in Table 1 and Table 2, we can see that the ADF method is essentially unusable at any of the sample sizes studied. At $n = 150$, the procedure only converges 445/500 or 472/500 times, and in all but one converged solution the true model is rejected. At $n = 300$, about 80% of true models are rejected. At $n = 1000$, the procedure works better, but about 20% are still rejected. Hu et al (1992) showed that the method behaves as expected at $n = 5000$. On the other hand, from the row labeled CADF, we see that the statistic using the inverse of the cross products of the fitted residuals as a weight matrix gives a great improvement over the ADF statistic. As expected, the correction is larger for smaller sample sizes and tends to be smaller as sample size becomes larger. When sample size is small, the rejection rate for CADF is usually less than .05. It seems that the statistic over-corrects. But, considering that the test statistic is only approximately chi-square distributed for a given sample size, and that an ideal test statistic would accept the model in all samples when the underlying null hypothesis is true, it is really an advantage rather than a flaw for us to use the corrected statistic in practice. The standard deviation of χ^2_{87} is $\sqrt{2 \times 87} \approx 13.19$. At small sample sizes, the empirical standard deviation of the test statistics is generally smaller than expected. Considering that the CADF means are around 87 at all sample sizes, small standard errors are also much better than larger standard errors with means much larger than 87. Notice that the Bartlett-type correction for a LRT statistic is based on correcting the mean of the statistic (e.g., Stuart and Ord, 1991, §23.9), making the mean of the corrected statistic nearer to the degrees of freedom of a chi-square. Our correction seems also to mainly correct the means.

The regression test statistic T_1 performs virtually the same as the ADF test statistic. That is, T_1 also greatly rejects the true model as can be seen in Tables 1 and 2. In fact, we can see that the differences between ADF and T_1 disappear when the sample size gets larger. Even though $\bar{X}$ and S are not asymptotically independent, using the marginal information S for estimating θ_0 when there is only a covariance structure does not lose information for this specific example. This is not surprising. As commented by Cox and Hinkley (1974, p. 18), before knowing μ_0, we can not extract any information from $\bar{X}$ about θ_0 when μ is unstructed. This phenomenon may occur in general for other skewed data.

On the other hand, the corrected regression statistic T_2 based on our residual GLS procedure, implemented via (24), also performs excellently. Its performance is virtually the same as that of the CADF. Hence, while our regression approach has yielded substantially enhanced test statistic performance based on residual weight matrices, in the unstructured mean case we have not found evidence that modeling the means yields any improvement in performance. Of course, in models with structured means $\mu = \mu(\theta_0)$, modeling the means can not be avoided. We would expect our regression approach using residual weight matrices also to outperform the classical approach in small samples for the reasons noted above.

6 CONCLUSIONS

Our approach yields a variety of estimators and tests depending on the initial consistent estimator chosen and the second-stage weight matrix used for final estimates and tests. Perhaps our most interesting result is that the two-stage approach can be avoided completely. Similarly, although we have emphasized the importance of residual weight matrices, it may not be necessary to compute such matrices. The estimators that result from the use of GLS weight matrices and residual-based weight matrices are numerically equal, and the test statistics that would be obtained from residual-based weight matrices can be computed as simple Bartlett-type corrections to the GLS statistics. Our small simulation study has shown that these corrected statistics work remarkably well. Although we have emphasized arbitrary distribution theory, our corrections also apply to data of any known distributional forms for which the appropriate GLS weight matrices are used.

It is obvious, from our discussion in Section 4, that T_2 will reject the null hypothesis for certain departing alternatives even when sample size is small. It is possible that for a very small misspecification of the model and a small sample size, T_2 may not be as powerful as T_1 against alternatives. This also occurs with standard Bartlett-type corrections. From our experience, T_1 almost always rejects alternatives for all sample sizes. Also, every model is probably wrong in practice and there is some consensus that a model need not to be totally correct before it becomes useful (e.g, Box, 1979). The possible leniency in T_2 for a very small misspecification may be an advantage for accepting a useful but not perfect model in practice.

