PSYCHOMETRIKA—VOL. 66, NO. 3, 421–436
SEPTEMBER 2001

STANDARD ERRORS OF FIT INDICES USING RESIDUALS IN STRUCTURAL EQUATION MODELING

HARUHIKO OGASAWARA

OTARU UNIVERSITY OF COMMERCE

The asymptotic standard errors of the correlation residuals and Bentler's standardized residuals in covariance structures are derived based on the asymptotic covariance matrix of raw covariance residuals. Using these results, approximations of the asymptotic standard errors of the root mean square residuals for unstandardized or standardized residuals are derived by the delta method. Further, in mean structures, approximations of the asymptotic standard errors of residuals, standardized residuals and their summary statistics are derived in a similar manner. Simulations are carried out, which show that the asymptotic standard errors of the various types of residuals and the root mean square residuals in covariance, correlation and mean structures are close to actual ones.

Key words: standardized residuals, correlation residuals, asymptotic standard errors, RMR, covariance structures, mean structures, structural means.

In structural equation modeling, various model fit indices have been proposed. This comes partly from the well-known inappropriateness of the classical likelihood ratio chi-square statistic: in practical situations, models reasonably close to sample covariance matrices with moderate to large sample sizes are frequently rejected. Most of the proposed model fit indices are overall model fit measures (Bollen 1989a, pp. 256–281) which represent model fit by single values. Some of them are based on residuals in a sample covariance matrix after fitting a covariance model (e.g., root mean square residual, RMR, Jöreskog & Sörbom, 1981, sec. I.41; standardized root mean square residual, SRMR, Bentler, 1989, pp. 90–91; see also Hu & Bentler, 1999). Some indices use baseline or null models to be compared with posited models (Bentler, 1990; Bentler & Bonett, 1980; Bollen, 1986, 1989b; McDonald & Marsh, 1990). The so-called goodness of fit indices (GFI; Jöreskog & Sörbom, 1981; see also Bentler, 1983; Tanaka & Huba, 1985, 1989) are based on the discrepancy functions for the estimation of the parameters describing models. Some of the fit indices using the estimates of the noncentrality parameters adopt baseline models (Bentler, 1990; McDonald & Marsh, 1990) or do not use baseline models (Browne & Cudeck, 1993; McDonald, 1989; Steiger, 1989, 1990; Steiger & Lind, 1980). The parsimony index (Mulaik, James, Van Alstine, Bennett, Lind & Stilwell, 1989) can be used with some of the above indices. Further, we have cross validation indices (Browne & Cudeck, 1993; Cudeck & Browne, 1983), information based indices (Akaike, 1973; Schwartz, 1978), the critical N (Hoelter, 1983) and so on.

However, the distributions of most of the above indices are unknown even in large samples except for some of GFIs (Maiti & Mukherjee, 1990) and the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger & Lind, 1980). The properties of these fit indices have been investigated mainly by simulation (see, e.g., Anderson & Gerbing, 1984; Gerbing & Anderson, 1993; Hu & Bentler, 1999; Marsh & Balla, 1994; Marsh, Balla & McDonald, 1988; Marsh & Hau, 1996).

On the other hand, the component fit measures (Bollen, 1989a, pp. 281–289) include parameter estimates, the asymptotic standard errors of parameter estimates and the asymptotic correlations for parameter estimates. The individual residuals (i.e., the (i, j)-th sample covariance

The author is indebted to the reviewers for their comments and suggestions which have led to an improvement of this work.

Requests for reprints should be sent to Haruhiko Ogasawara, Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 JAPAN. E-Mail: hogasa@res.otaru-uc.ac.jp

0033-3123/2001-3/1999-0780-A $00.75/0

minus the corresponding fitted covariance) or their transformations may be classified as component fit measures. These residuals and their summary indices are used to see whether a model is fitted to a sample covariance or correlation matrix, though the individual residuals tend not to be reported in articles since they often require large amount of space as is the case of sample covariance matrices and raw data matrices (see Steiger, 1988). If the model is not well fitted to the sample matrix, then we look for where and how the model does not fit the data. The Lagrange multiplier (LM) test performs a similar work, where a relatively restricted model is fitted and the plausibility of a more unrestricted model with, for example, a new path in the fitted structural model is tested. When the chi-square value for the test is significant, we see that the new parameter should be included in the restricted model. It is to be noted that by applying the LM test, a more unrestricted model should be specified though the parameter estimates in the more unrestricted model are not required. On the other hand, especially in the initial stage of model building, we often have to consider many possible alternatives to revise the present model which has left substantial residuals. In these situations, analysis of residuals would be helpful to clarify where the model should be revised without confining the ways of revision. In some cases, more appropriate models may not be found in the nested sequence of models. In such cases, the LM test is not informative.

The purpose of this article is to derive the asymptotic standard errors of transformed residuals and the summary statistics using residuals. If we have the standard errors, we can evaluate the sizes of the residuals and their overall statistics from a statistical viewpoint. As was addressed earlier, the residuals can be summarized as

$$\text{RMR} = \left\{ \frac{\sum_{i \geq j} (s_{ij} - \hat{\sigma}_{ij})^2}{p^*} \right\}^{1/2} \tag{1}$$

and

$$\text{SRMR} = \left\{ \frac{\sum_{i \geq j} \left(\frac{s_{ij} - \hat{\sigma}_{ij}}{(s_{ii} s_{jj})^{1/2}} \right)^2}{p^*} \right\}^{1/2} \tag{2}$$

where $s_{ij}, (i, j = 1, \ldots, p)$ is the (i, j)-th unbiased sample covariance;

$$\hat{\sigma}_{ij} = \sigma_{ij}(\hat{\boldsymbol{\theta}}), (i, j = 1, \ldots, p)$$

is the fitted (i, j)-th covariance with $\hat{\boldsymbol{\theta}}$ being the estimate of a q-vector of parameters; $p^* = p(p+1)/2$; and p is the number of observed variables. In addition to the above residuals used in RMR and SRMR the correlation residuals (Bollen, 1989a, p. 258)

$$r_{ij} - \hat{\rho}_{ij} = \frac{s_{ij}}{(s_{ii} s_{jj})^{1/2}} - \frac{\hat{\sigma}_{ij}}{(\hat{\sigma}_{ii} \hat{\sigma}_{jj})^{1/2}}, \quad (i, j = 1, \ldots, p) \tag{3}$$

may be used. For some of covariance structure models (e.g., the exploratory factor analysis model) with typical estimation methods such as maximum likelihood estimation, $s_{ii} = \hat{\sigma}_{ii}, (i = 1, \ldots, p)$ and in this case the standardized residuals (see (2)) and the correlation residuals (see (3)) are equivalent.

Jöreskog and Sörbom (1981, sec. I.42) proposed normalized residuals

$$\frac{s_{ij} - \hat{\sigma}_{ij}}{((\hat{\sigma}_{ii} \hat{\sigma}_{jj} + \hat{\sigma}_{ij}^2)/N)^{1/2}}, \quad (i, j = 1, \ldots, p), \tag{4}$$

where N is the number of observations and the denominator was regarded as an estimate of the asymptotic standard error of the numerator. However, they found that the estimate of the standard error was too high and provided more accurate standard errors (Jöreskog & Sörbom, 1989, sec. 1.15; see also Bentler & Dijkstra, 1985, p. 19; Bollen, 1989a, p. 259; Jöreskog, Sörbom, du Toit, & du Toit, Appendix A.1, 1999.) On the other hand, Bentler and Dijkstra (1985, Equation 1.7.5) gave the exact asymptotic covariance matrix of the residuals in means, covariances and frequencies etc. based on the generalized least squares (GLS) discrepancy function with or without restrictions on parameters.

