BIAS REDUCTION OF ESTIMATED STANDARD ERRORS IN FACTOR ANALYSIS

Haruhiko Ogasawara*

Formulas for the asymptotic biases of the estimators of the normal theory standard errors in factor analysis are given with and without the assumption of multivariate normality for observed variables. The biases are derived from the asymptotic variances of standard error estimators and the asymptotic biases of the estimated variances of parameter estimators. The latter biases are derived from the asymptotic variances/covariances and asymptotic biases of the parameter estimators. The formulas cover the cases for unstandardized and standardized variables. Numerical examples using factor analysis models show the accuracy of the formulas. The biases of standard error estimators are theoretically and empirically shown to be of the same order as that of the differences between the asymptotic standard errors neglecting higher-order terms and those considering them.

1. Introduction

Standard errors of parameter estimates are often used in structural equation modeling. In the familiar programs such as LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 1989) and Amos (Arbuckle & Wothke, 1999), the standard errors are available mostly as asymptotic ones for e.g., maximum likelihood estimators. For exploratory factor analysis with factor rotation, the standard errors of typical rotated solutions (e.g., the varimax solution) can be obtained by the programs CEFA (Browne, Cudeck, Tateneni & Mels, 1999), SAS (SAS, 2000) and ROSEF (Ogasawara, 2003) with the assumption of multivariate normality for observed variables. The assumption of multivariate normality can be relaxed without losing the validity of the normal theory standard errors for the models with asymptotic robustness against the violation of multivariate normality under some distributional conditions (see, e.g., Satorra, 2002 and the references therein).

The normal theory standard errors, however, have inaccuracy in finite samples even when the normality assumption or the condition for asymptotic robustness of a model is satisfied. Because the standard errors are obtained from a lower-order term of an asymptotic expansion, they are still approximate ones under the condition when the population parameters, on which the standard errors usually depend, are available as in experimental situations. Since in practice the population parameters are not available, they are estimated, which gives estimated asymptotic standard errors and consequently added inaccuracy of the standard errors. Ogasawara (2002) gave the general formulas for the asymptotic standard errors of estimated normal theory standard errors of maximum likelihood estimators in exploratory factor analysis and structural equation modeling.

∗ 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

The asymptotic biases of the estimated normal theory standard errors will be derived in this article. Using the results we can correct the estimated standard errors, which will reduce the inaccuracy of standard errors to some extent.

2. Order of the biases of the estimated standard errors

In this section it is shown that the asymptotic biases of the estimated standard errors are of the same order as that of the added accuracy of the estimated asymptotic standard errors by considering higher-order terms. Let θ_i be the i-th element of a $q \times 1$ population parameter vector $\boldsymbol{\theta}$ in a structural equation model and $\hat{\theta}_i$ be the corresponding estimator by some method e.g., maximum likelihood. Then, using the Taylor expansion of $\hat{\theta}_i$ about θ_i under the assumption that $\hat{\theta}_i$ is four-times differentiable, we have

$$
\hat{\theta}_i = \theta_i + \left.\frac{\partial \hat{\theta}_i}{\partial \mathbf{s}'}\right|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s}-\boldsymbol{\sigma}) + \frac{1}{2}(\mathbf{s}-\boldsymbol{\sigma})' \left.\frac{\partial^2 \hat{\theta}_i}{\partial \mathbf{s}\,\partial \mathbf{s}'}\right|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s}-\boldsymbol{\sigma})
$$

$$
+ \frac{1}{6}\sum_{a\geq b}\mathrm{tr}\left\{\left.\frac{\partial^3 \hat{\theta}_i}{\partial \mathbf{s}\,\partial \mathbf{s}'\partial s_{ab}}\right|_{\mathbf{s}=\boldsymbol{\sigma}}(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\right\}(s_{ab}-\sigma_{ab}) + O_p(n^{-2}),
$$

$$
(i=1,\ldots,q), \tag{1}
$$

where $\boldsymbol{\sigma} = \mathrm{v}(\boldsymbol{\Sigma})$; $\mathbf{s} = \mathrm{v}(\mathbf{S})$; $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\mathbf{S}$ are $p \times p$ population and unbiased sample covariance matrices, respectively; $\mathrm{v}(\cdot)$ denotes a $p(p+1)/2 \times 1$ vector taking nonduplicated elements of a $p \times p$ symmetric matrix (note that $\mathbf{s}$ is also used as a variable as well as an estimate for simplicity of notation); and $n+1=N$ is the number of observations in a sample. The usual asymptotic standard error of $\hat{\theta}_i$ is obtained from (1) as

$$
\mathrm{ase}(\hat{\theta}_i) = \sqrt{\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'}\mathrm{acov}(\mathbf{s}-\boldsymbol{\sigma})\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}}} = n^{-1/2}\sqrt{\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}}}, \tag{2}
$$

where $\boldsymbol{\Omega} = \mathrm{acov}\{\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})\}$ and $\partial\theta_i/\partial\boldsymbol{\sigma}$ denotes $\partial\hat{\theta}_i/\partial\mathbf{s}|_{\mathbf{s}=\boldsymbol{\sigma}}$ for simplicity of notation (similar notation will be used throughout this article). Since $\{\mathrm{ase}(\hat{\theta}_i)\}^2 = O(n^{-1})$ from (2), we write $\mathrm{ase}(\hat{\theta}_i) = \sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1})}$ when necessary. A more accurate asymptotic standard error can be obtained using higher-order terms as

$$
\mathrm{E}\{(\hat{\theta}_i - \theta_i)^2\} = n^{-1}\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'}\mathrm{E}\{n\,(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\}\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}}
$$

$$
+ \mathrm{E}\left[\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})\,\mathrm{tr}\left\{\frac{\partial^2 \theta_i}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\right\}\right.
$$

$$
+ \frac{1}{4}\left(\mathrm{tr}\left\{\frac{\partial^2 \theta_i}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\right\}\right)^2
$$

$$
\left. + \frac{1}{3}\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})\sum_{a\geq b}\mathrm{tr}\left(\frac{\partial^3 \theta_i}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'\partial \sigma_{ab}}(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\right)(s_{ab}-\sigma_{ab})\right]
$$

$$
+ O(n^{-3}), \quad (i=1,\ldots,q), \tag{3}
$$

where $\mathrm{E}\{n\,(\mathbf{s}-\boldsymbol{\sigma})(\mathbf{s}-\boldsymbol{\sigma})'\}$ should be evaluated up to order $O(n^{-1})$. The expectation of

the second term on the right-hand side of (3) is of order $O(n^{-2})$ because the term is the expectation of the sum of the cubic and quartic terms of $(s_{ab} - \sigma_{ab})$'s. Denote the asymptotic variance from (3) omitting the term of order $O(n^{-3})$ by $\mathrm{avar}(\hat{\theta}_i; n^{-1}, n^{-2})$. Then, the difference of the asymptotic standard errors using $\mathrm{avar}(\hat{\theta}_i; n^{-1}, n^{-2})$ and $\mathrm{avar}(\hat{\theta}_i; n^{-1})$ is

$$
\sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1}, n^{-2})} - \sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1})}
$$
$$
= \frac{\mathrm{avar}(\hat{\theta}_i; n^{-1}, n^{-2}) - \mathrm{avar}(\hat{\theta}_i; n^{-1})}{\sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1}, n^{-2})} + \sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1})}} = \frac{O(n^{-2})}{O(n^{-1/2})} = O(n^{-3/2}). \qquad (4)
$$

Equation (4) holds whether $\hat{\theta}_i$ is (first-order) bias corrected or not because the squared bias is of order $O(n^{-2})$ and we have the remaining terms of the same order in (3).

In practice, the usual asymptotic standard error is available only as a function of $\hat{\boldsymbol{\theta}}$, which gives an estimator $\widehat{\mathrm{ase}}(\hat{\theta}_i)$. Assume that $\widehat{\mathrm{ase}}(\hat{\theta}_i)$ like $\hat{\theta}_i$ is a function of only $\mathbf{s}$. Then, $\sqrt{n}$ times $\widehat{\mathrm{ase}}(\hat{\theta}_i)$ is expanded in a manner similar to (1) as

$$
\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i) = \sqrt{n}\,\mathrm{ase}(\hat{\theta}_i) + \frac{\partial \sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}'}(\mathbf{s} - \boldsymbol{\sigma})
$$
$$
+ \frac{1}{2}(\mathbf{s} - \boldsymbol{\sigma})' \frac{\partial^2 \sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'}(\mathbf{s} - \boldsymbol{\sigma}) + O_p(n^{-3/2}),
$$
$$
(i = 1, \ldots, q). \qquad (5)
$$

From (5) the asymptotic bias of $\widehat{\mathrm{ase}}(\hat{\theta}_i)$ is given as

$$
\mathrm{abis}\{\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\} = n^{-1/2}\frac{1}{2}\mathrm{E}\left[\mathrm{tr}\left\{\frac{\partial^2 \sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'}(\mathbf{s} - \boldsymbol{\sigma})(\mathbf{s} - \boldsymbol{\sigma})'\right\}\right]
$$
$$
= n^{-1/2}O(n^{-1}) = O(n^{-3/2}),
$$
$$
(i = 1, \ldots, q). \qquad (6)
$$

The result of (6) is obtained from the assumption that $\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i) = O_p(1)$ is an (implicit) function of $\mathbf{s}$ or $\mathbf{S}$. From (4) and (6), we find that the two sources of inaccuracy in $\sqrt{\widehat{\mathrm{avar}}(\hat{\theta}_i; n^{-1})}$ i.e.,

$$
\mathrm{E}\left\{\sqrt{\widehat{\mathrm{avar}}(\hat{\theta}_i; n^{-1})}\right\} - \sqrt{\mathrm{E}\{(\hat{\theta}_i - \theta_i)^2\}}
$$
$$
= \mathrm{E}\left\{\sqrt{\widehat{\mathrm{avar}}(\hat{\theta}_i; n^{-1})}\right\} - \sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1})}
$$
$$
- \left\{\sqrt{\mathrm{E}\{(\hat{\theta}_i - \theta_i)^2\}} - \sqrt{\mathrm{avar}(\hat{\theta}_i; n^{-1})}\right\} \qquad (7)
$$

are of the same order. It is to be noted that this equivalence does not depend on estimation methods or distributional assumptions of $\mathbf{S}$ as long as $\widehat{\mathrm{ase}}(\hat{\theta}_i)$ is a function of only $\mathbf{s}$. In a later section, it will be empirically shown that the absolute value of (6) can be greater than that of (4) and vice versa depending on parameters in a model.

