Economic Review (Otaru University of Commerce),
Vo.57, No.4, 131-160, March, 2007

Higher-order estimation error in structural equation modeling*

Haruhiko Ogasawara

A general formula of the higher-order asymptotic standard error is derived for the estimators of the parameters in structural equation modeling. The formula can be used for nonnormally distributed data as well as normally distributed ones. For this derivation, the third- and fourth-order asymptotic central moments of sample variances and covariances are provided for nonnormally and normally distributed cases. The formula requires the partial derivatives of an estimator up to the third order with respect to sample variances and covariances, which are shown for the case of the Wishart maximum likelihood estimator. To see the accuracy of the formula, simulations are performed using the factor/component analysis models. It is numerically shown that some of the added contributions of the higher-order asymptotic standard errors are substantial with small to modest sample sizes.

Key words: Mean square errors, asymptotic standard errors, asymptotic biases, higher-order asymptotic expansion, structural equation modeling, nonnormal distributions, asymptotic robustness.

1. Introduction

In structural equation modeling, errors of parameter estimates due to sampling variation and use of some estimation methods are evaluated typically by the square roots of the associated mean square errors. They are usually given as asymptotic standard errors and are available in the familiar programs e.g., Amos (Arbucle & Wothke, 1999), EQS (Bentler, 1989), LISREL (Jöreskog & Sörbom, 1996), Mplus (Muthén & Muthén, 2004) and RAMONA (Browne & Mels, 2000) with or without the assumption of multivariate

---

* Essential results of this article including the lemmas, theorems and corollary were first given in a paper published as Ogasawara (2004, October). Partially supported by Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology. Author's address: Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 Japan. Email: hogasa@res.otaru-uc.ac.jp

normality for observed variables. Under normality, the asymptotic standard errors are given from the inverse of the information matrix while under nonnormality they are given by the sandwich-like estimators for the asymptotic variances shown later. The asymptotic standard errors are of order $O(n^{-1/2})$, where n+1 is the number of observations, and are not necessarily equivalent to the corresponding exact root mean square errors. The difference stems from the existence of the biases of parameter estimates and from the remaining higher-order terms in an asymptotic expansion. Fortunately, the biases are usually of order $O(n^{-1})$ (see e.g., Siotani, Hayakawa & Fujikoshi, 1985, p.161), which is asymptotically smaller than that of the usual asymptotic standard errors. Therefore, the latter standard errors are asymptotically equivalent to the corresponding root mean square errors.

While the asymptotic standard errors usually give values accurate in practical sense with moderate sample sizes under some distributional conditions required, we often observe slightly but consistently different results between simulated (or true) standard errors and the asymptotic ones. This will be illustrated later in the sections of numerical examples. A method of improving the asymptotic standard errors has been known at least in principle as higher-order asymptotic expansion of parameter estimates. This method has been developed mostly from theoretical viewpoint with sparse applications in practice (for historical reviews and summary, see Rothenberg, 1984; Ghosh, 1994; and Kano, 1997, 1998). In structural equation modeling the topic was briefly discussed by Bentler (1983, p.502) in an early stage.

The purpose of this article is to derive a general formula of the higher-order (i.e., accurate up to order $O(n^{-2})$) asymptotic mean square error of an estimator with some discrepancy function and apply this to the case of the usual Wishart maximum likelihood estimators (MLEs) of structural parameters with and without the assumption of multivariate normality for observed variables. Similar results including numerical examples were given by Amemiya (1980) for the MLEs and the minimum chi-square estimators in the special case of logistic regression. Numerical examples of factor/component analysis models with simulations will be given to show the accuracy of the formula. It will be illustrated that the higher-order asymptotic standard errors well explain the differences of the usual asymptotic standard errors and the corresponding simulated ones.

2. The estimators of parameters and their higher-order asymptotic variances

Let $\boldsymbol{\theta}$ be the $q \times 1$ parameter vector in a $p \times p$ covariance matrix

$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ of p observed variables. The vector $\hat{\boldsymbol{\theta}}$ of parameter estimates is assumed to be obtained by minimizing a discrepancy function of the estimators with regularity conditions for such function (see e.g., Browne, 1982, p.81):

$$F = F(\mathbf{S}, \boldsymbol{\Sigma}) \text{ subject to } \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}, \qquad (1)$$

where $\mathbf{S}$ is a $p \times p$ unbiased sample covariance matrix and $\mathbf{h}(\boldsymbol{\theta})$ is a $r \times 1$ vector for restrictions on $\boldsymbol{\theta}$. Note that F is a generic one including those by maximum likelihood and weighted/unweighted least squares. Let $\hat{\theta}_i, (i = 1, ..., q)$ be the i-th element of $\hat{\boldsymbol{\theta}}$ and suppose that minimizing F gives $\hat{\theta}_i$ which is assumed to be three times differentiable with respect to sample variances and covariances, and can be expanded in the following Taylor series:

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

$$+ \frac{1}{6}\left(\frac{\partial}{\partial \mathbf{s}'}\right)^{<3>} \hat{\theta}_i \Big|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s} - \boldsymbol{\sigma})^{<3>} + O_p(n^{-2}), \qquad (2)$$

$$(i = 1, ..., q),$$

where $\boldsymbol{\sigma} = \mathrm{v}(\boldsymbol{\Sigma})$; $\mathbf{s} = \mathrm{v}(\mathbf{S})$; $\mathrm{v}(\cdot)$ is the vectorizing operator taking the nonduplicated elements of a symmetric matrix; and $\mathbf{s}^{<u>}$ denotes the u-fold (right) Kronecker product (see Kano, 1997) i.e., $\mathbf{s}^{<u>} = \mathbf{s} \otimes \mathbf{s} \otimes \cdots \otimes \mathbf{s}$ (u times). Note that the elementwise expression of e.g., the third-order term on the right-hand side of (2) is

$$\frac{1}{6}\sum_{a \ge b}\sum_{c \ge d}\sum_{e \ge f} \frac{\partial^3 \hat{\theta}_i}{\partial s_{ab} \partial s_{cd} \partial s_{ef}}\Big|_{\mathbf{s}=\boldsymbol{\sigma}} (s_{ab} - \sigma_{ab})(s_{cd} - \sigma_{cd})(s_{ef} - \sigma_{ef}), \qquad (3)$$

where $\sum_{a \ge b} = \sum_{p \ge a \ge b \ge 1}$.

In (2), θ_i and $\boldsymbol{\sigma}$ are used as true or population values while they were variables in (1), and $\hat{\theta}_i$ and $\mathbf{s}$ as variables in (2), which were previously estimates, to avoid complicated expression. We assume that the required finite higher-order moments of $\mathbf{s}$ exist. Then, the mean square error of $\hat{\theta}_i$ is given from (2) as

$$\mathrm{E}\{(\hat{\theta}_i-\theta_i)^2\}=\left(\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\mathrm{E}\left\{(\mathbf{s}-\boldsymbol{\sigma})^{<2>}\right\}$$

$$+\mathrm{E}\left[\left\{\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\otimes\left(\left(\frac{\partial}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\theta_i\right)\right\}(\mathbf{s}-\boldsymbol{\sigma})^{<3>}+\frac{1}{4}\left\{\left(\left(\frac{\partial}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\theta_i\right)(\mathbf{s}-\boldsymbol{\sigma})^{<2>}\right\}^2\right. \tag{4}$$

$$\left.+\frac{1}{3}\left\{\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\otimes\left(\left(\frac{\partial}{\partial\boldsymbol{\sigma}'}\right)^{<3>}\theta_i\right)\right\}(\mathbf{s}-\boldsymbol{\sigma})^{<4>}\right]+O(n^{-3}),\ (i=1,\ldots,q),$$

where $\partial\theta_i/\partial\boldsymbol{\sigma}=\partial\hat{\theta}_i/\partial\mathbf{s}\,|_{\mathbf{s}=\boldsymbol{\sigma}}$ with other similar expressions is used for simplicity of notation;

$$\left(\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\mathrm{E}\left\{(\mathbf{s}-\boldsymbol{\sigma})^{<2>}\right\}=n^{-1}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}}-n^{-2}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\mathbf{K}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}}+O(n^{-3}) \tag{5}$$

(see Kaplan, 1952, Equation (3)) with

$$\mathrm{avar}(\hat{\theta}_i;n^{-1})\equiv n^{-1}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}}=n^{-1}\left(\frac{\partial\theta_i}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\mathrm{vec}\,\boldsymbol{\Omega},\quad (i=1,\ldots,q);$$