The reason of the use of the denominator of (4) by Jöreskog and Sörbom (1981) is apparent: when $\hat{\sigma}_{ij}$ is replaced with its population counterpart, the estimate of the asymptotic standard error of the residual with the assumption of multivariate normality is represented by the denominator. However, $\hat{\sigma}_{ij}$ is a random variable depending on sample variances and covariances, and is generally correlated with s_{ij}. The correlation may be usually positive since when s_{ij} is high, $\hat{\sigma}_{ij}$ should be influenced by the high value of s_{ij} as well as other s_{ij}'s, which yields a relatively high value of $\hat{\sigma}_{ij}$.

The asymptotic standard errors of various types of residuals and their summary statistics will be derived in the following sections, which may be directly used for testing or constructing confidence intervals. But, more realistic application may be for standardization of residuals as was first intended by Jöreskog and Sörbom's (1981) normalized residuals. Bentler's standardized residuals and the correlation residuals include standardization by the scales of observed variables. But, these standardization methods do not consider the statistical variability for residuals. We should note that the same values for residuals with different corresponding standard errors do not give the same conclusion.

Asymptotic Standard Errors of Covariance Residuals

Let $S(p \times p)$ and $\Sigma = \Sigma(\boldsymbol{\theta})(p \times p)$ be an unbiased sample covariance matrix and a covariance matrix derived by a covariance structure model with the parameter vector $\boldsymbol{\theta}$, respectively. We assume that p observed variables are multivariate normally distributed and employ the maximum Wishart likelihood estimation of the parameters in the model. Then, the discrepancy function is

$$F = \log|\Sigma| - \log|S| + \mathrm{tr}(S\Sigma^{-1}) - p. \tag{5}$$

The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is obtained by minimizing (5), whose necessary conditions are given by the gradient vector of F with respect to $\boldsymbol{\theta}$ set equal to zero:

$$g_i = \frac{\partial F}{\partial \theta_i} = \mathrm{tr}\{(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\dot{\Sigma}_i\} = 0, \quad (i = 1, \ldots, q), \tag{6}$$

where g_i and θ_i are the i-th elements of the gradient vector $\mathbf{g}$ and the parameter vector $\boldsymbol{\theta}$, respectively; and $\dot{\Sigma}_i = \partial\Sigma/\partial\theta_i$. The q equations of (6) represent the relationships between S and $\hat{\boldsymbol{\theta}}$, though usually $\hat{\boldsymbol{\theta}}$ is not an explicit function of S.

Let $u_{ij} = s_{ij} - \hat{\sigma}_{ij}$. Then, since $\hat{\boldsymbol{\theta}}$ is an (implicit) function of S, we see that u_{ij} is a function of S. Let

$$\mathbf{s} = \mathrm{v}(S), \quad \hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}) = \mathrm{v}(\Sigma(\hat{\boldsymbol{\theta}})) \quad \text{and} \quad \mathbf{u} = \mathrm{v}(S - \Sigma(\hat{\boldsymbol{\theta}})),$$

where $\mathrm{v}(\cdot)$ is the vectorizing operator taking nonduplicated elements of a matrix, for example,

$$\mathbf{s} = (s_{11}, s_{21}, s_{22}, \ldots, s_{p,p-1}, s_{pp})'.$$

Then, using the delta method we have the following results first given by Bentler and Dijkstra (1985, Equation 1.7.5); see also Satorra, 1989, Lemma 1; Satorra & Bentler, 1990, Equation 2.17):

$$\text{acov}(\mathbf{u}) = \text{acov}(\mathbf{s} - \hat{\boldsymbol{\sigma}}) = \left(I_{p^*} - \frac{\partial \hat{\boldsymbol{\sigma}}}{\partial \mathbf{s}'}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}} \right) \text{acov}(\mathbf{s}) \left(I_{p^*} - \frac{\partial \hat{\boldsymbol{\sigma}}'}{\partial \mathbf{s}}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}} \right) \tag{7}$$

where I_{p^*} is the $p^* \times p^*$ identity matrix and the partial derivative in (7) is obtained as follows.

$$\frac{\partial \hat{\boldsymbol{\sigma}}}{\partial \mathbf{s}'} = \frac{\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'} \frac{\partial \hat{\boldsymbol{\theta}}}{\partial \mathbf{s}'} = -\frac{\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'} \left(\frac{\partial \mathbf{g}}{\partial \hat{\boldsymbol{\theta}}'} \right)^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{s}'}. \tag{8}$$

The derivation of $\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})/\partial \mathbf{s}'$ in (8) stems from the formula of partial derivatives in implicit functions (see (6)). The partial derivative $\partial \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})/\partial \hat{\boldsymbol{\theta}}'$ is easily obtained since $\boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})$ is usually an explicit function of $\hat{\boldsymbol{\theta}}$.

Since $\text{E}(S) = \Sigma$, $\partial \mathbf{g}/\partial \hat{\boldsymbol{\theta}}'$ in (8) is approximated by

$$\frac{\partial g_i}{\partial \theta_j}\bigg|_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}} \cong \text{tr}(\Sigma^{-1}\dot{\Sigma}_i \Sigma^{-1}\dot{\Sigma}_j), \quad (i, j = 1, \ldots, q), \tag{9}$$

which is well-known in covariance structure analysis. From (6), the matrix $\partial \mathbf{g}/\partial \mathbf{s}'$ is

$$\frac{\partial g_i}{\partial s_{jk}} = -(2 - \delta_{jk})(\Sigma^{-1}\dot{\Sigma}_i \Sigma^{-1})_{jk}, (i = 1, \ldots, q;\ p \geq j \geq k \geq 1), \tag{10}$$

where δ_{jk} is the Kronecker delta and $(\cdot)_{jk}$ is the (j, k)-th element of the parenthesized matrix. With (8) through (10), the estimate of (7) is obtained by replacing $\boldsymbol{\theta}$ with its estimate $\hat{\boldsymbol{\theta}}$.

As an application of (7), we can derive an approximation of the asymptotic variance of RMR in the following way.

$$\text{avar(RMR)} \cong \frac{\text{avar}\left(\sum_{i \geq j} u_{ij}^2\right)}{(2 \times \text{E(RMR)} \times p^*)^2} \cong \frac{\text{tr}[\{\text{acov}(\mathbf{u})\}^2]}{\text{E}(\mathbf{u}'\mathbf{u}) \times 2p^*} = \frac{\text{tr}[\{\text{acov}(\mathbf{u})\}^2]}{\text{tr}\{\text{acov}(\mathbf{u})\} \times 2p^*}, \tag{11}$$

where the expectation is taken in large samples and

$$\text{avar}\left(\sum_{i \geq j} u_{ij}^2 \right) = 2\text{tr}[\{\text{acov}(\mathbf{u})\}^2]$$

is used, whose derivation is provided in Appendix.

Asymptotic Standard Errors of Standardized Residuals

Let b_{ij} be Bentler's (1989) standardized residual, that is,

$$b_{ij} = \frac{u_{ij}}{(s_{ii}s_{jj})^{1/2}}, \quad (i, j = 1, \ldots, p). \tag{12}$$

Noting that in large samples $\text{E}(u_{ij}) \cong 0$, we have the following simple result:

$$\text{acov}(b_{ij}, b_{kl}) = \frac{\text{acov}(u_{ij}, u_{kl})}{(\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{ll})^{1/2}}, \quad (p \geq i \geq j \geq 1;\ p \geq k \geq l \geq 1). \tag{13}$$

which is easily obtained when $\text{acov}(\mathbf{u})$ is given (see (7) and associated equations). Also, an approximation of the asymptotic variance of SRMR is derived similarly to (11):

$$\text{avar(SRMR)} \cong \frac{\text{tr}[\{\text{acov}(\mathbf{b})\}^2]}{\text{tr}\{\text{acov}(\mathbf{b})\} \times 2p^*}, \tag{14}$$

where $\mathbf{b} = (b_{11}, b_{21}, b_{22}, \ldots, b_{p,p-1}, b_{pp})'$.