3. Asymptotic biases of normal theory standard errors

The asymptotic biases of estimated asymptotic standard errors can be obtained from (6) when the partial derivatives and the expectation in (6) are available. In this section, we derive the asymptotic biases of the normal theory standard errors with and without the assumption of multivariate normality for observed variables.

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator which minimizes

$$F = \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \text{tr}\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}\} \tag{8}$$

with r restrictions $\mathbf{h} = \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}\,(r \times 1)$. Then, the usual normal theory asymptotic variance (i.e., $\text{avar}(\hat{\theta}_i) = \text{avar}(\hat{\theta}_i; n^{-1})$) of $\hat{\theta}_i, (i = 1, \ldots, q)$ is given by $n\,\text{avar}(\hat{\theta}_i) = 2(\mathbf{H}^*)_{ii}$ with

$$\mathbf{H}_A^{-1} = \left[\begin{array}{cc} \text{E}\left(\dfrac{\partial^2 F}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right) & \dfrac{\partial\mathbf{h}'}{\partial\boldsymbol{\theta}} \\ \dfrac{\partial\mathbf{h}}{\partial\boldsymbol{\theta}'} & \mathbf{O} \end{array} \right]^{-1} = \left[\begin{array}{cc} \mathbf{H}^* & \mathbf{H}_{\mathbf{h}'} \\ \mathbf{H}_{\mathbf{h}} & \# \end{array} \right], \tag{9}$$

where $(\cdot)_{ij}$ indicates the (i,j)th element of a matrix, $(1/2)\mathbf{H}_A$ is the augmented information matrix per observation with

$$\left\{\text{E}\left(\dfrac{\partial^2 F}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right)\right\}_{ij} = \text{tr}\left\{\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right\}, \quad (i,j = 1, \ldots, q), \tag{10}$$

(see e.g., Jennrich, 1974), $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ and the hash mark denotes a submatrix of $\mathbf{H}_A^{-1}$ which will not be used. Using the above notation, we have

Lemma 1. The asymptotic biases of estimated normal theory asymptotic variances of the maximum likelihood estimators $\hat{\boldsymbol{\theta}}$ are given by

$$\begin{aligned}
\text{abis}\{\widehat{\text{avar}}(\hat{\theta}_i)\} = n^{-1}\sum_{k=1}^{q}&\left[\sum_{l=1}^{q}\left(2\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\theta_k}\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\theta_l}\mathbf{H}_A^{-1}\right.\right. \\
&\left.\left. - \mathbf{H}_A^{-1}\dfrac{\partial^2\mathbf{H}_A}{\partial\theta_k\partial\theta_l}\mathbf{H}_A^{-1}\right)_{ii}\text{acov}(\hat{\theta}_k,\hat{\theta}_l) - \left(2\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\theta_k}\mathbf{H}_A^{-1}\right)_{ii}\text{abis}(\hat{\theta}_k)\right], \\
&(i = 1, \ldots, q).
\end{aligned} \tag{11}$$

Proof. Consider the expansion of $n\,\widehat{\text{avar}}(\hat{\theta}_i)$ similar to (5), which yields

$$\begin{aligned}
\text{abis}\{n\,\widehat{\text{avar}}(\hat{\theta}_i)\} &= \dfrac{n^{-1}}{2}\text{tr}\left\{\dfrac{\partial^2 n\,\text{avar}(\hat{\theta}_i)}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\right\} = \dfrac{n^{-1}}{2}\text{tr}\left\{\dfrac{\partial^2 2(\mathbf{H}^*)_{ii}}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\right\} \\
&= n^{-1}\sum_{a\geq b}\sum_{c\geq d}\left[\partial\left(-\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\sigma_{ab}}\mathbf{H}_A^{-1}\right)_{ii}\Big/\partial\sigma_{cd}\right](\boldsymbol{\Omega})_{ab,cd} \\
&= n^{-1}\sum_{a\geq b}\sum_{c\geq d}\left(2\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\sigma_{ab}}\mathbf{H}_A^{-1}\dfrac{\partial\mathbf{H}_A}{\partial\sigma_{cd}}\mathbf{H}_A^{-1} - \mathbf{H}_A^{-1}\dfrac{\partial^2\mathbf{H}_A}{\partial\sigma_{ab}\partial\sigma_{cd}}\mathbf{H}_A^{-1}\right)_{ii}(\boldsymbol{\Omega})_{ab,cd} \\
&= n^{-1}\sum_{a\geq b}\sum_{c\geq d}\left[2\mathbf{H}_A^{-1}\sum_{k=1}^{q}\dfrac{\partial\mathbf{H}_A}{\partial\theta_k}\dfrac{\partial\theta_k}{\partial\sigma_{ab}}\mathbf{H}_A^{-1}\sum_{l=1}^{q}\dfrac{\partial\mathbf{H}_A}{\partial\theta_l}\dfrac{\partial\theta_l}{\partial\sigma_{cd}}\mathbf{H}_A^{-1}\right.
\end{aligned}$$

$$- \mathbf{H}_A^{-1} \sum_{k=1}^{q} \left(\sum_{l=1}^{q} \left\{ \frac{\partial^2 \mathbf{H}_A}{\partial \theta_k \partial \theta_l} \frac{\partial \theta_k}{\partial \sigma_{ab}} \frac{\partial \theta_l}{\partial \sigma_{cd}} \right\} + \frac{\partial \mathbf{H}_A}{\partial \theta_k} \frac{\partial^2 \theta_k}{\partial \sigma_{ab} \partial \sigma_{cd}} \right) \mathbf{H}_A^{-1} \Bigg]_{ii} (\mathbf{\Omega})_{ab,cd}$$

$$= n^{-1} \sum_{k=1}^{q} \Bigg[\sum_{l=1}^{q} \left\{ \left(2\mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_k} \mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_l} \mathbf{H}_A^{-1} - \mathbf{H}_A^{-1} \frac{\partial^2 \mathbf{H}_A}{\partial \theta_k \partial \theta_l} \mathbf{H}_A^{-1} \right)_{ii} \right.$$

$$\times \sum_{a \geq b} \sum_{c \geq d} \frac{\partial \theta_k}{\partial \sigma_{ab}} \frac{\partial \theta_l}{\partial \sigma_{cd}} (\mathbf{\Omega})_{ab,cd} \Bigg\}$$

$$-2 \left(\mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_k} \mathbf{H}_A^{-1} \right)_{ii} \frac{1}{2} \sum_{a \geq b} \sum_{c \geq d} \frac{\partial^2 \theta_k}{\partial \sigma_{ab} \partial \sigma_{cd}} (\mathbf{\Omega})_{ab,cd} \Bigg], \tag{12}$$

which gives (11). Q.E.D.

In (11) the first and second partial derivatives with respect to the parameters are given as follows:

$$\left(\frac{\partial \mathbf{H}_A}{\partial \theta_k} \right)_{ij} = \text{tr} \left\{ -2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \right.$$

$$+ \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_k} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} + \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_k} \right\},$$

$$(i,j,k = 1,\ldots,q), \tag{13a}$$

$$\left(\frac{\partial \mathbf{H}_A}{\partial \theta_k} \right)_{q+i,j} = \left(\frac{\partial \mathbf{H}_A}{\partial \theta_k} \right)_{j,q+i} = \frac{\partial^2 h_i}{\partial \theta_j \partial \theta_k},$$

$$(i = 1,\ldots,r; \ j,k = 1,\ldots,q), \tag{13b}$$

where $h_i = h_i(\boldsymbol{\theta})$ is the i-th element of $\mathbf{h}(\boldsymbol{\theta})$,

$$\left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_k \partial \theta_l} \right)_{ij} = \text{tr} \left\{ 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_l} \right.$$

$$+ 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k}$$

$$+ 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} - 2\mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k}$$

$$- 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} - 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_k \partial \theta_l}$$

$$- 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_k} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} - 2\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_k}$$

$$+ \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_k} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_l} + \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_k}$$

$$+ \mathbf{\Sigma}^{-1} \frac{\partial^3 \mathbf{\Sigma}}{\partial \theta_i \partial \theta_k \partial \theta_l} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} + \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_i} \mathbf{\Sigma}^{-1} \frac{\partial^3 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\},$$

$$(i,j,k,l = 1,\ldots,q) \tag{13c}$$

and

$$\left(\frac{\partial^2 \mathbf{H}_A}{\partial\,\theta_k \partial\,\theta_l}\right)_{q+i,j} = \left(\frac{\partial^2 \mathbf{H}_A}{\partial\,\theta_k \partial\,\theta_l}\right)_{j,q+i} = \frac{\partial^3 h_{\,i}}{\partial\,\theta_j \partial\,\theta_k \partial\,\theta_l},$$
$$(i = 1, \ldots, r; \ j, k, l = 1, \ldots, q). \tag{13d}$$