We have discussed the effect of the correction in Section 4. Now let us look at it from another point of view. When the null hypothesis is not true and the departure is small, we have $ET_1 = nF_{10} = d + \delta$ where d is the degrees of freedom of the model and δ is the noncentrality parameter of χ^2_d distribution. So the multiplier $1/(1 + F_{10}) = n/(n + d + \delta)$. This point of view also is interesting because it gives us another rationale for our correction eq. (24). Suppose we decided to base our correction on $n/(n + d + \hat{\delta})$ using an estimated noncentrality parameter. Then if we take as the estimator $\hat{\delta} = (n\hat{F}_1 - d)$, we obtain our correction. From this point of view we also could consider other alternatives. First, we could use other noncentrality parameter estimators. For example, if we used $\max\{(n\hat{F}_1 - d), 0\}$, our correction would be modified when $T_1 < d$. This may be useful since the variance of this corrected statistic would be greater than that of T_2 or CADF, which, according to Tables 1 and 2, tends to be too small in small samples. Also, since under the null hypothesis $\delta = 0$, we immediately get the correction $n/(n+d)$. Unlike eq. (24), the latter does not depend on model fit. This is an attractive feature, but since the variance of such a corrected statistic would be the same as variance of the uncorrected statistic, this variance would be substantially too large as shown in Tables 1 and 2. Hence, this correction probably would not work well. Further, these alternatives do not have the virtue of arising precisely from our residual GLS approach.

Although we have not discussed the standard error estimators obtained from our approach, it is a simple matter to show algebraically that the residual-based standard error estimator is identical to its typical ADF estimator. That is $\{\dot{\xi}^T(\hat{\theta}_n)S_z^{-1}\dot{\xi}(\hat{\theta}_n)\}^{-1} = \{\dot{\xi}^T(\hat{\theta}_n)\hat{V}_n^{-1}\dot{\xi}(\hat{\theta}_n)\}^{-1}$. However, since empirical evidence shows that ADF standard error estimates substantially underestimate the empirical sampling variability of the estimators at smaller sample sizes (Henly, 1993; West et al 1995), it would be desirable to find a way to correct the standard errors as well. One way to do this is to recognize from (24) that the divisor $(1 + F_n(\hat{\theta}_n))$ could be used to define yet a different weight matrix, namely, $(1 + F_n(\hat{\theta}_n))^{-1}S_z^{-1}$. The resulting covariance matrix of the estimator will become $(1 + F_n(\hat{\theta}_n))\{\dot{\xi}^T(\hat{\theta}_n)S_z^{-1}\dot{\xi}(\hat{\theta}_n)\}^{-1}$. The estimated variances would be increased, though asymptotically they would be the same. We will discuss this issue elsewhere.

In practice most applications do not involve a structured mean. We have shown that substantial improvements in test statistic accuracy can be obtained applying our approach to this situation with both normal and nonnormal data. Although the sample mean may not be the most efficient estimator of μ with nonnormal data, we have found no evidence that simultaneous estimation of the unstructured mean and the covariance parameters improves the performance of the resulting estimators and test statistics in this situation. Further study of this problem is indicated (see e.g., Kano, Bentler, and Mooijaart, 1993).

Finally, several obvious extensions of our results should be mentioned. In order to keep our presentation simple, we did not discuss constraints on parameters. These are handled in the usual way. Also, all of our results assumed a single model evaluation based on one sample from a given population. In reality, researchers may compare nested models and may use χ^2 difference, Wald, or Lagrange multiplier (score) tests for this purpose (e.g., Satora, 1989). ADF and generalized variants of these tests will be plagued by the same small-sample problems discussed above, and, as we will discuss elsewhere, our new methods can be adopted directly to this situation. Similarly, many applications of structural models are to multiple independent samples from possibly the same population (e.g., Bentler, Lee, and Weng, 1987; Muthén, 1989). Extensions to this situation are direct and need not be detailed.