Asymptotic Standard Errors of Correlation Residuals

Let v_{ij} be the correlation residual corresponding to the (i, j)-th element of a covariance matrix (see (3)). Then, the asymptotic variance of v_{ij} is obtained by

$$\begin{aligned}\text{avar}(v_{ij}) &= \text{avar}(r_{ij} - \hat{\rho}_{ij}) \\ &= \text{avar}(r_{ij}) - 2\text{acov}(\hat{\rho}_{ij}, r_{ij}) + \text{avar}(\hat{\rho}_{ij}), \quad (p \geq i > j \geq 1),\end{aligned} \tag{15}$$

where the values for $i = j = 1, \ldots, p$ are not included since they are always zero. The three terms on the right-hand side of (15) are given from (3) and the following results using the delta method. Let

$$\boldsymbol{\sigma}_{ij}^{+} = \left(\frac{1}{\sigma_{ij}}, -\frac{1}{2\sigma_{ii}}, -\frac{1}{2\sigma_{jj}}\right)', \quad \mathbf{s}_{ij}{}^{*} = (s_{ij}, s_{ii}, s_{jj})' \quad \text{and} \quad \boldsymbol{\sigma}_{ij}{}^{*} = (\sigma_{ij}, \sigma_{ii}, \sigma_{jj})', \tag{16}$$

then

$$\text{acov}(r_{ij}, r_{kl}) = \rho_{ij}\rho_{kl}\boldsymbol{\sigma}_{ij}^{+'}\text{acov}(\mathbf{s}_{ij}{}^{*}, \mathbf{s}_{kl}{}^{*'})\boldsymbol{\sigma}_{kl}^{+} \tag{17}$$

(see also Girshick, 1939; Steiger & Hakstian, 1982),

$$\text{acov}(\hat{\rho}_{ij}, r_{kl}) = \rho_{ij}\rho_{kl}\boldsymbol{\sigma}_{ij}^{+'}\frac{\partial \boldsymbol{\sigma}_{ij}{}^{*}}{\partial \boldsymbol{\theta}'}\frac{\partial \hat{\boldsymbol{\theta}}}{\partial \mathbf{s}'}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}}\text{acov}(\mathbf{s}, \mathbf{s}_{kl}{}^{*'})\boldsymbol{\sigma}_{kl}^{+}, \tag{18}$$

$$\begin{aligned}\text{acov}(\hat{\rho}_{ij}, \hat{\rho}_{kl}) &= \rho_{ij}\rho_{kl}\boldsymbol{\sigma}_{ij}^{+'}\frac{\partial \boldsymbol{\sigma}_{ij}{}^{*}}{\partial \boldsymbol{\theta}'}\text{acov}(\hat{\boldsymbol{\theta}})\frac{\partial \boldsymbol{\sigma}_{kl}{}^{*'}}{\partial \boldsymbol{\theta}}\boldsymbol{\sigma}_{kl}^{+}, \\ &\qquad (p \geq i > j \geq 1;\ p \geq k > l \geq 1).\end{aligned} \tag{19}$$

One of the advantages of the correlation residuals is that $|\hat{\rho}_{ij}| \leq 1$ while $|\hat{\sigma}_{ij}/(s_{ii}s_{jj})^{1/2}|$ in the standardized residual can be more than one. The second advantage of the correlation residual is that when the covariance structure model is for standardized observed variables, $\hat{\rho}_{ij}$ becomes somewhat simpler, because in this case

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{1/2}} = \rho_{ij}(\hat{\boldsymbol{\theta}}), \tag{20}$$

where $\rho_{ij}(\hat{\boldsymbol{\theta}})$ is the estimate of the (i, j)-th element of a fitted correlation matrix. When (20) holds, (18) and (19) become simpler: for example, (19) is simply $(\partial\rho_{ij}/\partial\boldsymbol{\theta}')\text{acov}(\hat{\boldsymbol{\theta}})(\partial\rho_{kl}/\partial\boldsymbol{\theta})$.

The correlation root mean square residual, an overall fit index using the correlation residuals, similar to RMR and SRMR (see (1) and (2)), is defined by

$$\text{CRMR} = \left\{\sum_{i>j}\left(\frac{s_{ij}}{(s_{ii}s_{jj})^{1/2}} - \frac{\hat{\sigma}_{ij}}{(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{1/2}}\right)^2 \Big/ p^-\right\}^{1/2}, \tag{21}$$

where $p^- = p(p-1)/2$. Let $\mathbf{v} = (v_{21}, v_{31}, \ldots, v_{p,p-1})'$. Then, an approximation of the asymptotic variance of CRMR is obtained similarly to those for RMR and SRMR:

$$\text{avar(CRMR)} \cong \frac{\text{tr}[\{\text{acov}(\mathbf{v})\}^2]}{\text{tr}\{\text{acov}(\mathbf{v})\} \times 2p^-}. \tag{22}$$

Asymptotic Standard Errors of Mean Residuals

While structural models are mainly concerned with covariance or correlation structures, the means of observed variable are sometimes assumed to have structured forms in structural equation modeling. A typical model with structured means is the model with factor means (see, e.g., Sörbom, 1974; also Jöreskog & Sörbom, 1996, chap. 10; Yung & Bentler, 1999). In such models, the sample means of observed variables are not necessarily equal to the corresponding means reproduced by the models, which yields mean residuals (the vector of sample means minus the corresponding means from a model). In usual covariance structure models, the means are unconstrained: we have always zero mean residuals. When the means are structured, the mean residuals become component fit measures, which have meanings similar to covariance/correlation residuals to evaluate the goodness-of-fit of a model in means. To assess the sizes of mean residuals, the distributions of the residuals will be helpful. In this section we give the asymptotic standard errors of the mean residuals with the assumption of multivariate normality for observed variables. Note that the results of Bentler and Dijkstra (1985, Equation 1.7.5) in GLS estimation cover the residuals in means and gives the asymptotically equivalent ones to those given in this section. However, since their formulas are somewhat abstract, we present the following results which can be used in actual computation.

When the model means are saturated, the Wishart likelihood is utilized. On the other hand, in case with structured or restricted means, we have to use the likelihood of the original multivariate normal distribution. Let $\boldsymbol{\mu}(p \times 1) = \boldsymbol{\mu}(\boldsymbol{\theta})$ and $\Sigma(p \times p) = \Sigma(\boldsymbol{\theta})$ be the structured mean vector and the structured covariance matrix, respectively, where $\boldsymbol{\theta}(q \times 1)$ is the vector of parameters. The vector $\boldsymbol{\mu}$ and the matrix Σ may or may not have common parameters in $\boldsymbol{\theta}$. Then, the log likelihood of $\boldsymbol{\theta}$ is

$$l = -\frac{N}{2}\{\ln|\Sigma| + \mathrm{tr}(\tilde{S}\Sigma^{-1}) + (\bar{\mathbf{x}} - \boldsymbol{\mu})'\Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) + p\ln(2\pi)\}, \tag{23}$$

where

$$\tilde{S} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'; \mathbf{x}_i(p \times 1)$$

is the vector of the i-th observation; and

$$\bar{\mathbf{x}} = \left(\frac{1}{N}\right)\sum_{i=1}^{N}\mathbf{x}_i.$$

Note that $\tilde{S} = ((N-1)/N)S$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by maximizing (23), which is given from the following equations:

$$g_i \equiv -\frac{2}{N}\frac{\partial l}{\partial \theta_i} = \mathrm{tr}\{(\Sigma^{-1} - \Sigma^{-1}\tilde{S}\Sigma^{-1})\dot{\Sigma}_i\} - 2(\bar{\mathbf{x}} - \boldsymbol{\mu})'\Sigma^{-1}\dot{\boldsymbol{\mu}}_i = 0, \quad (i = 1, \ldots, q), \tag{24}$$

where $\dot{\boldsymbol{\mu}}_i = \partial\boldsymbol{\mu}/\partial\theta_i$.