In (13a) and (13c), the partial derivatives of $\boldsymbol{\Sigma}$ with respect to the parameters are easily obtained because usually $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is an explicit function of $\boldsymbol{\theta}$. The second derivatives vanish for linear models of $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\theta}$. For exploratory factor analysis, the third derivatives in (13c) vanish for the model with fixed variances-covariances of common factors as the usual orthogonal factor model. The results of (13b) and (13d) depend on the restrictions $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ employed. One of the remaining factors to be evaluated in (11) is

$$n\operatorname{acov}(\hat{\theta}_k, \hat{\theta}_l) = \sum_{a \geq b} \sum_{c \geq d} \frac{\partial\,\theta_k}{\partial\,\sigma_{ab}} \frac{\partial\,\theta_l}{\partial\,\sigma_{cd}} (\boldsymbol{\Omega})_{ab,cd} = \frac{\partial\,\theta_k}{\partial\,\boldsymbol{\sigma}'} \boldsymbol{\Omega} \frac{\partial\,\theta_l}{\partial\,\boldsymbol{\sigma}},$$
$$(k, l = 1, \ldots, q). \tag{14}$$

For nonnormal distributions, (14) is given from the partial derivatives in implicit functions

$$\frac{\partial\,\boldsymbol{\theta}}{\partial\,\boldsymbol{\sigma}'} = -\mathbf{H}^* \frac{\partial\,\mathbf{g_s}}{\partial\,\mathbf{s}'}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}} \quad \text{and} \quad (\boldsymbol{\Omega})_{ab,cd} = \sigma_{abcd} - \sigma_{ab}\sigma_{cd} \tag{15}$$

with σ_{abcd} being the fourth-order central moment of the (a, b, c, d)th variables. In (15), $\mathbf{g_s}$ is the gradient vector with

$$(\mathbf{g_s})_{i\cdot} = \frac{\partial\,F}{\partial\,\hat{\theta}_i} = \operatorname{tr}\left\{\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\,\hat{\boldsymbol{\Sigma}}}{\partial\,\hat{\theta}_i}\right\}, \quad (i = 1, \ldots, q), \tag{16}$$

where $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ and $(\cdot)_{i\cdot}$ is the i-th element (row) of a vector (matrix) and the subscript $\mathbf{s}$ in $\mathbf{g_s}$ denotes that the subscripted term is regarded as a function of $\mathbf{s}$ or $\mathbf{S}$ explicitly involved in $\mathbf{g_s}$ while $\hat{\boldsymbol{\theta}}$ is temporarily regarded as a constant. That is,

$$\left(\frac{\partial\,\mathbf{g_s}}{\partial\,\mathbf{s}'}\right)_{i\cdot} = -\left(\operatorname{vec}\left(\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\,\hat{\boldsymbol{\Sigma}}}{\partial\,\hat{\theta}_i}\hat{\boldsymbol{\Sigma}}^{-1}\right)\right)' \mathbf{D}_p = -\frac{\partial\,(\operatorname{vec}\hat{\boldsymbol{\Sigma}})'}{\partial\,\hat{\theta}_i}(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1})\mathbf{D}_p, \tag{17}$$

where $\operatorname{vec}\mathbf{X} = \operatorname{vec}(\mathbf{X})$ is a vectorizing operator stacking the columns of a matrix $\mathbf{X}$; $\mathbf{D}_p$ is the duplication matrix for a $p \times p$ symmetric matrix with $\operatorname{vec}(\mathbf{S}) = \mathbf{D}_p\operatorname{v}(\mathbf{S})$; and $\otimes$ denotes the right Kronecker product.

From (15) and (17), we have

$$\frac{\partial\,\boldsymbol{\theta}}{\partial\,\boldsymbol{\sigma}'} = \mathbf{H}^* \frac{\partial\,(\operatorname{vec}\boldsymbol{\Sigma})'}{\partial\,\boldsymbol{\theta}}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{D}_p. \tag{18}$$

For normal data, (14) becomes simply, as was shown before,

$$n\operatorname{acov}(\hat{\boldsymbol{\theta}}) = 2\mathbf{H}^*. \tag{19}$$

The remaining factor $\operatorname{abis}(\hat{\theta}_i)$ to be evaluated in (11) is obtained from Theorems 1 and 2

of Ogasawara (2004) for nonnormal and normal data, respectively. That is, for nonnormal data,

$$
\begin{aligned}
\text{abis}(\hat{\theta}_i) = & -\frac{n^{-1}}{2} \sum_{a \geq b} \sum_{c \geq d} \sum_{j=1}^{q} \sum_{k=1}^{q} (2 - \delta_{cd}) \left\{ \left(\mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_k} \mathbf{H}_A^{-1} \right)_{ij} \left(\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \right)_{cd} \right. \\
& + (\mathbf{H}^*)_{ij} \left(\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_j \partial \theta_k} \mathbf{\Sigma}^{-1} \right. \\
& + \left. \left. \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{\Sigma}^{-1} \right)_{cd} \right\} \frac{\partial \theta_k}{\partial \sigma_{ab}} (\sigma_{abcd} - \sigma_{ab}\sigma_{cd}), \quad (i = 1, \ldots, q), \quad (20)
\end{aligned}
$$

where δ_{cd} is the Kronecker delta, and for normal data

$$
\begin{aligned}
\text{abis}(\hat{\theta}_i) = -n^{-1} \sum_{k=1}^{q} \left[\mathbf{H}_A^{-1} \left(\begin{array}{c} \left\{ \text{tr} \left(\mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_u} \mathbf{\Sigma}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_k \partial \theta_v} \right) \right\}_{(u,v)\text{th}} \\ \frac{\partial^2 \mathbf{h}}{\partial \boldsymbol{\theta}' \partial \theta_k} \end{array} \right) \mathbf{H}^* \right]_{ik}, \\
(i = 1, \ldots, q), \quad (21)
\end{aligned}
$$

where $\{\cdot\}_{(u,v)\text{th}}$ is the $q \times q$ matrix whose (u,v)th element is given by the expression in braces. The special cases of (21) for the estimators of the variances of unique factors in exploratory factor analysis have also been given by Ihara (1985) when $p = 3$ and Ichikawa and Konishi (2002) for general p's.

The asymptotic biases of estimated standard errors $\widehat{\text{ase}}(\hat{\theta}_i) = \sqrt{\widehat{\text{avar}}(\hat{\theta}_i; n^{-1})}$ are given by the following theorem with Lemma 1 for $\text{abis}\{n \, \widehat{\text{avar}}(\hat{\theta}_i)\}$.

Theorem 1. The asymptotic biases of estimated normal theory standard errors of maximum likelihood estimators $\hat{\theta}_i, (i = 1, \ldots, q)$ are

$$
\begin{aligned}
\text{abis}\{\sqrt{n} \, \widehat{\text{ase}}(\hat{\theta}_i)\} = \frac{1}{2\sqrt{n} \, \text{ase}(\hat{\theta}_i)} \left[-\text{avar}\{\sqrt{n} \, \widehat{\text{ase}}(\hat{\theta}_i)\} + \text{abis}\{n \, \widehat{\text{avar}}(\hat{\theta}_i)\} \right], \\
(i = 1, \ldots, q). \quad (22)
\end{aligned}
$$

Proof 1.

$$
\begin{aligned}
\text{abis}\{\sqrt{n} \, \widehat{\text{ase}}(\hat{\theta}_i)\} = & \frac{n^{-1}}{2} \text{tr} \left\{ \frac{\partial^2 \sqrt{n \, \text{avar}(\hat{\theta}_i)}}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\sigma}'} \mathbf{\Omega} \right\} \\
= & \frac{n^{-1}}{2} \text{tr} \left\{ \left[\frac{\partial}{\partial \boldsymbol{\sigma}} \left(\frac{\{n \, \text{avar}(\hat{\theta}_i)\}^{-1/2}}{2} \frac{\partial n \, \text{avar}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}'} \right) \right] \mathbf{\Omega} \right\} \\
= & \frac{n^{-1}}{2} \text{tr} \left\{ \left[-\frac{\{n \, \text{avar}(\hat{\theta}_i)\}^{-3/2}}{4} \frac{\partial n \, \text{avar}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}} \frac{\partial n \, \text{avar}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma}'} \right. \right. \\
& + \left. \left. \frac{\{n \, \text{avar}(\hat{\theta}_i)\}^{-1/2}}{2} \frac{\partial^2 n \, \text{avar}(\hat{\theta}_i)}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\sigma}'} \right] \mathbf{\Omega} \right\} \\
= & -\frac{\{n \, \text{avar}(\hat{\theta}_i)\}^{-3/2}}{8} \text{avar}\{n \, \widehat{\text{avar}}(\hat{\theta}_i)\} + \frac{\{n \, \text{avar}(\hat{\theta}_i)\}^{-1/2}}{2} \text{abis}\{n \, \widehat{\text{avar}}(\hat{\theta}_i)\}, \\
& (i = 1, \ldots, q). \quad (23)
\end{aligned}
$$

which yields (22). Q. E. D.