7 APPENDIX

Proof of Corollary 2: We rewrite Ω^{-1} in Corollary 1 as

$$\Omega^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} \\ \Omega^{21} & \Omega^{22} \end{pmatrix}, \tag{28}$$

where

$$\Omega^{11} = A^{-1} - \dot{\tau}^T(\mu)B^{-1}V_{21}\Sigma^{-1} + (\dot{\tau}^T(\mu) - \Sigma^{-1}V_{12})B^{-1}\dot{\tau}(\mu),$$

$$\Omega^{12} = \Omega^{21T} = (\dot{\tau}^T(\mu) - \Sigma^{-1}V_{12})B^{-1}\dot{\sigma}(\theta),$$

and $\Omega^{22} = \dot{\sigma}^T(\theta)B^{-1}\dot{\sigma}(\theta)$ with $A = \Sigma - V_{12}V_{22}^{-1}V_{21}$. Under the hypothesis that all the third central moments are zero, that is

$$E(x_i - \mu_{i0})(x_j - \mu_{j0})(x_k - \mu_{k0}) = 0, \quad all \quad i, j, k, \tag{29}$$

we have

$$\begin{aligned} v_{i,jk} &= cov(x_i, y_{jk}) \\ &= E(x_i - \mu_{i0})x_j x_k \\ &= \mu_{j0}\sigma_{ik} + \mu_{k0}\sigma_{ij}. \end{aligned} \tag{30}$$

It can be shown that $\dot{\tau}^T(\mu) = \Sigma^{-1}V_{12}$ by a direct and tedious computation. So in (27) the off diagonal matrices become zero and the first diagonal matrix becomes Σ^{-1}. This proves the independence of $\hat{\mu}_n$ and $\hat{\theta}_n$.

Let $V_{22} = (v_{ij,kl})$ with

$$\begin{aligned} v_{ij,kl} &= cov(y_{ij}, y_{kl}) \\ &= E(x_i x_j - E x_i x_j)(x_k x_l - E x_k x_l) \\ &= E x_i x_j x_k x_l - E x_i x_j E x_k x_l. \end{aligned} \tag{31}$$

Since

$$\begin{aligned} E x_i x_j x_k x_l &= \sigma_{ijkl} + \{\sigma_{ijk}\mu_l + \sigma_{ijl}\mu_k + \sigma_{ikl}\mu_j + \sigma_{jkl}\mu_i\} \\ &+ \{\sigma_{ij}\mu_k\mu_l + \sigma_{ik}\mu_j\mu_l + \sigma_{il}\mu_j\mu_k + \sigma_{jk}\mu_i\mu_l + \sigma_{jl}\mu_i\mu_k + \sigma_{kl}\mu_i\mu_j\} \\ &+ \mu_i\mu_j\mu_k\mu_l, \end{aligned} \tag{32}$$

and

$$\begin{aligned} E x_i x_j E x_k x_l &= (\sigma_{ij} + \mu_i\mu_j)(\sigma_{kl} + \mu_k\mu_l) \\ &= \sigma_{ij}\sigma_{kl} + \sigma_{ij}\mu_k\mu_l + \sigma_{kl}\mu_i\mu_j + \mu_i\mu_j\mu_k\mu_l, \end{aligned} \tag{33}$$

we have from (30), (31) and (32)

$$\begin{aligned} v_{ij,kl} &= \sigma_{ijkl} + \{\sigma_{ijk}\mu_l + \sigma_{ijl}\mu_k + \sigma_{ikl}\mu_j + \sigma_{jkl}\mu_i\} \\ &+ \{\sigma_{ik}\mu_j\mu_l + \sigma_{il}\mu_j\mu_k + \sigma_{jk}\mu_i\mu_l + \sigma_{jl}\mu_i\mu_k\} \\ &- \sigma_{ij}\sigma_{kl}. \end{aligned} \tag{34}$$