Let $m_j = \bar{x}_j - \hat{\mu}_j$ be the j-th mean residual, where $\bar{x}_j$ and $\hat{\mu}_j$ are the j-th elements of $\bar{\mathbf{x}}$ and $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$, respectively. Then, the asymptotic covariance of m_j and m_k is obtained as follows:

$$\begin{aligned}\mathrm{acov}(m_j, m_k) &= \mathrm{cov}(\bar{x}_j, \bar{x}_k) - \mathrm{acov}(\bar{x}_j, \hat{\mu}_k) - \mathrm{acov}(\bar{x}_k, \hat{\mu}_j) + \mathrm{acov}(\hat{\mu}_j, \hat{\mu}_k) \\ &= \frac{\sigma_{jk}}{N} - \mathrm{cov}(\bar{x}_j, \bar{\mathbf{x}}')\frac{\partial\hat{\boldsymbol{\theta}}'}{\partial\bar{\mathbf{x}}}\bigg|_{\bar{\mathbf{x}}=\boldsymbol{\mu}}\frac{\partial\mu_k}{\partial\boldsymbol{\theta}} - \mathrm{cov}(\bar{x}_k, \bar{\mathbf{x}}')\frac{\partial\hat{\boldsymbol{\theta}}'}{\partial\bar{\mathbf{x}}}\bigg|_{\bar{\mathbf{x}}=\boldsymbol{\mu}}\frac{\partial\mu_j}{\partial\boldsymbol{\theta}} \\ &\quad + \frac{\partial\mu_j}{\partial\boldsymbol{\theta}'}\mathrm{acov}(\hat{\boldsymbol{\theta}})\frac{\partial\mu_k}{\partial\boldsymbol{\theta}}, \quad (p \geq j \geq 1; p \geq k \geq 1),\end{aligned} \tag{25}$$

where $\text{acov}(\bar{x}_j, \tilde{\mathbf{s}}') = \mathbf{0}'$ with $\tilde{\mathbf{s}} = \text{v}(\tilde{S})$ is used;

$$\text{cov}(\bar{x}_j, \bar{\mathbf{x}}') = (\sigma_{j1}, \sigma_{j2}, \ldots, \sigma_{jp})/N, \quad (j = 1, \ldots, p); \text{ and } \partial\mu_j/\partial\boldsymbol{\theta}', \quad (j = 1, \ldots, p)$$

are easily obtained since μ_j, $(j = 1, \ldots, p)$ are usually explicit functions of $\boldsymbol{\theta}$. The remaining partial derivatives in (25) are given by

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{t}'} = -\left(\frac{\partial\mathbf{g}}{\partial\hat{\boldsymbol{\theta}}'}\right)^{-1}\frac{\partial\mathbf{g}}{\partial\mathbf{t}'} \quad \text{with} \quad \mathbf{t} = (\tilde{\mathbf{s}}', \bar{\mathbf{x}}')' \quad \text{and} \quad \mathbf{g} = (g_1, \ldots, g_q)'. \tag{26}$$

For covariance residuals in the case of structured means, $\partial\hat{\boldsymbol{\theta}}/\partial\tilde{\mathbf{s}}'$ in (26) is the same in form as the corresponding term in (8). However, we should note that $\partial\mathbf{g}/\partial\hat{\boldsymbol{\theta}}'$ is to be obtained from (24), which is generally different from (6). That is,

$$\left.\frac{\partial g_i}{\partial\hat{\theta}_j}\right|_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}} \cong \text{tr}(\Sigma^{-1}\dot{\Sigma}_i\Sigma^{-1}\dot{\Sigma}_j) + 2\dot{\boldsymbol{\mu}}_i'\Sigma^{-1}\dot{\boldsymbol{\mu}}_j, \quad (i, j = 1, \ldots, q) \tag{27}$$

and other partial derivatives in (26) are

$$\left.\frac{\partial g_i}{\partial\tilde{s}_{jk}}\right|_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}} = -(2 - \delta_{jk})(\Sigma^{-1}\dot{\Sigma}_i\Sigma^{-1})_{jk}, \quad (i = 1, \ldots, q;\ p \geq j \geq k \geq 1), \tag{28}$$

$$\left.\frac{\partial g_i}{\partial\bar{\mathbf{x}}}\right|_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}} = -2\Sigma^{-1}\dot{\boldsymbol{\mu}}_i, (i = 1, \ldots, q). \tag{29}$$

To derive (27), we used the relationships $\tilde{S} \cong \Sigma$ and $\bar{\mathbf{x}} \cong \boldsymbol{\mu}$ in large samples. The summary fit index for mean residuals (root mean square of mean residuals, RMMR) is defined similarly to RMR:

$$\text{RMMR} = \left(\sum_{j=1}^{p}\frac{m_j^2}{p}\right)^{1/2}, \tag{30}$$

whose asymptotic variance is approximated by

$$\text{avar(RMMR)} \cong \frac{\text{tr}[\{\text{acov}(\mathbf{m})\}^2]}{\text{tr}\{\text{acov}(\mathbf{m})\} \times 2p} \quad \text{with} \quad \mathbf{m} = (m_1, \ldots, m_p)', \tag{31}$$

which is obtained in a manner similar to (11).

The mean residuals m_j, $(j = 1, \ldots, p)$ depend on scales and cannot directly be compared. Therefore, the standardized mean residuals and their summary statistics are defined by

$$c_j = \frac{m_j}{(\hat{\sigma}_{jj})^{1/2}}, (j = 1, \ldots, p) \text{ and SRMMR (standardized RMMR)} = \left(\sum_{j=1}^{p}\frac{c_j^2}{p}\right)^{1/2}, \tag{32}$$

where $\hat{\sigma}_{jj}$ may be replaced by s_{jj}. Their asymptotic standard errors can be obtained in a manner similar to the previous formulas.

Numerical Examples

Numerical examples are based on two real correlation matrices. One is from Harman's (1976, p. 22) eight physical variables ($N = 305$). Another is also from Harman's (1976, p. 401) twelve psychological tests ($N = 355$). We assumed that the correlation matrices were sample covariance matrices. In the first stage, unrestricted orthogonal factor analysis models for unstandardized observed variables were fitted with 2 and 3 common factors to the two (assumed) sample

covariance matrices, respectively, by the maximum likelihood method. In the second stage, which will be explained later, we fitted restricted factor analysis models to these data. Because we have the rotational indeterminacy for the models in the first stage, we set $(k^2 - k)/2$ elements of the factor loading matrices equal to zero without loss of generality, where k is the number of common factors. From the property of the maximum likelihood estimates for the unrestricted factor analysis model, we have $s_{ii} = \hat{\sigma}_{ii}$, $(i = 1, \ldots, p)$ and from the assumption of $s_{ii} = 1$, $(i = 1, \ldots, p)$ for these data, it follows that $u_{ij} = b_{ij} = v_{ij}$ and $\hat{\mathrm{SE}}(u_{ij}) = \hat{\mathrm{SE}}(b_{ij}) = \hat{\mathrm{SE}}(v_{ij})$.

Tables 1 and 2 show the results for the first stage. For the twelve psychological tests, Table 2 shows the results for every other observed variables to save space. The tables contain LISREL5 SEs obtained from the denominator of (8) multiplied by $(N/(N-1))^{1/2}$ without changing the asymptotic property, and the theoretical SEs which are the estimated standard errors (i.e., $\hat{\mathrm{SE}}(u_{ij})$, $\hat{\mathrm{SE}}(b_{ij})$ and $\hat{\mathrm{SE}}(v_{ij})$ from (7), (13) and (15)). The tables also contain simulated results, which have been obtained in the following way. First, the fitted covariance matrices were regarded as population covariance matrices. (Consequently, the theoretical SEs in the tables become population ones in this simulation study.) Then, with the assumption of multivari-