Proof 2. Let u_i and v_i be the linear and quadratic terms of $\mathbf{s} - \boldsymbol{\sigma}$ in the Taylor expansion of $\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)$, respectively. Then, omitting higher-order terms when necessary, we have

$$
\begin{aligned}
n\,\widehat{\mathrm{avar}}(\hat{\theta}_i) &= \{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\}^2 = \{\sqrt{n}\,\mathrm{ase}(\hat{\theta}_i) + u_i + v_i\}^2 \\
&= n\,\mathrm{avar}(\hat{\theta}_i) + 2\sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)u_i + u_i^2 + 2\sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)v_i, \quad (i = 1, \ldots, q). \quad (24)
\end{aligned}
$$

Taking the expectation of (24) with respect to the distribution of $\mathbf{s}$ in large samples,

$$
\begin{aligned}
&n\,\mathrm{avar}(\hat{\theta}_i) + \mathrm{abis}\{n\,\widehat{\mathrm{avar}}(\hat{\theta}_i)\} \\
&= n\,\mathrm{avar}(\hat{\theta}_i) + \mathrm{avar}\{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\} + 2\sqrt{n}\,\mathrm{ase}(\hat{\theta}_i)\,\mathrm{abis}\{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\}, \\
&(i = 1, \ldots, q),
\end{aligned} \quad (25)
$$

which gives (22). Q. E. D.

The above result is also given by taking the terms of order $O(n^{-1})$ in both sides of the following exact equation:

$$
\mathrm{E}(\hat{g} - g) = \frac{1}{2g}[-\mathrm{E}\{(\hat{g} - g)^2\} + \mathrm{E}(\hat{g}^2) - g^2], \quad (26)
$$

where $g = g(\boldsymbol{\theta})$ is any function of $\boldsymbol{\theta}$ with $\hat{g} = g(\hat{\boldsymbol{\theta}})$ (the author is indebted to a reviewer for this point).

In (22) $\mathrm{avar}\{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\}$ was derived by Ogasawara (2002) for normal data. In our case, however, $\mathrm{avar}\{n\,\widehat{\mathrm{avar}}(\hat{\theta}_i)\}$ in (23) is slightly more convenient for evaluation of $\mathrm{abis}\{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\}$ than $\mathrm{avar}\{\sqrt{n}\,\widehat{\mathrm{ase}}(\hat{\theta}_i)\}$ in (22). That is,

$$
n\,\mathrm{avar}\{n\,\widehat{\mathrm{avar}}(\hat{\theta}_i)\} = \frac{\partial\,2(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\,2(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\sigma}}, \quad (i = 1, \ldots, q), \quad (27)
$$

where

$$
\begin{aligned}
\frac{\partial\,(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\sigma}'} &= \frac{\partial\,(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\theta}'}\frac{\partial\,\boldsymbol{\theta}}{\partial\,\boldsymbol{\sigma}'} \\
&= -\sum_{k=1}^{q}\left(\mathbf{H}_A^{-1}\frac{\partial\,\mathbf{H}_A}{\partial\,\theta_k}\mathbf{H}_A^{-1}\right)_{ii}(\mathbf{H}^*)_{k\cdot}\frac{\partial\,(\mathrm{vec}\,\boldsymbol{\Sigma})'}{\partial\,\boldsymbol{\theta}}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\mathbf{D}_p. \quad (28)
\end{aligned}
$$

In (28), $\partial\mathbf{H}_A/\partial\theta_k$ is available from (13a) and (13b). Equation (27) holds for nonnormal data as well as normal ones, which is a slight generalization of Ogasawara (2002). For normal data, as was derived by Ogasawara (2002), (27) becomes simply

$$
n\,\mathrm{avar}\{n\,\widehat{\mathrm{avar}}(\hat{\theta}_i)\} = 8\frac{\partial\,(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\theta}'}\mathbf{H}^*\frac{\partial\,(\mathbf{H}^*)_{ii}}{\partial\,\boldsymbol{\theta}}, \quad (i = 1, \ldots, q). \quad (29)
$$

4. Asymptotic biases of the estimated standard errors for models of standardized observed variables

In the behavioral sciences observed variables are frequently standardized. That is,

sample correlation matrices are often used instead of sample covariance matrices. In exploratory factor analysis, it is known that the estimator $\hat{\boldsymbol{\theta}}$ which minimizes (8) with identification restrictions , when $\mathbf{S}$ is replaced by the corresponding sample correlation matrix $\mathbf{R}$, is still a maximum likelihood estimator with the covariance structure model:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{D}\mathbf{P}(\boldsymbol{\theta}_\rho)\mathbf{D}, \ \mathbf{D} = \operatorname{diag}(d_1, \ldots, d_p) \quad \text{and} \quad \operatorname{Diag}\{\mathbf{P}(\boldsymbol{\theta}_\rho)\} = \mathbf{I}_{(p)}, \qquad (30)$$

where $\operatorname{diag}(\cdot)$ is the diagonal matrix whose diagonal elements are given in parentheses, $\operatorname{Diag}(\cdot)$ is the diagonal matrix whose diagonal elements are those of the argument matrix, and $\mathbf{I}_{(p)}$ is the $p \times p$ identity matrix. The parameter vector $\boldsymbol{\theta} = (d_1, \ldots, d_p, \boldsymbol{\theta}_\rho')'$ includes the nuisance parameters $d_1, \ldots, d_p$ the standard deviations of the observed variables, and the parameters $\boldsymbol{\theta}_\rho(q \times 1)$ of interest in the correlation structure $\mathbf{P}(\boldsymbol{\theta}_\rho)$. The results in the previous section hold for the models of (30) when we consider the nuisance parameters and associated partial derivatives. That is, the matrix $\mathbf{H}_A$ should be expanded to a $(p+q+r+p) \times (p+q+r+p)$ matrix, where the increase of the size by $2p$ comes from the added p nuisance parameters and the additional p restrictions in (30) on $\mathbf{P}(\boldsymbol{\theta}_\rho)$.

Let the expanded $\mathbf{H}_A$ be partitioned with submatrices for the expected second derivatives of F with respect to $\boldsymbol{\theta}$ followed by those of the restrictions $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$ and $[\{\mathbf{P}(\boldsymbol{\theta}_\rho)\}_{11} - 1, \ldots, \{\mathbf{P}(\boldsymbol{\theta}_\rho)\}_{pp} - 1]' = \mathbf{0}$ in this order. Then, fortunately, it is known that the diagonal elements of $\mathbf{H}_A^{-1}$ associated with the asymptotic variances of $\hat{\boldsymbol{\theta}}_\rho$, i.e., $(\mathbf{H}_A^{-1})_{ii}, (i = p+1, \ldots, p+q)$ do not depend on the values of $d_1, \ldots, d_p$ (see Yuan & Bentler, 2000 and Ogasawara, 2002) though $\hat{\boldsymbol{\theta}}_\rho$ is correlated with $\hat{d}_1, \ldots, \hat{d}_p$. Consequently, we can simplify the result of Lemma 1 as follows.

Lemma 2. For the model of (30) for standardized variables, the asymptotic biases of the estimated variances of the maximum likelihood estimators in a correlation structure are given by

$$
\begin{aligned}
&\operatorname{abis}\{\widehat{\operatorname{avar}}(\hat{\theta}_{\rho i})\} \\
&= n^{-1} \sum_{k=1}^{q} \left[\sum_{l=1}^{q} \left(2\mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_{\rho l}} \mathbf{H}_A^{-1} \right. \right. \\
&\quad \left. \left. - \mathbf{H}_A^{-1} \frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \mathbf{H}_A^{-1} \right)_{p+i,\, p+i} \operatorname{acov}(\hat{\theta}_{\rho k}, \hat{\theta}_{\rho l}) - \left(2\mathbf{H}_A^{-1} \frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \mathbf{H}_A^{-1} \right)_{p+i,\, p+i} \operatorname{abis}(\hat{\theta}_{\rho k}) \right], \\
&(i = 1, \ldots, q),
\end{aligned} \qquad (31)
$$

where $\theta_{\rho i} = (\boldsymbol{\theta}_\rho)_i.$ and $\mathbf{H}_A$ is the $(2p+q+r) \times (2p+q+r)$ augmented matrix of expected second derivatives.

Proof. The proof will be given in Appendix.