When the third and fourth central moments of X satisfy

$$\sigma_{ijk} = \sigma_{ijl} = \sigma_{ikl} = \sigma_{jkl} = 0, \tag{35}$$

and

$$\sigma_{ijkl} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}, \tag{36}$$

we have from (33), (34) and (35)

$$\begin{aligned} v_{ij,kl} &= \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \\ &+ \{\sigma_{ik}\mu_j\mu_l + \sigma_{il}\mu_j\mu_k + \sigma_{jk}\mu_i\mu_l + \sigma_{jl}\mu_i\mu_k\}. \end{aligned} \tag{37}$$

Now denote

$$V_{21}\Sigma^{-1}V_{12} = \dot{\tau}(\mu)\Sigma\dot{\tau}^T(\mu) = (\delta_{ij,kl}) \tag{38}$$

As $\dot{\tau}(\mu)$ is a $p^* \times p$ matrix with the ijth row given by

$$\frac{\partial \tau_{ij}}{\partial \mu^T} = \frac{\partial(\mu_i\mu_j)}{\partial \mu^T} = (0, \ldots, 0, \mu_j, 0, \ldots, 0, \mu_i, 0, \ldots, 0), \tag{39}$$

where μ_j is the ith element, and μ_i is the jth element.
We have from (37) and (38)

$$\delta_{ij,kl} = \sigma_{ik}\mu_j\mu_l + \sigma_{il}\mu_j\mu_k + \sigma_{jk}\mu_i\mu_l + \sigma_{jl}\mu_i\mu_k. \tag{40}$$

From (36) and (39) we have

$$B = V_{22} - \dot{\tau}(\mu)\Sigma\dot{\tau}^T(\mu) = (b_{ij,kl}) \tag{41}$$

with

$$b_{ij,kl} = v_{ij,kl} - \delta_{ij,kl} = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}. \tag{42}$$

Using Browne's (1974) notation, we have from (40) and (41)

$$B = 2K_m^T(\Sigma \otimes \Sigma)K_m, \tag{43}$$

where K_m is the matrix such that

$$\sigma(\theta) = K_m vec(\Sigma(\theta)).$$

Since K_m has the properties

$$[K_m^T(\Sigma \otimes \Sigma)K_m]^{-1} = K_m^-(\Sigma^{-1} \otimes \Sigma^{-1})K_m^{-T},$$

and

$$vec(\Sigma(\theta)) = K_m^{-T}\sigma(\theta),$$

with

$$K_m^- = (K_m^T K_m)^{-1} K_m^T,$$

we have

$$\begin{aligned} \dot{\sigma}^T(\theta)B^{-1}\dot{\sigma}(\theta) &= \frac{1}{2}\dot{\sigma}^T(\theta)K_m^-(\Sigma^{-1} \otimes \Sigma^{-1})K_m^{-T}\dot{\sigma}(\theta) \\ &= \frac{1}{2}\dot{\sigma}_a^T(\theta)(\Sigma^{-1} \otimes \Sigma^{-1})\dot{\sigma}_a(\theta). \end{aligned}$$

The proof is finished.

References

[1] Anderson, T.W. (1989), "Linear latent variable models and covariance structures," *Journal of Econometrics,* **41**, 91–119.

[2] —— (1994), "Inference in linear models," in *Multivariate Analysis and Its Applications,* eds. T. W. Anderson, K. T. Fang, and I. Olkin, Hayward, CA: Institute of Mathematical Statistics, pp. 1–20.

[3] Arminger, G., and Schoenberg, R. J. (1989), "Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models," *Psychometrika,* **54**, 409–425.

[4] Bentler, P. M. (1983a), "Some contributions to efficient statistics for structural models: Specification and estimation of moment structures," *Psychometrika,* **48**, 493–517.

[5] —– (1983b), "Simultaneous equation systems as moment structure models: With an introduction to latent variable models," *Journal of Econometrics,* **22**, 13–42.

[6] —– (1989), *EQS Structural Equations Program Manual,* Los Angeles: BMDP Statistical Software.