TABLE 1.
Unrestricted factor analysis for the eight physical bariables ($N = 305$; Harman, 1976, p. 22)

i	j	s_{ij}	$\hat{\sigma}_{ij}$	$s_{ij} - \hat{\sigma}_{ij}$	LISREL5 SE	Theoretical SE	Simulated SE	z
2	1	.846	.854	−.008	.0754	.0037	.0037	−2.2
3	1	.805	.826	−.021	.0744	.0060	.0061	−3.5
3	2	.881	.863	.018	.0758	.0033	.0034	5.5
4	1	.859	.815	.044	.0740	.0074	.0075	5.9
4	2	.826	.842	−.016	.0750	.0043	.0044	−3.7
4	3	.801	.814	−.013	.0739	.0069	.0071	−1.9
5	1	.473	.467	.006	.0633	.0033	.0033	1.8
5	2	.376	.384	−.008	.0614	.0023	.0023	−3.4
5	3	.380	.375	.005	.0612	.0032	.0033	1.7
5	4	.436	.433	.003	.0625	.0037	.0036	.8
6	1	.398	.390	.008	.0616	.0118	.0115	.7
6	2	.326	.320	.006	.0602	.0082	.0079	.7
6	3	.319	.313	.006	.0601	.0115	.0113	.6
6	4	.329	.362	−.033	.0610	.0131	.0131	−2.5
6	5	.762	.761	.001	.0721	.0022	.0023	.3
7	1	.301	.332	−.031	.0604	.0128	.0128	−2.4
7	2	.277	.260	.017	.0593	.0089	.0091	1.9
7	3	.237	.255	−.018	.0592	.0124	.0127	−1.4
7	4	.327	.304	.023	.0600	.0142	.0143	1.6
7	5	.730	.727	.003	.0709	.0027	.0029	1.0
7	6	.583	.608	−.025	.0671	.0143	.0145	−1.7
8	1	.382	.418	−.036	.0622	.0150	.0153	−2.4
8	2	.415	.370	.045	.0612	.0104	.0105	4.3
8	3	.345	.360	−.015	.0610	.0146	.0148	−1.0
8	4	.365	.395	−.030	.0617	.0166	.0163	−1.8
8	5	.629	.640	−.011	.0681	.0048	.0051	−2.2
8	6	.577	.535	.042	.0650	.0209	.0209	2.0
8	7	.539	.505	.034	.0642	.0231	.0233	1.5
			RMR = .0230			.00252	.00253	9.1
			SRMR/CRMR = .0230			.00252	.00262	9.1

Note: $z = (s_{ij} - \hat{\sigma}_{ij})/$ Theoretical SE, SE = Standard Error, RMR = Root Mean Square Residual, SRMR = Standardized RMR, CRMR = Correlation RMR.

TABLE 2.
Unrestricted factor analysis for the twelve psychological tests ($N = 355$; Harman, 1976, p. 401)

i	j	s_{ij}	$\hat{\sigma}_{ij}$	$s_{ij} - \hat{\sigma}_{ij}$	LISREL5 SE	Theoretical SE	Simulated SE	z
4	2	.557	.583	−.026	.0615	.0118	.0118	−2.2
6	2	.300	.314	−.014	.0557	.0148	.0147	−1.0
6	4	.200	.212	−.012	.0543	.0232	.0225	−.5
8	2	.448	.423	.025	.0577	.0103	.0103	2.4
8	4	.310	.292	.018	.0554	.0161	.0157	1.1
8	6	.545	.591	−.046	.0617	.0143	.0135	−3.2
10	2	.377	.371	.006	.0567	.0167	.0169	.4
10	4	.286	.289	−.003	.0553	.0264	.0253	−.1
10	6	.407	.356	.051	.0564	.0262	.0260	2.0
10	8	.385	.411	−.026	.0575	.0183	.0183	−1.4
12	2	.200	.189	.011	.0541	.0168	.0166	.7
12	4	.145	.150	−.005	.0537	.0264	.0256	−.2
12	6	.236	.252	−.016	.0548	.0261	.0253	−.6
12	8	.285	.265	.020	.0550	.0182	.0180	1.1
12	10	.213	.281	−.068	.0552	.0268	.0272	−2.6
			RMR = .0257			.00228	.00240	11.3
			SRMR/CRMR = .0257			.00228	.00243	11.3

Note: $z = (s_{ij} - \hat{\sigma}_{ij})/$ Theoretical SE, SE = Standard Error, RMR = Root Mean Square Residual, SRMR = Standardized RMR, CRMR = Correlation RMR.

ate normality, independent observations with the sample sizes equal to the real ones ($N = 305$ and $N = 355$ for the two data sets, respectively) were generated. Based on these observations, the parameters of the factor models were estimated, and the residuals were calculated. From the residuals, we had RMR, SRMR and CRMR. Note that sample variances are not necessarily unities in simulated data and hence RMR and SRMR/CRMR are not necessarily equal in simulation. On the other hand, SRMR and CRMR are equal even in simulation because we always have $\hat{\sigma}_{ii} = s_{ii}$, $(i = 1, \ldots, p)$ for the unrestricted factor analysis model in case of the maximum likelihood estimation. The asymptotic standard errors of the residuals, RMR and SRMR/CRMR were estimated by our methods. As was explained above, since the diagonal elements of fitted covariance matrices are always equal to the corresponding elements of sample covariance matrices in Tables 1 and 2, the values p^* in (1) and (2) were tentatively replaced by the numbers of the nonduplicated off-diagonal elements p^- of the covariance matrices.

The above procedure was replicated until 1,000 regular sets of parameter estimates were obtained, where the Heywood cases (one case each for the first and second examples) had been excluded from the regular sets of estimates. Then, we had 1,000 values for each (i, j)-th resudual, RMR and SRMR/CRMR. The simulated SEs in the tables show the standard deviations of the residuals, RMR and SRMR/CRMR over 1,000 replications, which are regarded as true or actual standard errors. The z's in the tables are the values of $u_{ij} (= b_{ij} = v_{ij})$, RMR and SRMR/CRMR divided by their corresponding theoretical SEs in the tables. It can be shown that the corresponding z-values in the three types of residuals become equal even when s_{ii}, $(i = 1, \ldots, p)$ are not unities if $\hat{\sigma}_{ii} = s_{ii}$ and associated parameters are scale-free.

It is apparent that the LISREL5 SEs are too high by comparison with the actual SEs and that the theoretical standard errors of the residuals are close to the actual SEs, which supports the appropriateness of our method. The z values for the raw residuals have the asymptotic standard normal distributions and some of the absolute z values show the significant values such as greater than 1.96 with the two-tailed probability less than 0.05. These values may be associated with the fact that the likelihood ratio χ^2 values are 78.0 (d.f. = 13, $p < 0.001$) and 73.7 (d.f. = 33, $p <$

0.001) for the results for Tables 1 and 2, respectively. We should note that the summary statistics of the residuals (e.g., RMR) take only nonnegative values and are not normally distributed even in large samples. However, in some cases, we can use variable transformations such as a logarithmic transformation, which may yield less skewed distributions. The asymptotic standard error of the transformed statistic is easily obtained (see e.g., Browne, 1982, p. 96).

In the second stage, the restricted factor models with equal uniquenesses (the variances of unique factors) were fitted to the same (assumed) sample covariance matrices used in the first stage. The numbers of the common factors are the same as those in the first stage. Note that even with the restriction for the unique factors, we still have the rotational indeterminacy for the common factors. So, we used the same parameter patterns (i.e., $(k^2 - k)/2$ fixed zero loadings) for the identification of the loading matrices as those in the first stage. Tables 3 and 4 show the results for every other observed variables. Note that with the restriction for the unique factors, $\hat{\sigma}_{ii}$'s are no longer equal to the corresponding s_{ii}'s. Therefore, RMR and SRMR in the second stage were defined over p^* elements as in the original definitions.

The tables contain the estimates of the asymptotic standard errors of the three types of residuals, which are accompanied by their corresponding simulated standard errors. The simulated values have been obtained similarly as in the first stage with 1,000 replications, where the reproduced covariance matrices ($\hat{\sigma}_{ij}$'s) were regarded as population ones. The simulation in the second stage had no Heywood case until 1,000 regular samples were generated.

From the tables, we find that the standard errors of correlation residuals tend to be larger than those of the corresponding residuals or standardized residuals in these data. (The equivalence of RMR and SRMR except in simulation comes from $s_{ii} = 1, (i = 1, \ldots, p)$.) This can be seen both in theoretical and simulated values, which also supports the appropriateness of our procedure of estimating the standard errors.