When we use Lemma 2, the following additional partial derivatives of the submatrices in $\mathbf{H}_A$ for the nuisance parameters $d_1, \ldots, d_p$ are required. The first group is for the submatrix for $\operatorname{E}(\partial^2 F / \partial \mathbf{d} \partial \mathbf{d}')$ (note that F corresponds to (8) not (A1)), where $\mathbf{d} = (d_1, \ldots, d_p)'$:

$$d_i d_j \left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{ij} = \frac{\partial 2(\rho^{ij} \rho_{ij} + \delta_{ij})}{\partial \theta_{\rho k}} = -2 \left(\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \right)_{ij} \rho_{ij} + 2\rho^{ij} \frac{\partial \rho_{ij}}{\partial \theta_{\rho k}},$$

$$(i, j = 1, \ldots, p;\ k = 1, \ldots, q), \tag{32}$$

where $\rho^{ij} = (\mathbf{P}^{-1})_{ij}$, $\mathbf{P}^{-1} = \mathbf{P}(\boldsymbol{\theta}_\rho)^{-1}$ and δ_{ij} is the Kronecker delta;

$$
\begin{aligned}
d_i d_j \left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{ij} = \ & 2 \left(\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}}\, \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} + \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}} \mathbf{P}^{-1} \right. \\
& \left. - \mathbf{P}^{-1} \frac{\partial^2 \mathbf{P}}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \mathbf{P}^{-1} \right)_{ij} \rho_{ij} - 2 \left(\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \right)_{ij} \frac{\partial \rho_{ij}}{\partial \theta_{\rho l}} \\
& - 2 \left(\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}} \mathbf{P}^{-1} \right)_{ij} \frac{\partial \rho_{ij}}{\partial \theta_{\rho k}} + 2 \rho^{ij} \frac{\partial^2 \rho_{ij}}{\partial \theta_{\rho k} \partial \theta_{\rho l}}, \\
& (i, j = 1, \ldots, p;\ k, l = 1, \ldots, q).
\end{aligned}
\tag{33}
$$

The second group is for the submatrix $\mathrm{E}(\partial^2 F / \partial \boldsymbol{\theta}_\rho \partial \mathbf{d}')$ and its transpose:

$$
\begin{aligned}
d_j \left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{p+i,j} = \ & d_j \left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{j,p+i} = \frac{\partial\, 2 (\mathbf{P}^{-1} \partial \mathbf{P} / \partial \theta_{\rho i})_{jj}}{\partial \theta_{\rho k}} \\
= \ & 2 \left(-\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho i}} + \mathbf{P}^{-1} \frac{\partial^2 \mathbf{P}}{\partial \theta_{\rho i} \partial \theta_{\rho k}} \right)_{jj}, \\
& (i, k = 1, \ldots, q;\ j = 1, \ldots, p),
\end{aligned}
\tag{34}
$$

$$
\begin{aligned}
d_j \left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{p+i,j} = \ & d_j \left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{j,p+i} \\
= \ & 2 \left(\mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho i}} + \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho i}} \right. \\
& - \mathbf{P}^{-1} \frac{\partial^2 \mathbf{P}}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho i}} - \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho k}} \mathbf{P}^{-1} \frac{\partial^2 \mathbf{P}}{\partial \theta_{\rho i} \partial \theta_{\rho l}} \\
& \left. - \mathbf{P}^{-1} \frac{\partial \mathbf{P}}{\partial \theta_{\rho l}} \mathbf{P}^{-1} \frac{\partial^2 \mathbf{P}}{\partial \theta_{\rho i} \partial \theta_{\rho k}} + \mathbf{P}^{-1} \frac{\partial^3 \mathbf{P}}{\partial \theta_{\rho i} \partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{jj}, \\
& (i, k, l = 1, \ldots, q;\ j = 1, \ldots, p).
\end{aligned}
\tag{35}
$$

Note that d_i and d_j in (32)–(35) can be set to one without loss of generality after taking the partial derivatives.

The third group is for the restrictions $\mathrm{Diag}\{\mathbf{P}(\boldsymbol{\theta}_\rho)\} = \mathbf{I}_{(p)}$,:

$$
\begin{aligned}
\left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{p+q+r+i,p+j} = \ & \left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{p+j,p+q+r+i} = \frac{\partial \{\partial (\rho_{ii} - 1) / \partial \theta_{\rho j}\}}{\partial \theta_{\rho k}} = \frac{\partial^2 \rho_{ii}}{\partial \theta_{\rho j} \partial \theta_{\rho k}}, \\
& (i = 1, \ldots, p;\ j, k = 1, \ldots, q),
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
\left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{p+q+r+i,p+j} = \ & \left(\frac{\partial^2 \mathbf{H}_A}{\partial \theta_{\rho k} \partial \theta_{\rho l}} \right)_{p+j,p+q+r+i} = \frac{\partial^3 \rho_{ii}}{\partial \theta_{\rho j} \partial \theta_{\rho k} \partial \theta_{\rho l}}, \\
& (i = 1, \ldots, p;\ j, k, l = 1, \ldots, q).
\end{aligned}
\tag{37}
$$

Note that $\left(\frac{\partial \mathbf{H}_A}{\partial \theta_{\rho k}} \right)_{p+i,p+j}$, $(i, j = 1, \ldots, q + r;\ k = 1, \ldots, q)$ are obtained from (13a)–(13d) by replacing $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with $\mathbf{P}(\boldsymbol{\theta}_\rho)$ because

$$(\mathbf{H}_A)_{p+i,p+j} = \text{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_{\rho i}}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_{\rho j}}\right) = \text{tr}\left(\mathbf{P}^{-1}\frac{\partial\mathbf{P}}{\partial\theta_{\rho i}}\mathbf{P}^{-1}\frac{\partial\mathbf{P}}{\partial\theta_{\rho j}}\right),$$
$$(i,j = 1,\ldots,q), \tag{38}$$

when $\boldsymbol{\Sigma} = \mathbf{D}\mathbf{P}(\boldsymbol{\theta}_\rho)\mathbf{D}$.

5. The raw orthomax solution

As an example in exploratory factor analysis, the case of the raw orthomax solution with m common factors for unstandardized/standardized observed variables is presented. The covariance structure model for unstandardized variables is written by

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}, \tag{39}$$

where $\boldsymbol{\Lambda}$ is a $p \times m$ loading matrix and $\boldsymbol{\Psi}$ is a covariance matrix of uncorrelated unique factors. For the correlation structure model, $\boldsymbol{\Sigma}$ in (39) should be replaced by $\mathbf{P}$. The partial derivatives $\partial\boldsymbol{\Sigma}/\partial\boldsymbol{\theta}$, $\partial^2\boldsymbol{\Sigma}/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ in (13a), (13c), and $\partial\mathbf{P}/\partial\boldsymbol{\theta}_\rho$, $\partial^2\mathbf{P}/\partial\boldsymbol{\theta}_\rho\partial\boldsymbol{\theta}_\rho'$ in (32)–(36) and (38) are easily obtained (note that the third derivatives in (35) and (37) vanish).

The partial derivatives in (13b) and (13d) of the identification restrictions for the raw orthomax solution are given as follows. The restrictions were given by Archer and Jennrich (1973) as

$$\left(\boldsymbol{\Lambda}'\frac{\partial f}{\partial\boldsymbol{\Lambda}} - \frac{\partial f}{\partial\boldsymbol{\Lambda}'}\boldsymbol{\Lambda}\right)_{st} = 0, \quad (1 \le s < t \le m) \tag{40}$$

with $f = \frac{1}{4}\sum_{j=1}^{m}\left\{p\sum_{i=1}^{p}\lambda_{ij}^4 - w\left(\sum_{i=1}^{p}\lambda_{ij}^2\right)^2\right\}$, $\lambda_{ij} = (\boldsymbol{\Lambda})_{ij}$, and w being the orthomax weight. The first derivatives of (40) with respect to $\lambda_{ij}, (i = 1,\ldots,p;\ j = 1,\ldots,m)$ were also given by Archer and Jennrich (1973). For the second derivatives, see Ogasawara (2002). The third derivatives required in (13d) (note that $h_i, (i = 1,\ldots,(m^2 - m)/2)$ in (8) correspond to (40)) are given as follows:

$$\partial^3\left(\boldsymbol{\Lambda}'\frac{\partial f}{\partial\boldsymbol{\Lambda}} - \frac{\partial f}{\partial\boldsymbol{\Lambda}'}\boldsymbol{\Lambda}\right)_{st}\Big/\partial\lambda_{ab}\partial\lambda_{cd}\partial\lambda_{ef}$$
$$= (\delta_{sb}\delta_{td}\delta_{tf} + \delta_{tb}\delta_{sd}\delta_{tf} + \delta_{tb}\delta_{td}\delta_{sf})\frac{\partial^3 f}{\partial\lambda_{at}\partial\lambda_{ct}\partial\lambda_{et}}$$
$$+ \delta_{tb}\delta_{td}\delta_{tf}\left(\boldsymbol{\Lambda}'\frac{\partial^4 f}{\partial\boldsymbol{\Lambda}\,\partial\lambda_{at}\partial\lambda_{ct}\partial\lambda_{et}}\right)_{st}$$
$$- (\delta_{tb}\delta_{sd}\delta_{sf} + \delta_{sb}\delta_{td}\delta_{sf} + \delta_{sb}\delta_{sd}\delta_{tf})\frac{\partial^3 f}{\partial\lambda_{as}\partial\lambda_{cs}\partial\lambda_{es}}$$
$$- \delta_{sb}\delta_{sd}\delta_{sf}\left(\boldsymbol{\Lambda}'\frac{\partial^4 f}{\partial\boldsymbol{\Lambda}\,\partial\lambda_{as}\partial\lambda_{cs}\partial\lambda_{es}}\right)_{ts},$$
$$(1 \le s < t \le m;\ a,c,e = 1,\ldots,p;\ b,d,f = 1,\ldots,m), \tag{41}$$

where

$$\frac{\partial^3 f}{\partial \lambda_{at} \partial \lambda_{ct} \partial \lambda_{et}} = \partial \left(\frac{\partial \{p\lambda_{at}^3 - w(\sum_{i=1}^{p} \lambda_{it}^2)\lambda_{at}\}}{\partial \lambda_{ct}} \right) \Big/ \partial \lambda_{et}$$

$$= \partial \{3\delta_{ac} p \lambda_{at}^2 - w(2\lambda_{ct}\lambda_{at} + \delta_{ac} \sum_{i=1}^{p} \lambda_{it}^2)\}/\partial \lambda_{et}$$

$$= 6\delta_{ac}\delta_{ae} p \lambda_{at} - 2w(\delta_{ce}\lambda_{at} + \delta_{ae}\lambda_{ct} + \delta_{ac}\lambda_{et}) \tag{42}$$

and

$$\frac{\partial^4 f}{\partial \lambda_{at} \partial \lambda_{ct} \partial \lambda_{et} \partial \lambda_{gt}} = 6\delta_{ac}\delta_{ae}\delta_{ag} p - 2w(\delta_{ce}\delta_{ga} + \delta_{ae}\delta_{gc} + \delta_{ac}\delta_{ge}),$$

$$(1 \le s < t \le m; \ a,c,e,g = 1,\ldots,p). \tag{43}$$

In practice, the normal orthomax solution (i.e., the orthomax solution with row or Kaiser's normalization) is often used. However, the partial derivatives corresponding to the above results become involved and are not pursued here (for the first derivatives, see Ogasawara, 2000 and Yuan & Bentler, 2000).