[7] —– (1993), Structural equation models as nonlinear regression models," in *Statistical Modelling and Latent Variables,* eds. K. Haagen, D. J. Bartholomew, and M. Deistler, Amsterdam: North Holland, pp. 51–64.

[8] —– (1994), "A testing method for covariance structure analysis," in *Multivariate Analysis and Its Applications,* eds. T. W. Anderson, K. T. Fang, and I. Olkin, Hayward, CA: Institute of Mathematical Statistics, pp. 123–136.

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

[10] Bentler, P. M., and Dudgeon, P. (in press), "Covariance structure analysis: Statistical practice, theory, and directions," *Annual Review of Psychology,* Palo Alto: Annual Reviews.

[11] Bentler, P. M., Lee, S.Y., and Weng, J. (1987), "Multiple population covariance structure analysis under arbitrary distribution theory," *Communications in Statistics - Theory and Methods,* **16**, 1951–1964.

[12] Bentler, P. M., and Wu, E. J. C. (1995a), *EQS for Macintosh User's Guide,* Encino, CA: Multivariate Software.

[13] —— (1995b), *EQS for Windows User's Guide,* Encino, CA: Multivariate Software.

[14] Bollen, K. A., and Long, J. S. (eds) (1993), *Testing Structural Equation Models,* Newbury Park, CA: Sage.

[15] Box, G. E. P. (1979), "Robustness in the strategy of scientific model building," *Robustness in Statistics,* eds. R. L. Launer, and G. N. Wilkinson, New York: Academic Press, pp. 201–236.

[16] Browne, M. W. (1974), "Generalized least-squares estimators in the analysis of covariance structures," *South African Statistical Journal,* **8**, 1–24.

[17] —— (1982), "Covariance structure analysis," in *Topics in Applied Multivariate Analysis,* ed. D. M. Hawkins, England: Cambridge University Press, pp. 72–141.

[18] —— (1984), "Asymptotic distribution-free methods for the analysis of covariance structures," *British Journal of Mathematical and Statistical Psychology,* **37**, 62–83.

[19] Browne, M. W., and Arminger G. (1995), "Specification and estimation of mean and covariance models," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences,* eds. G. Arminger, C. C. Clogg, and M. E. Sobel, New York: Plenum, pp. 185–249.

[20] Byrne, B. M. (1994), *Structural Equation Modeling with EQS and EQS/Windows,* Thousand Oaks, CA: Sage.

[21] Carroll, R. J., Wu, C. F. J., and Ruppert, D. (1988), "The effect of estimating weights in weighted least squares," *Journal of American Statistical Association,* **83**, 1045–1054.

[22] Chamberlain, G. (1982), "Multivariate regression models for panel data," *Journal of Econometrics,* **18**, 5–46.

[23] Chan, W. (1995), *Covariance Structure Analysis of Ipsative Data,* Ph.D. thesis, University of California, Los Angeles.

[24] Chou, C.-P., Bentler, P. M., and Satorra, A. (1991), "Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study," *British Journal of Mathematical and Statistical Psychology,* **44**, 347–357.

[25] Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics,* London: Chapman and Hall.

[26] Fang, K.-T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions,* London: Chapman and Hall.

[27] Ferguson, T. (1958), "A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities," *Annal of Mathematical Statistics,* **29**, 1046–1062.

[28] Fuller, W. A. (1987), *Measurement Error Models,* New York: Wiley.

[29] Gierl, M. J., and Mulvenon S. (1995), "Evaluating the application of fit indices to structural equation models in educational research: A review of the literature from 1990 through 1994," presented at *Annual Meetings of the American Educational Research Association,* San Francisco.

[30] Gourieroux, C., Monfort, A., and Trognon, A. (1984), "Pseudo maximum likelihood methods: theory," *Econometrica,* **52**, 681–700.

[31] Henly, S. J. (1993), "Robustness of some estimators for the analysis of covariance structures," *British Journal of Mathematical and Statistical Psychology,* **46**, 313–338.