A numerical example for the model with structural means is based on a factor analysis model with nonzero factor means for six unstandardized observed variables:

$$\boldsymbol{\mu} = \Lambda\boldsymbol{\gamma}, \Sigma = \Lambda\Lambda' + \Psi \text{ with } \boldsymbol{\gamma} = (1, 1)',$$
$$\Lambda = \begin{bmatrix} 011445 \\ 543321 \end{bmatrix}', \Psi = \text{diag}(2, 2, 2, 2, 2, 2), \tag{33}$$

TABLE 3.
Restricted factor analysis for the eight physical variables ($N = 305$; Harman, 1976, p.22)

i	j	$s_{ij} - \hat{\sigma}_{ij}$	SE of Residual Theor.	SE of Residual Simul.	SE of Stand. Resi. Theor.	SE of Stand. Resi. Simul.	SE of Corr. Resi. Theor.	SE of Corr. Resi. Simul.
2	2	−.094	.0142	.0141	.0130	.0131	*	*
4	2	.010	.0116	.0118	.0107	.0110	.0159	.0165
4	4	−.059	.0146	.0149	.0138	.0142	*	*
6	2	−.009	.0110	.0107	.0109	.0107	.0112	.0112
6	4	−.026	.0112	.0109	.0113	.0110	.0117	.0114
6	6	.069	.0143	.0148	.0153	.0161	*	*
8	2	.039	.0115	.0115	.0120	.0122	.0128	.0130
8	4	−.026	.0116	.0114	.0124	.0123	.0134	.0133
8	6	−.041	.0118	.0119	.0134	.0137	.0189	.019
8	8	.166	.0155	.0154	.0185	.0188	*	*
		RMR = .0510	.00196	.00196				
		SRMR = .0510			.00200	.00214		
		CRMR = .0652					.00240	.00251

Note: SE = Standard Error, RMR = Root Mean Square Residual, SRMR = Standardized RMR, CRMR = Correlation RMR. The values denoted by asterisks are zero by definition.

TABLE 4.
Restricted factor analysis for the twelve psychological tests ($N = 355$; Harman, 1976, p. 401)

i	j	$s_{ij} - \hat{\sigma}_{ij}$	SE of Residual Theor.	SE of Residual Simul.	SE of Stand. Resi. Theor.	SE of Stand. Resi. Simul.	SE of Corr. Resi. Theor.	SE of Corr. Resi. Simul.
2	2	−.096	.0244	.0257	.0222	.0237	*	*
4	2	−.017	.0187	.0182	.0180	.0176	.0245	.0240
4	4	.011	.0235	.0232	.0238	.0238	*	*
6	2	−.018	.0186	.0184	.0177	.0176	.0181	.0181
6	4	.004	.0183	.0181	.0184	.0184	.0180	.0181
6	6	−.003	.0251	.0250	.0250	.0251	*	*
8	2	.029	.0187	.0189	.0171	.0174	.0184	.0187
8	4	.020	.0184	.0180	.0177	.0174	.0179	.0176
8	6	−.038	.0195	.0190	.0187	.0182	.0247	.0244
8	8	−.088	.0255	.0257	.0234	.0239	*	*
10	2	.002	.0199	.0203	.0209	.0216	.0232	.0241
10	4	−.010	.0195	.0188	.0216	.0209	.0233	.0224
10	6	.021	.0202	.0199	.0222	.0219	.0250	.0246
10	8	−.049	.0203	.0203	.0215	.0215	.0246	.0253
10	10	.174	.0287	.0278	.0347	.0339	*	*
12	2	.015	.0151	.0149	.0152	.0152	.0153	.0154
12	4	−.004	.0148	.0145	.0157	.0155	.0160	.0160
12	6	.002	.0153	.0151	.0161	.0160	.0162	.0162
12	8	.035	.0154	.0157	.0156	.0160	.0156	.0161
12	10	−.099	.0168	.0170	.0195	.0198	.0235	.0237
12	12	.101	.0166	.0171	.0184	.0193	*	*
		RMR = .0455	.00211	.00215				
		SRMR = .0455			.00217	.00226		
		CRMR = .0490					.00232	.00238

Note: SE = Standard Error, RMR = Root Mean Square Residual, SRMR = Standardized RMR, CRMR = Correlation RMR. The values denoted by asterisks are zero by definition.

where $\boldsymbol{\gamma}$ is a population factor mean vector; Λ is a population loading matrix; and Ψ is a diagonal matrix with the diagonal elements being the population variances of unique factors. The zero element in Λ is a fixed loading. Note that the loading matrix Λ appears both in the mean vector and the covariance matrix. Table 5 shows the theoretical and simulated standard errors of the mean

TABLE 5.
Standard errors of mean residuals ($N = 400$)

i	SE of $\bar{x}_i - \hat{\mu}_i$ Theor.	SE of $\bar{x}_i - \hat{\mu}_i$ Simul.	SE of $(\bar{x}_i - \hat{\mu}_i)/\sqrt{\hat{\sigma}_{ii}}$ Theor.	SE of $(\bar{x}_i - \hat{\mu}_i)/\sqrt{\hat{\sigma}_{ii}}$ Simul.
1	.0274	.0279	.0053	.0054
2	.0343	.0353	.0079	.0081
3	.0374	.0371	.0108	.0107
4	.0339	.0340	.0065	.0066
5	.0343	.0347	.0073	.0074
6	.0287	.0290	.0054	.0055
RMMR	.0116	.0113		
SRMMR			.00276	.00263

Note SE = Srandard Error, RMMR = Root Mean Square of Mean Residuals, SRMMR = Standardized RMMR.

TABLE 6.
Correlations between mean residuals ($N = 400$)

1	1.00	−.66	−.27	−.06	.04	.26
2	−.63	1.00	−.38	−.23	−.07	.20
3	−.36	−.30	1.00	−.15	−.07	.03
4	−.17	−.15	−.13	1.00	−.48	−.52
5	.06	−.12	−.03	−.39	1.00	−.46
6	.35	.09	−.02	−.49	−.55	1.00

Note: The elements above and below the main diagonal of the matrix indicate the theoretical and simulated correlations, respectively.

residuals, the standardized mean residuals and their summary indices. The theoretical standard errors were obtained by using the above population values with the assumption of $N = 400$. The simulated standard errors were obtained using the population values in a manner similar to the simulation in the previous examples. The number of replications in the simulation was 1,000. From Table 5, we see that the theoretical values are close to their corresponding simulated values. It is to be noted that while the standard errors of raw mean residuals are similar to each other though they depend on scales, the standard errors of the standardized mean residuals are substantially different. Table 6 shows the theoretical and simulated correlations between the raw mean residuals. The theoretical correlations are similar to their corresponding simulated correlations, which also shows the appropriateness of our procedure.

Discussion

In this article, we used the maximum likelihood method with the assumption of multivariate normality for the estimation of the parameters in mean and covariance structure models. However, it is to be noted that in principle, the assumption and the procedure can be relaxed or replaced by other ones. For instance, the basic equation (7) for covariance structure models without structured means holds irrespective of different distributional assumptions and different estimation methods. The normal theory discrepancy function (5) may be replaced by other ones, for example, those for elliptical distributions, if necessary. When we use the GLS estimation method and the unweighted least squares (LS) method, we have typical elements of the gradient vectors corresponding to (6) in the following way:

$$g_{\mathrm{GLS}i} = \frac{\partial F_{\mathrm{GLS}}}{\partial \theta_i} = \mathrm{tr}\{(S^{-1}\Sigma S^{-1} - S^{-1})\dot{\Sigma}_i\} = 0 \tag{34}$$

and

$$g_{\mathrm{LS}i} = \frac{\partial F_{\mathrm{LS}}}{\partial \theta_i} = \mathrm{tr}\{(\Sigma - S)\dot{\Sigma}_i\} = 0, (i = 1, \ldots, q), \tag{35}$$

with fit functions F_{GLS} and F_{LS} for the GLS and LS estimations, respectively. The equations (34) and (35) evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ represent (implicit) functions between S and $\hat{\boldsymbol{\theta}}$. It is to be noted that the asymptotic distribution of the residuals in covariances by the GLS method is equivalent to that by the ML method since the GLS estimator is asymptotically equivalent to the ML estimator (Browne, 1974/1977).