6. Numerical examples with simulations

Three numerical examples in exploratory factor analysis under normality are presented in the first half of this section, and an example in confirmatory factor analysis in the second section. The first example is an artificial one. The second and third examples are based on real data. The first example employs a one-factor model with equal population loadings and unique variances for six observed variables as

$$\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Psi} \tag{44}$$

with $\boldsymbol{\lambda} = (.6, .6, .6, .6, .6, .6)'$ and $\boldsymbol{\Psi} = \mathrm{diag}(.64, .64, .64, .64, .64, .64)$. When the model for standardized variables is used, $\boldsymbol{\Sigma}$ in (44) is replaced by $\mathbf{P}$ with $\mathbf{D} = \mathbf{I}_{(6)}$ in the population.

The results are shown in Table 1, where the standard errors (SEs), the standard errors of estimated standard errors (SEs of SE), and the biases of estimated standard errors (SE biases) are given. These values have each theoretical and simulated ones. The theoretical values are asymptotic ones using the population parameters with the assumption of $N = 300$ while the simulated values have been given in the following way. First, independent observations of $N = 300$ following the multivariate normal distribution with (44) were randomly generated. From these observations, sample covariance/correlation matrices were calculated followed by estimation of the parameters in the model by maximum likelihood. The Heywood case, if any, was removed. The model has a minor indeterminacy of the sign of the single common factor, which was removed by reversing the sign when the sum of the loading estimates was negative. This procedure was replicated until regular 1,000,000 sets of estimates were obtained. No Heywood case had occurred until the required number of replications was obtained. The simulated SEs were given from the square roots of the usual unbiased sample variances of the 1,000,000 estimates for each parameter. The simulated SEs of SE are the simulated SEs of estimated asymptotic

Table 1: SE biases of the one-factor model under normality ($N = 300$; Number of replications $= 1,000,000$)

	Population parameter value	A	A′	B	B′	C	C′
		SE $\times 10^5$		SE of SE $\times 10^5$		SE bias $\times 10^6$	
		Th.	Sim.	Th.	Sim.	Th.	Sim.
Unstandardized variables							
Ψ	1 .64	6303	6318	496	496	−327	−328
	2 .64	6303	6323	496	496	−327	−318
$(A - \bar{A}')/\bar{A}'$	3 .64	6303	6314	496	495	−327	−324
$= -.0026$	4 .64	6303	6325	496	496	−327	−333
$C/A = -.0052$	5 .64	6303	6317	496	496	−327	−317
$C/B = -.066$	6 .64	6303	6320	496	496	−327	−319
λ	1 .6	5999	6024	238	238	−206	−209
	2 .6	5999	6027	238	238	−206	−204
$(A - \bar{A}')/\bar{A}'$	3 .6	5999	6024	238	238	−206	−207
$= -.0041$	4 .6	5999	6019	238	238	−206	−211
$C/A = -.0034$	5 .6	5999	6025	238	238	−206	−203
$C/B = -.087$	6 .6	5999	6020	238	238	−206	−204
Standardized variables							
Ψ	1 .64	5662	5674	157	172	−599	−604
	2 .64	5662	5676	157	172	−599	−604
$(A - \bar{A}')/\bar{A}'$	3 .64	5662	5672	157	172	−599	−605
$= -.0021$	4 .64	5662	5677	157	172	−599	−604
$C/A = -.011$	5 .64	5662	5672	157	172	−599	−601
$C/B = -.38$	6 .64	5662	5669	157	172	−599	−601
λ	1 .6	4718	4761	352	350	−196	−194
	2 .6	4718	4764	352	350	−196	−194
$(A - \bar{A}')/\bar{A}'$	3 .6	4718	4761	352	350	−196	−190
$= -.0089$	4 .6	4718	4763	352	351	−196	−195
$C/A = -.0042$	5 .6	4718	4760	352	350	−196	−198
$C/B = -.056$	6 .6	4718	4757	352	350	−196	−197

Note. SE = standard error, Th. = theoretical (asymptotic) values, Sim. = simulated values.

standard errors using parameter estimates. The simulated SE biases are the means of the SE estimates minus their corresponding true values.

In Table 1, the simulated values are fairly unchanged over the six variables with the same population values, which satisfies a condition when the values are used as true values for this model. The theoretical SE biases are all negative and close to their corresponding simulated values showing the accuracy of the formulas in this article. The theoretical SEs (column A) are slightly but consistently smaller than the simulated or true SEs (column A′). The amount of the relative difference is given as the index $(A - \bar{A}')/\bar{A}'$ in the table where $\bar{A}'$ denotes an average of A′ over the six observed variables. The index shows inaccuracy when neglecting higher-order terms. The corresponding index for SE biases is C/A (theoretical SE bias/theoretical SE). We find that the absolute value of C/A $(-.011)$ in **Ψ** for standardized variables is much larger than the corresponding value of $(A - \bar{A}')/\bar{A}'$ $(-.0021)$, which shows that SE biases should not be neglected when we

Table 2: SE biases of the raw varimax solutions of Emmett's (1949) data under normality ($N = 211$; Number of replications = 1,000,000)

	Parameter value	Unstandardized variables				Standardized variables			
		SE $\times 10^3$ Th.	SE of SE $\times 10^4$ Th.	SE bias $\times 10^5$ Th.	Sim.	SE $\times 10^3$ Th.	SE of SE $\times 10^4$ Th.	SE bias $\times 10^5$ Th.	Sim.
Ψ	1 .46	52	50	−59	−59	53	31	−58	−59
	2 .46	51	49	−56	−57	53	30	−53	−55
	3 .67	70	67	−73	−75	59	15	−69	−69
	4 .19	38	39	−30	−29	41	46	−10	−10
	5 .41	48	45	−37	−37	51	36	−36	−37
	6 .22	42	43	−36	−36	45	49	−19	−18
	7 .40	50	46	−63	−64	53	37	−61	−63
	8 .74	76	74	−76	−77	57	27	−56	−57
	9 .22	41	38	−65	−64	44	46	−47	−47
I	1 .32	59	29	−30	−30	53	23	−30	−32
	2 .38	61	30	−31	−32	53	23	−36	−37
	3 .20	64	29	−45	−46	62	26	−49	−51
	4 .85	59	26	−10	−10	29	33	5	5
	5 .74	62	27	−9	−9	36	37	1	2
	6 .85	60	25	−12	−12	29	33	−1	−1
	7 .27	56	28	−24	−24	51	24	−19	−20
	8 .18	67	31	−49	−50	65	25	−58	−59
	9 .32	52	31	−12	−12	44	22	−3	−4
II	1 .66	64	30	−23	−24	44	40	−17	−18
	2 .63	64	32	−23	−24	45	38	−19	−22
	3 .54	68	32	−35	−37	54	42	−36	−38
	4 .29	45	29	37	37	37	23	22	22
	5 .21	50	27	10	12	46	29	3	4
	6 .23	43	28	37	38	38	24	19	20
	7 .73	63	29	−22	−23	39	40	−15	−16
	8 .47	70	33	−40	−41	59	40	−44	−46
	9 .82	60	28	−17	−18	32	37	−4	−5

Note. SE = standard error, Th. = theoretical (asymptotic) values, Sim. = simulated values, I = loadings of Factor I, II = loadings of Factor II.

consider asymptotic standard errors with higher-order terms. The table includes the index C/B (theoretical SE bias/theoretical SE of SE), which shows the relative size of the SE bias compared to the stability of the SE estimate. While the absolute values of $(A−\bar{A}')/\bar{A}'$ and C/A are mostly less than 1% in the table, the largest absolute value of C/B is as large as 38% in this example.