[32] Hoyle, R. (ed) (1995), *Structural Equation Modeling: Concepts, Issues, and Applications,* Thousand Oaks, CA: Sage.

[33] Hu, L., Bentler, P. M., and Kano, Y. (1992), "Can test statistics in covariance structure analysis be trusted?" *Psychological Bulletin,* **112**, 351–362.

[34] Jöreskog, K. G., and Sörbom, D. (1993), *LISREL 8 User's Reference Guide,* Chicago: Scientific Software International.

[35] Kano, Y. (1992), "Robust statistics for test-of-independence and related structural models," *Statistics and Probability Letters,* **15**, 21-26

[36] Kano, Y., Bentler, P. M., and Mooijaart, A. (1993), "Additional information and precision of estimators in multivariate structural models," in *Statistical Sciences and Data Analysis,* eds. K. Matusita, M. L. Puri, T. Hayakawa, Zeist: VSP International Science Publishers, pp. 187–196.

[37] Lee, S.-Y., and Jennrich, R. I. (1984), "The analysis of structural equation models by means of derivative free nonlinear least squares," *Psychometrika,* **49**, 521–528.

[38] Micceri, T. (1989), "The unicorn, the normal curve, and other improbable creatures," *Psychological Bulletin,* **105**, 156–166.

[39] Muthén, B. (1989), "Multiple group structural modelling with non-normal continuous variables," *British Journal of Mathematical and Statistical Psychology,* **42**, 55–62.

[40] Muthén, B., and Kaplan, D. (1992), "A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model," *British Journal of Mathematical and Statistical Psychology,* **45**, 19–30.

[41] Pierce, D. A. (1982), "The asymptotic effect of substituting estimators for parameters in certain types of statistics," *The Annals of Statistics,* **10**, 475–478.

[42] Rudin, W. (1976), *Principles of Mathematical Analysis,* 3rd ed. New York: McGraw-Hill.

[43] Satorra, A. (1989), "Alternative test criteria in covariance structure analysis: A unified approach," *Psychometrika,* **54**, 131–151.

[44] —– (1992), "Asymptotic robust inferences in the analysis of mean and covariance structures," in *Sociological Methodology,* ed. P. V. Marsden, Blackwell, Oxford, pp. 249–278.

[45] Satorra, A., and Bentler, P. M. (1988), "Scaling corrections for chi-square statistic in covariance structure analysis," *Proceedings of American Statistical Association,* 308–313.

[46] —– "Corrections to test statistics and standard errors in covariance structure analysis," in *Latent Variables Analysis: Applications for Developmental Research,* eds. A. von Eye, and C. C. Clogg, Thousand Oaks, CA: Sage, pp. 399-419.

[47] Satorra, A., and Neudecker, H. (1994), "On the asymptotic optimality of alternative minimum-distance estimators in linear latent-variable models," *Econometric Theory,* **10**, 867–883.

[48] Shapiro, A. (1986), "Asymptotic theory of overparameterized structural models," *Journal of the American Statistical Association,* **81**, 142–149.

[49] Shapiro, A., and Browne, M. W. (1987), "Analysis of covariance structures under elliptical distributions," *Journal of the American Statistical Association,* **82**, 1092–1097.

[50] Stuart, A., and Ord, J. K. (1991), *Kendall's Advanced Theory of Statistics, Vol. 2, 5th ed.,* New York: Oxford University Press.

[51] West, S. G., Finch, J. F., and Curran, P. J. (1995), "Structural equation models with nonnormal variables," in *Structural Equation Modeling: Concepts, Issues, and Applications,* ed. R. Hoyle, Thousand Oaks, CA: Sage, pp. 56–75.

[52] Yung, Y.F., and Bentler, P. M. (1994), "Bootstrap-corrected ADF test statistics in covariance structure analysis," *British Journal of Mathematical and Statistical Psychology,* **47**, 63-84.