In case with structured means, we can use the minimum chi-square estimation with or without normality assumption for observed variables (see Ferguson, 1958; also Bentler, 1989; Kano, Bentler & Mooijaart, 1993). The discrepancy function to be minimized is

$$F_{\mathrm{MC}} = \{\mathbf{t}' - (\mathrm{v}'(\Sigma(\boldsymbol{\theta})), \boldsymbol{\mu}'(\boldsymbol{\theta}))\}\hat{V}^{-1}\{\mathbf{t} - (\mathrm{v}'(\Sigma(\boldsymbol{\theta})), \boldsymbol{\mu}'(\boldsymbol{\theta}))'\}, \tag{36}$$

where $\hat{V}$ is a consistent estimator of the asymptotic covariance matrix of $\mathbf{t}$. (Note that $\mathbf{t} = (\tilde{\mathbf{s}}', \overline{\mathbf{x}}')'$ may be replaced by $(\mathbf{s}', \overline{\mathbf{x}}')'$ without changing the asymptotic property of the estimator of $\boldsymbol{\theta}$. See, e.g., Bentler, 1989, p. 224.) Let

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{37}$$

where the submatrices correspond to the subvectors in $\mathbf{t} = (\tilde{\mathbf{s}}', \overline{\mathbf{x}}')'$. In the asymptotically distribution-free case, V_{11} and V_{12} consist of the fourth- and third-order moments, respectively (Browne, 1984). For the normal distribution, we have the well-known simple results: a typical element of V_{11} is $(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})/N$ and $V_{12} = O$ ($V_{22} = \{\sigma_{ij}/N\}$ is unchanged). The implicit function between $\mathbf{t}$ and the estimate $\hat{\boldsymbol{\theta}}$ is obtained from $\partial F_{\mathrm{MC}}/\partial\boldsymbol{\theta} = \mathbf{0}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

In the case of structured means, overall statistics for the residuals in covariances/correlations and means may be constructed to evaluate the overall goodness-of-fit of a model. For instance, $\mathrm{T} = w_{\mathrm{T}1}\mathrm{RMR} + w_{\mathrm{T}2}\mathrm{RMMR}$ or $\mathrm{ST} = w_{\mathrm{S}1}\mathrm{SRMR} + w_{\mathrm{S}2}\mathrm{SRMMR}$ may be used for such indices, where $w_{\mathrm{T}i}, w_{\mathrm{S}i}$, $(i = 1, 2)$ are the relative weights to be determined by researchers. The asymptotic standard errors of T and ST are derived straightforwardly by using again the delta method. However, in this case the asymptotic covariances between the mean and covariance residuals are required. With the assumption of normality, they are given in the following way.

$$\begin{aligned}
\mathrm{acov}(m_i, u_{jk}) &= \mathrm{cov}(\overline{x}_i, \tilde{s}_{jk}) - \mathrm{acov}(\overline{x}_i, \hat{\sigma}_{jk}) - \mathrm{acov}(\hat{\mu}_i, \tilde{s}_{jk}) + \mathrm{acov}(\hat{\mu}_i, \hat{\sigma}_{jk}) \\
&= -\mathrm{cov}(\overline{x}_i, \overline{\mathbf{x}}')\frac{\partial\hat{\sigma}_{jk}}{\partial\overline{\mathbf{x}}}\bigg|_{\overline{\mathbf{x}}=\boldsymbol{\mu}} - \frac{\partial\hat{\mu}_i}{\partial\tilde{\mathbf{s}}'}\bigg|_{\tilde{\mathbf{s}}=\boldsymbol{\sigma}}\mathrm{acov}(\tilde{\mathbf{s}}, \tilde{s}_{jk}) + \frac{\partial\mu_i}{\partial\boldsymbol{\theta}'}\mathrm{acov}(\hat{\boldsymbol{\theta}})\frac{\partial\sigma_{jk}}{\partial\boldsymbol{\theta}} \\
&= -\frac{(\sigma_{i1}, \ldots, \sigma_{ip})}{N}\frac{\partial\hat{\boldsymbol{\theta}}'}{\partial\overline{\mathbf{x}}}\bigg|_{\overline{\mathbf{x}}=\boldsymbol{\mu}}\frac{\partial\sigma_{jk}}{\partial\boldsymbol{\theta}} \\
&\quad - \frac{\partial\hat{\mu}_i}{\partial\tilde{\mathbf{s}}'}\bigg|_{\tilde{\mathbf{s}}=\boldsymbol{\sigma}}\frac{(2\sigma_{1j}\sigma_{1k}, \sigma_{2j}\sigma_{1k} + \sigma_{2k}\sigma_{1j}, \ldots, 2\sigma_{pj}\sigma_{pk})'}{N} \\
&\quad + \frac{\partial\mu_i}{\partial\boldsymbol{\theta}'}\mathrm{acov}(\hat{\boldsymbol{\theta}})\frac{\partial\sigma_{jk}}{\partial\boldsymbol{\theta}}, \qquad (i, j, k = 1, \ldots, p),
\end{aligned} \tag{38}$$

where $\mathrm{cov}(\overline{x}_i, \tilde{\mathbf{s}}') = \mathbf{0}'$ is used.

The objective of the present paper was to derive the asymptotic standard errors of the various types of residuals and their summary indices, and not to choose appropriate type(s) of residuals for model evaluation. However, from the limited numerical examples, it seems that Bentler's standardized residuals in covariances are more stable than the correlation residuals though the correlation residuals have some advantages mentioned earlier. (Recall that we cannot directly compare raw residuals with standardized and correlation residuals because of the scale dependency of the raw residuals.)

In previous sections, we assumed that a model was true. However, in practice, structural models may be at most approximations to realities as is often discussed in structural equation modeling. In such cases, the asymptotic results derived in this article may be meaningless. But, our results can be extended to such cases, when a model is slightly misspecified in the sense that $\mathrm{E}(s_{ij}) - \sigma_{ij} = O(1/\sqrt{N})$ (see, e.g., Bentler & Dijkstra, 1985; Satorra, 1989). When the relaxed assumption holds, it can be shown that $u_{ij}^2/\mathrm{avar}(u_{ij})$ has the asymptotic noncentral chi-square distribution with df = 1 and the noncentrality parameter being say, $\delta_{ij} \equiv \{\mathrm{E}(u_{ij})\}^2/\mathrm{avar}(u_{ij})$. In this case, the asymptotic variance of fit indices using residuals e.g., RMR can also be derived by replacing asymptotic expectation of RMR and acov($\mathbf{u}$) by corresponding appropriate ones. The former, E(RMR), is easily obtained as follows:

$$\mathrm{E(RMR)} \cong (\mathrm{E}(\mathbf{u}'\mathbf{u})/p^*)^{1/2} = \left(\left[\mathrm{tr}\{\mathrm{acov}(\mathbf{u})\} + \sum_{i\geq j}\delta_{ij}\mathrm{avar}(u_{ij})\right]\Big/ p^*\right)^{1/2}. \tag{39}$$

The latter, acov(**u**), is provided in the Appendix.

Appendix

The Asymptotic Variance of the Sum of Squared Residuals

In this appendix, we deal with the case in which models are slightly misspecified in the sense that

$$\mathrm{E}(u_{ij}) = O(1/\sqrt{N}). \tag{A1}$$

Let $\mathbf{u}^* = (u_{11}^2, u_{21}^2, u_{22}^2, \ldots, u_{p,p-1}^2, u_{pp}^2)'$. Then, we have

$$\mathrm{avar}\left(\sum_{i\geq j} u_{ij}^2\right) = \mathrm{avar}(\mathbf{u}'\mathbf{u}) = \mathbf{1}_{p^*}'\mathrm{acov}(\mathbf{u}^*)\mathbf{1}_{p^*} \tag{A2}$$

where $\mathbf{1}_{p^*}$ is the $p^* \times 1$ vector consisting of ones. The (ij, kl)-th element of acov($\mathbf{u}^*$), acov(u_{ij}^2, u_{kl}^2), ($p \geq i \geq j \geq 1$; $p \geq k \geq l \geq 1$), is obtained as follows.