Tables 2 and 3 show the results of the raw varimax solutions based on real data with simulations by the method similar to that in Table 1. Table 2 is given from the correlation matrix of Emmett's (1949, p. 90) nine variables ($N = 211$; the correlation coefficients rounded to three decimal places by Lawley & Maxwell, 1971, p. 43 were used) while Table 3 is for Harman's (1976, p. 22) eight physical variables ($N = 305$). In both data sets, two common factors are assumed. The population covariance matrices were given by the fitted correlation matrices. The sample sizes in the simulation were set equal to the real ones in each data set. The numbers of Heywood cases until regular 1,000,000 replications in the

Table 3: SE biases of the raw varimax solutions of Harman's (1976) eight physical variables under normality ($N = 305$; Number of replications $= 1,000,000$)

	Parameter value	Unstandardized variables				Standardized variables			
		SE $\times 10^3$	SE of SE $\times 10^4$	SE bias $\times 10^5$		SE $\times 10^3$	SE of SE $\times 10^4$	SE bias $\times 10^5$	
		Th.	Th.	Th.	Sim.	Th.	Th.	Th.	Sim.
Ψ	1 .17	18	14	−14	−14	21	20	−8	−8
	2 .11	15	11	−14	−13	17	18	−6	−6
	3 .17	18	14	−15	−15	21	20	−8	−8
	4 .20	20	16	−16	−16	23	22	−9	−9
	5 .09	29	23	−24	−24	30	30	−5	−5
	6 .36	36	27	−27	−27	39	25	−25	−25
	7 .42	40	30	−30	−30	42	23	−29	−29
	8 .54	46	37	−33	−32	45	14	−32	−32
I	1 .87	44	16	−6	−5	14	13	1	1
	2 .93	43	16	−5	−5	10	10	1	1
	3 .90	44	16	−6	−6	12	12	1	1
	4 .86	45	16	−6	−6	15	14	1	1
	5 .25	36	15	−7	−7	32	11	−4	−4
	6 .21	42	14	−11	−11	39	14	−8	−8
	7 .15	43	14	−12	−12	41	15	−9	−10
	8 .29	48	17	−14	−14	44	17	−13	−14
II	1 .27	35	13	−2	−2	30	11	−3	−3
	2 .16	31	12	1	1	29	9	−3	−3
	3 .16	33	12	−1	−1	31	11	−4	−4
	4 .23	35	13	−3	−3	32	12	−4	−4
	5 .92	45	16	−7	−7	19	18	2	3
	6 .77	49	17	−10	−10	26	22	−3	−4
	7 .75	50	17	−11	−11	28	24	−4	−5
	8 .61	52	19	−13	−13	36	25	−10	−10

Note. SE = standard error, Th. = theoretical (asymptotic) values, Sim. = simulated values, I = loadings of Factor I, II = loadings of Factor II.

simulation were obtained are 28 and 2,146 in Tables 2 and 3, respectively (the same set of generated sample covariance matrices was used both for unstandardized and standardized variables while for standardized variables the covariance matrices were transformed to sample correlation matrices). The theoretical biases in Tables 2 and 3 are close to their corresponding simulated values. The SE biases in **Ψ** are all negative, as was shown in Table 1. The SE biases in standardized variables are, as a whole, somewhat smaller than those in unstandardized variables in these data. The absolute values of the SE biases for **Ψ** in Table 2 (approximately 1% of their corresponding SEs on average) are greater than the corresponding values in Table 3. This can be partially explained by the difference of sample size.

Table 4 shows the results of an artificial example in confirmatory factor analysis with two common factors for six unstandardized observed variables using normally/nonnormally distributed data with $N = 300$. The population covariance matrix was constructed as follows:

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

Table 4: Results of the confirmatory factor analysis model for unstandardized variables ($N = 300$; Number of replications = 1,000,000)

	Thoretical SE	Theoretical SE of SE	Theoretical SE bias	Simulated SE bias			
					Nonnormal (df)		
				Normal	(10)	(3)	(1)
ψ_1	1.1	.09	$-.0032$	$-.0031$	$-.0026$	$-.0032$	$-.0038$
ψ_2	1.9	.17	.0057	.0076	.0065	.0038	$-.0032$
ψ_3	3.6	.47	.0388	.0440	.0425	.0340	.0102
ψ_4	1.4	.11	$-.0036$	$-.0036$	$-.0031$	$-.0040$	$-.0051$
ψ_5	2.3	.20	.0028	.0040	.0027	$-.0016$	$-.0128$
ψ_6	3.1	.29	.0101	.0125	.0103	.0023	$-.0168$
ϕ_{11}	.15	.014	.00067	.00079	.00083	.00073	.00049
ϕ_{21}	.08	.007	$-.00035$	$-.00035$	$-.00038$	$-.00048$	$-.00075$
ϕ_{22}	.14	.012	.00008	.00014	.00008	$-.00006$	$-.00044$
λ_{11}	.27	.026	.0010	.0012	.0017	.0029	.0063
λ_{21}	.44	.060	.0049	.0057	.0065	.0092	.0164
λ_{42}	.32	.031	.0017	.0017	.0022	.0029	.0051
λ_{52}	.47	.059	.0046	.0049	.0054	.0063	.0092

Note. SE = standard error.

$$\boldsymbol{\Lambda} = \left[\begin{array}{cccccc} 2 & 4 & 6^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 3 & 5 & 6^* \end{array} \right]', \ \boldsymbol{\Phi} = \left[\begin{array}{cc} 1 & .4 \\ .4 & 1 \end{array} \right], \ \boldsymbol{\Psi} = \mathrm{diag}(12, 13, 14, 15, 16, 17), \quad (45)$$

where the population values with asterisks were used as fixed parameters in estimation while the remaining values were estimated as free parameters. Note that the unit population variances of the two common factors are free parameters in estimation, which satisfies the condition of a model with asymptotic robustness against the violation of multivariate normality for observed variables. Observations were randomly generated with or without normality for factors. Nonnormal cases were generated by using independently chi-square distributed variables with affine transformation to have zero means and unit variances for the first common factor and the six unique factors with $df = 10$, 3 and 1 while the second common factor was generated as a weighted sum of two independently chi-square distributed variables and a constant to have the required covariance of the two common factors and the unit variance of the second factor. The degrees of freedom 10, 3 and 1 represent moderate, substantial and strong nonnormality, respectively. The numbers of Heywood cases until regular 1,000,000 replications were obtained are 1,000, 1,042, 1,233 and 1,782 for normal and nonnormal data with $df = 10$, 3 and 1, respectively.

Table 4 shows that when data are normally distributed or nonnormality is modest, the theoretical SE biases are fairly close to their corresponding simulated values. However, when nonnormality is strong, the theoretical SE biases become different from their corresponding simulated ones. (It is to be noted that the theoretical SE biases for nonnormal data can be derived, if necessary, by our formulas when the fourth-order central moments of sample covariances are available.) This shows that the SE biases do not have asymptotic robustness which is enjoyed by the asymptotic SEs of the loading estimates in this example (this robustness was empirically confirmed though not shown in the table). It is

of interest to find that the absolute values of the simulated SE biases in nonnormal cases are not necessarily larger than the corresponding theoretical or simulated SE biases with normality assumption. It is also of interest to see that some of the theoretical SE biases in $\boldsymbol{\Psi}$ are positive, which is different from those in Tables 1–3.

7. Conclusive remarks

From the results of the numerical examples with moderate sample sizes, we see that the asymptotic biases of estimated standard errors are fairly small such as 1% of the corresponding asymptotic standard error. However, note that the sample sizes were employed to have stable simulated values as true values and that the relative size is of order $O(n^{-1})$. That is, when the sample size is reduced to a quarter of the original size, the relative size of the asymptotic bias of the estimated standard error becomes 4%, which may not be neglected.

In factor analysis, it is difficult to have stable simulated results with small sample sizes e.g., $N = 50$ and 100 due to anomalous cases such as non-convergence. In exploratory factor analysis, we have an additional problem of indeterminacy of the orders and signs of the columns of a rotated loading matrix. For the numerical examples with raw varimax rotation, the indeterminacy was removed by searching the pattern most similar to the population one. This was successful because the number of common factors was only two both for the two real data examples. When the number of common factors is relatively large, the pattern may depend on the searching methods and/or the definition of similarity. In such cases, the asymptotic standard errors (preferably the higher-order asymptotic standard errors) with bias correction may be only indices available.

8. Discussion

For derivation of the asymptotic biases of estimated standard errors, true models were assumed. In practice, this assumption is not necessarily a realistic one. However, we can relax the assumption by considering the following slight model misspecification (see e.g., Bentler & Dijkstra, 1985; Satorra, 1989, Ogasawara, 2002) without changing the essential results:

$$\boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}(\boldsymbol{\theta}) = O(n^{-1/2}) \tag{46}$$

where $\boldsymbol{\Sigma}_t$ is a true population covariance matrix while $\boldsymbol{\theta}$ has the value obtained by minimizing (8) with possible restrictions and $\mathbf{S}$ being replaced by $\boldsymbol{\Sigma}_t$.

In this article, the normal theory standard error estimators have been focused on while their asymptotic biases were derived with and without the normality assumption for observed variables. More general standard errors are those for nonnormal data which are given as

$$\{\text{ase}(\hat{\theta}_i)\}^2 = n^{-1} \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\sigma}'} \boldsymbol{\Omega} \frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{\sigma}} \right)_{ii}, \quad (i = 1, \dots, q), \tag{47}$$

where $(\boldsymbol{\Omega})_{ab,cd} = \sigma_{abcd} - \sigma_{ab}\sigma_{cd}$. The asymptotic biases as well as asymptotic standard

errors of the estimators of the nonnormal theory standard errors using (47) can, in principle, be obtained when the joint asymptotic distributions of $\hat{\sigma}_{abcd}$'s and $\hat{\sigma}_{ab}$'s are available. In this case, $\mathbf{s}$ and $\boldsymbol{\sigma}$ in (5) should be replaced by those including fourth-order moments, which gives somewhat involved results.

Appendix. Proof of Lemma 2

Suppose that $\hat{\boldsymbol{\theta}}_\rho$ is given by minimizing

$$F_\rho = \ln |\mathbf{P}(\boldsymbol{\theta}_\rho)| + \operatorname{tr}\{\mathbf{P}(\boldsymbol{\theta}_\rho)^{-1}\mathbf{R}\} \tag{A1}$$

with possible restrictions $\mathbf{h}(\boldsymbol{\theta}_\rho) = \mathbf{0}$ as required in exploratory factor analysis. Then, we have the expansion of $n\,\widehat{\operatorname{avar}}(\hat{\theta}_{\rho i})$ with respect to $\mathbf{r} - \boldsymbol{\rho}$ where $\mathbf{r} = \operatorname{vb}(\mathbf{R})$, $\boldsymbol{\rho} = \boldsymbol{\rho}(\boldsymbol{\theta}_\rho) = \operatorname{vb}\{\mathbf{P}(\boldsymbol{\theta}_\rho)\}$ and $\operatorname{vb}(\,\cdot\,)$ is a vector consisting of the elements below the main diagonal of a square matrix:

$$
\begin{aligned}
n\,\widehat{\operatorname{avar}}(\hat{\theta}_{\rho\,i}) &= n\,\operatorname{avar}(\hat{\theta}_{\rho\,i}) + \frac{\partial\, n\,\operatorname{avar}(\hat{\theta}_{\rho\,i})}{\partial\,\boldsymbol{\rho}'} \\
&\quad \times \left\{ \frac{\partial\,\boldsymbol{\rho}}{\partial\,\boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma}) + \frac{1}{2}\begin{bmatrix}(\mathbf{s}-\boldsymbol{\sigma})'\dfrac{\partial^2 \rho_{21}}{\partial\,\boldsymbol{\sigma}\,\partial\,\boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma}) \\ \vdots \\ (\mathbf{s}-\boldsymbol{\sigma})'\dfrac{\partial^2 \rho_{p(p-1)}}{\partial\,\boldsymbol{\sigma}\,\partial\,\boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})\end{bmatrix}\right\} \\
&\quad + \frac{1}{2}(\mathbf{r}-\boldsymbol{\rho})'\frac{\partial^2 n\,\operatorname{avar}(\hat{\theta}_{\rho\,i})}{\partial\,\boldsymbol{\rho}\,\partial\,\boldsymbol{\rho}'}(\mathbf{r}-\boldsymbol{\rho}) + O_p(n^{-3/2}), \\
&\quad (i = 1,\ldots,q),
\end{aligned}
\tag{A2}
$$

where the terms in braces on the right-hand side of (A2) is obtained by the expansion of $\mathbf{r} - \boldsymbol{\rho}$ with respect to $\mathbf{s} - \boldsymbol{\sigma}$. By taking expectations of both sides of (A2), we have

$$
\begin{aligned}
\operatorname{abis}\{n\,\widehat{\operatorname{avar}}(\hat{\theta}_{\rho\,i})\} &= \frac{1}{2}\frac{\partial\, n\,\operatorname{avar}(\hat{\theta}_{\rho\,i})}{\partial\,\boldsymbol{\rho}'}\mathrm{E}\begin{bmatrix}(\mathbf{s}-\boldsymbol{\sigma})'\dfrac{\partial^2 \rho_{21}}{\partial\,\boldsymbol{\sigma}\,\partial\,\boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma}) \\ \vdots \\ (\mathbf{s}-\boldsymbol{\sigma})'\dfrac{\partial^2 \rho_{p(p-1)}}{\partial\,\boldsymbol{\sigma}\,\partial\,\boldsymbol{\sigma}'}(\mathbf{s}-\boldsymbol{\sigma})\end{bmatrix} \\
&\quad + \frac{n^{-1}}{2}\sum_{a>b}\sum_{c>d}2\left(2\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\rho_{ab}}\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\rho_{cd}}\mathbf{H}_A^{-1} - \mathbf{H}_A^{-1}\frac{\partial^2\mathbf{H}_A}{\partial\rho_{ab}\partial\rho_{cd}}\mathbf{H}_A^{-1}\right)_{p+i,p+i} \\
&\quad \times (\boldsymbol{\Pi})_{ab,cd} + O(n^{-2}), \\
&\quad (i = 1,\ldots,q),
\end{aligned}
\tag{A3}
$$

where $\rho_{ab} = \{\mathbf{P}(\boldsymbol{\theta}_\rho)\}_{ab}, (a > b)$ and $(\boldsymbol{\Pi})_{ab,cd} = n\,\operatorname{acov}(r_{ab}, r_{cd}), (a > b;\ c > d)$. In (A3), $(1/2)\,\mathrm{E}[\cdot]$ is the asymptotic bias of the sample correlation vector (for actual expression, see Ogasawara, 2004). Rewrite the term of order $O(n^{-1})$ in $(1/2)\,\mathrm{E}[\cdot]$ as $\operatorname{abis}(\mathbf{r})$. Then,

noting $\partial n \operatorname{avar}(\hat{\theta}_{\rho\,i})/\partial\boldsymbol{\rho}' = -2\sum_{k=1}^{q}\left(\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho k}}\mathbf{H}_A^{-1}\right)_{p+i,p+i}\frac{\partial\theta_{\rho k}}{\partial\boldsymbol{\rho}'}$, (A3) becomes

$$
\begin{aligned}
&= \frac{\partial n\operatorname{avar}(\hat{\theta}_{\rho\,i})}{\partial\boldsymbol{\rho}'}\operatorname{abis}(\mathbf{r}) \\
&\quad + n^{-1}\sum_{k=1}^{q}\left[\sum_{l=1}^{q}\left(2\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho k}}\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho l}}\mathbf{H}_A^{-1} - \mathbf{H}_A^{-1}\frac{\partial^2\mathbf{H}_A}{\partial\theta_{\rho k}\partial\theta_{\rho l}}\mathbf{H}_A^{-1}\right)_{p+i,p+i}\right.\\
&\quad\qquad \times n\operatorname{acov}(\hat{\theta}_{\rho k},\hat{\theta}_{\rho l}) - 2\left(\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho k}}\mathbf{H}_A^{-1}\right)_{p+i,p+i}\frac{1}{2}\operatorname{tr}\left(\frac{\partial^2\theta_{\rho k}}{\partial\boldsymbol{\rho}\,\partial\boldsymbol{\rho}'}\boldsymbol{\Pi}\right)\right]\\
&= \sum_{k=1}^{q}\left[\sum_{l=1}^{q}\left(2\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho k}}\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho l}}\mathbf{H}_A^{-1} - \mathbf{H}_A^{-1}\frac{\partial^2\mathbf{H}_A}{\partial\theta_{\rho k}\partial\theta_{\rho l}}\mathbf{H}_A^{-1}\right)_{p+i,p+i}\right.\\
&\quad \times\operatorname{acov}(\hat{\theta}_{\rho k},\hat{\theta}_{\rho l}) - 2\left(\mathbf{H}_A^{-1}\frac{\partial\mathbf{H}_A}{\partial\theta_{\rho k}}\mathbf{H}_A^{-1}\right)_{p+i,p+i}\left\{\frac{n^{-1}}{2}\operatorname{tr}\left(\frac{\partial^2\theta_{\rho k}}{\partial\boldsymbol{\rho}\,\partial\boldsymbol{\rho}'}\boldsymbol{\Pi}\right) + \frac{\partial\theta_{\rho k}}{\partial\boldsymbol{\rho}'}\operatorname{abis}(\mathbf{r})\right\}\right].
\end{aligned}
$$
(A4)

In (A4), the factor in braces on the right-hand side of the last equation is $\operatorname{abis}(\hat{\theta}_{\rho\,i})$ (see Ogasawara, 2004, Theorem 3), which gives (31). Q.E.D.

Acknowledgements

REFERENCES

Arbuckle, J.L., & Wothke, W. (1999). *Amos 4.0 user's guide.* Chicago: SPSS Inc.

Archer, C.O., & Jennrich, R.I. (1973). Standard errors for rotated factor loadings. *Psychometrika*, **38**, 581–592.

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

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

Browne, M.W., Cudeck, R., Tateneni, K., & Mels, G. (1999). *CEFA: Comprehensive exploratory factor analysis* (Version 1.03). URL: http://quantrm2.psy.ohio-state.edu/browne/

Emmett, W.G. (1949). Factor analysis by Lawley's method of maximum likelihood. *British Journal of Psychology, Statistical Section*, **2**, 90–97.

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

Ichikawa, M., & Konishi, S. (2002). Asymptotic expansions and bootstrap approximations in factor analysis. *Journal of Multivariate Analysis*, **81**, 47–66.

Ihara, M. (1985). Asymptotic bias of estimators of the uniqueness in factor analysis. *Mathematica Japonica*, **30**, 885–889.

Jennrich, R.I. (1974). Simplified formulae for standard errors in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, **27**, 122–131.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL8: User's reference guide* (2nd ed.). Chicago: Scientific Software International.

Lawley, D.N., & Maxwell, A.E. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworths.

Ogasawara, H. (2000). On the standard errors of rotated factor loadings with weights for observed variables. *Behaviormetrika*, **27**, 1–14.

Ogasawara, H. (2002). Asymptotic standard errors of estimated standard errors in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, **55**, 213–229.

Ogasawara, H. (2003). *ROSEF Version 2.0 user's guide: A subroutine library for the rotated solutions with their asymptotic standard errors in factor analysis*. URL: http://www.res.otaru-uc.ac.jp/~hogasa/

Ogasawara, H. (2004). Asymptotic biases in exploratory factor analysis and structural equation modeling. *Psychometrika*, **69**, 235–256.

SAS Institute Inc. (2000). *SAS/STAT software: Changes and enhancements*, Release 8.1. Cary, NC: Author.

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

Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models. *Econometric Theory*, **18**, 297–312.

Yuan, K.-H., & Bentler, P. (2000). On equivalence and invariance of standard errors in three exploratory factor models. *Psychometrika*, **65**, 121–133.