Let $z_1 = u_1/\mathrm{ase}(u_1)$ and $z_2 = u_2/\mathrm{ase}(u_2)$, where the subscripts 1 and 2 denote pairs (i, j) and (k, l), respectively, and ase$(\cdot)$ denotes the asymptotic standard error of the parenthesized variable. Then, from (A1) z_1 and z_2 are asymptotically distributed according to $N(\delta_1, 1)$ and $N(\delta_2, 1)$ respectively, where δ_1 and δ_2 are of order $O(1)$. We assume that the asymptotic correlation between z_1 and z_2 is ρ_{12}.

First we derive the asymptotic covariance between u_1^2 and u_2^2. Write z_2 as

$$z_2 = \rho_{12}(z_1 - \delta_1) + \sqrt{1-\rho_{12}^2}e_{12} + \delta_2, \tag{A3}$$

where e_{12} has the asymptotic distribution $N(0, 1)$ and acov$(e_{12}, z_1) = 0$. Then,

$$\begin{aligned}
\mathrm{acov}(u_1^2, u_2^2) &= \mathrm{E}(u_1^2u_2^2) - \mathrm{E}(u_1^2)\mathrm{E}(u_2^2) \\
&= \mathrm{avar}(u_1)\mathrm{avar}(u_2)\big[\mathrm{E}\big[\{(z_1-\delta_1)^2 + 2(z_1-\delta_1)\delta_1 + \delta_1^2\} \\
&\quad \times \{\rho_{12}^2(z_1-\delta_1)^2 + 2\rho_{12}(1-\rho_{12}^2)^{1/2}(z_1-\delta_1)e_{12} + (1-\rho_{12}^2)e_{12}^2 \\
&\quad + 2\rho_{12}(z_1-\delta_1)\delta_2 + 2(1-\rho_{12}^2)^{1/2}e_{12}\delta_2 + \delta_2^2\}\big] - (1+\delta_1^2)(1+\delta_2^2)\big] \\
&= 2\mathrm{avar}(u_1)\mathrm{avar}(u_2)\rho_{12}(\rho_{12} + 2\delta_1\delta_2) \\
&= 2\{\mathrm{acov}(u_1, u_2)\}^2 + 4\mathrm{acov}(u_1, u_2)\mathrm{ase}(u_1)\mathrm{ase}(u_2)\delta_1\delta_2,
\end{aligned} \tag{A4}$$

where $\mathrm{E}\{(z_1-\delta_1)^4\} = 3$, $\mathrm{E}\{(z_1-\delta_1)^3e_{12}\} = 0$ and $\mathrm{E}\{(z_1-\delta_1)^2e_{12}^2\} = 1$ are used.

From (A2) and (A4),

$$\begin{aligned}
\mathrm{avar}\left(\sum_{i\geq j} u_{ij}^2\right) &= 2\mathbf{1}_{p^*}'(\mathrm{acov}(\mathbf{u}) \odot \mathrm{acov}(\mathbf{u}))\mathbf{1}_{p^*} + 4(\mathrm{ase}(\mathbf{u}) \odot \boldsymbol{\delta})'\mathrm{acov}(\mathbf{u})(\mathrm{ase}(\mathbf{u}) \odot \boldsymbol{\delta}) \\
&= 2\mathrm{tr}[\mathrm{acov}(\mathbf{u})\{\mathrm{acov}(\mathbf{u}) + 2(\mathrm{ase}(\mathbf{u}) \odot \boldsymbol{\delta})(\mathrm{ase}(\mathbf{u}) \odot \boldsymbol{\delta})'\}],
\end{aligned} \tag{A5}$$

where $\odot$ denotes a Hadamard (elementwise) product,

$$\mathrm{ase}(\mathbf{u}) = (\mathrm{ase}(u_{11}), \mathrm{ase}(u_{21}), \mathrm{ase}(u_{22}), \ldots, \mathrm{ase}(u_{p,p-1}), \mathrm{ase}(u_{pp}))'$$

and

$$\boldsymbol{\delta} = (\delta_{11}, \delta_{21}, \delta_{22}, \ldots, \delta_{p,p-1}, \delta_{pp})'.$$

When a model is true,

$$\boldsymbol{\delta} = \mathbf{0} \quad \text{and} \quad \text{avar}\left(\sum_{i \geq j} u_{ij}^2\right) = 2\text{tr}[\{\text{acov}(\mathbf{u})\}^2].$$

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Anderson, J.C., & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155–173.

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

Bentler, P.M. (1989). *EQS: A structural equations program.* Los Angeles, CA: BMDP Statistical Software.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Bentler, P.M., & Dijkstra, T. (1985). Efficient estimation via linealization in structural models. In P.R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 9–42). New York, NY: Elsevier.

Bollen K.A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51*, 375–377.

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

Bollen K.A. (1989b). A new incremental fit index for general equation models. *Sociological Methods & Research, 17*, 303–316.

Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*, 1–24. (Reprinted in D.J. Aigner & A.S. Goldberger (Eds.), *Latent variables in socioeconomic models*, 1977, pp. 205–226, Amsterdam: North Holland)

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

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

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147–167.

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

Gerbing, D.W., & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Girshick, M.A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics, 10*, 203–224.

Harman, H.H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.

Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research, 11*, 325–344.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Jöreskog, K.G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relations by the method of maximum likelihood.* Chicago, IL: International Educational Services.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.

Jöreskog, K.G., Sörbom, D., du Toit, S. & du Toit, M. (1999). *LISREL8: New statistical features.* Chicago, IL: Scientific Software International.

Kano, Y., Bentler, P.M., & Mooijaart, A. (1993). Additional information and precision of estimators in multivariate structural models. In K. Matusita, M.L. Puri, & T. Hayakawa (Eds.), *Statistical science and data analysis* (pp. 187–196). Zeist, The Netherlands: VSP International Science Publishers.

Maiti, S.S., & Mukherjee, B.N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika, 55*, 721–726.

Marsh, H.W., & Balla, J.R. (1994). Goodness of fit in confirmatory factor analysis: The effect of sample size and model parsimony. *Quality & Quantity, 28*, 185–217.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis: Effects of sample size. *Psychological Bulletin, 103*, 391–411.

Marsh, H.W., & Hau, K.T. (1996). Assessing goodness of fit: Is parsimony always desirable? *The Journal of Experimental Education, 64*, 364–390.

McDonald, R.P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97–103.
McDonald, R.P., & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247–255.
Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C.D. (1989). An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin*, *105*, 430–445.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, *54*, 1–21.
Satorra, A., & Bentler, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis*, *10*, 235–249.
Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
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.
Steiger, J.H. (1988). Aspects of person-machine communication in structural modeling of correlations and covariances. *Multivariate Behavioral Research*, *23*, 281–290.
Steiger, J.H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanstone, IL: SYSTAT.
Steiger, J.H. (1990). Structural model evaluation and modification.: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.
Steiger, J.H., & Hakstian, A.R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, *35*, 208–215.
Steiger, J.H., & Lind, J.C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual Spring Meeting of the Psychometric Society, Iowa City, IA.
Tanaka, J.S., & Huba, G.J. (1985). A fit index for covariance structural models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *38*, 197–201.
Tanaka, J.S., & Huba, G.J. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *42*, 233–239.
Yung, Y.F., & Bentler, P.M. (1999). On added information for ML factor analysis with mean and covariance structures. *Journal of Educational and Behavioral Statistics*, *24*, 1–20.

Manuscript received 30 NOV 1999
Final version received 10 OCT 2000