$\mathrm{avar}(\hat{\theta}_i;n^{-1})$ is the usual asymptotic variance of $\hat{\theta}_i$ up to order $O(n^{-1})$; $\mathrm{vec}(\cdot)$ is the vectorizing operator stacking the columns of a matrix; and $\boldsymbol{\Omega}=(\omega_{ab,cd})=(\sigma_{abcd}-\sigma_{ab}\sigma_{cd})=n\,\mathrm{acov}(\mathbf{s})$, using double subscript notation for elements, is the $p^*\times p^*$ asymptotic covariance matrix of $\sqrt{n}\,\mathbf{s}$ with $p^*=p(p+1)/2$; σ_{abcd} is the fourth multivariate central moment of the variables X_a, X_b, X_c and X_d; and $\mathbf{K}$ is the $p^*\times p^*$ matrix of the fourth multivariate cumulants of X_j's. Let the sum of the terms of order $O(n^{-2})$ in (4) be denoted by $\Delta\mathrm{avar}(\hat{\theta}_i;n^{-2})+\{\mathrm{abis}(\hat{\theta}_i;n^{-1})\}^2$, where $\Delta\mathrm{avar}(\hat{\theta}_i;n^{-2})$ is the added asymptotic variance of $\hat{\theta}_i$ with correction of the asymptotic bias of $\hat{\theta}_i$ up to order $O(n^{-1})$ given by

$$\mathrm{abis}(\hat{\theta}_i;n^{-1})=\frac{n^{-1}}{2}\left\{\left(\frac{\partial}{\partial\boldsymbol{\sigma}'}\right)^{<2>}\theta_i\right\}\mathrm{vec}\,\boldsymbol{\Omega},\quad (i=1,\ldots,q). \tag{6}$$

Similarly, let the asymptotic variance of $\hat{\theta}_i$ including the terms of order

$O(n^{-1})$ and $O(n^{-2})$ be denoted by $\text{avar}(\hat{\theta}_i; n^{-1}, n^{-2})$. Then,

$$\begin{aligned}\text{E}\{(\hat{\theta}_i - \theta_i)^2\} &= \text{avar}(\hat{\theta}_i; n^{-1}, n^{-2}) + \{\text{abis}(\hat{\theta}_i; n^{-1})\}^2 + O(n^{-3}) \\ &= \text{avar}(\hat{\theta}_i; n^{-1}) + \Delta\text{avar}(\hat{\theta}_i; n^{-2}) + \{\text{abis}(\hat{\theta}_i; n^{-1})\}^2 + O(n^{-3}).\end{aligned} \quad (7)$$

The higher-order asymptotic standard error with bias correction and the higher-order asymptotic root mean square error of $\hat{\theta}_i$ are defined as

$$\begin{aligned}&\text{HASE}(\hat{\theta}_i) = \sqrt{\text{avar}(\hat{\theta}_i; n^{-1}, n^{-2})}, \\ &\text{HARME}(\hat{\theta}_i) = \sqrt{\text{avar}(\hat{\theta}_i; n^{-1}, n^{-2}) + \{\text{abis}(\hat{\theta}_i; n^{-1})\}^2}, \\ &(i = 1, \ldots, q).\end{aligned} \quad (8)$$

The problem is to have the expectations of $(\mathbf{s} - \boldsymbol{\sigma})^{<u>}, (u = 3, 4)$ in (4) and the partial derivatives of $\hat{\theta}_i$ with respect to **s** up to the third order. The former results are given by Lemmas 1 to 3 in the appendix and the latter in the next section. Before expanding the partial derivatives, we have the following general results.

Theorem 1. The bias-corrected higher-order asymptotic variance of $\hat{\theta}_i$ with the assumption of the existence of the associated moments of observed variables and three times differentiability of $\hat{\theta}_i$ with respect to sample variances and covariances is given as

$$\begin{aligned}\Delta\text{avar}(\hat{\theta}_i; n^{-2}) = n^{-2} \Bigg\{ &-\frac{\partial \theta_i}{\partial \boldsymbol{\sigma}'} \mathbf{K} \frac{\partial \theta_i}{\partial \boldsymbol{\sigma}} + \sum_{a \ge b} \sum_{c \ge d} \sum_{e \ge f} \frac{\partial \theta_i}{\partial \sigma_{ab}} \frac{\partial^2 \theta_i}{\partial \sigma_{cd} \partial \sigma_{ef}} \\ &\times (\sigma_{abcdef} - \sigma_{ab}\sigma_{cdef} - \sigma_{cd}\sigma_{abef} - \sigma_{ef}\sigma_{abcd} \\ &- \sigma_{acd}\sigma_{bef} - \sigma_{bcd}\sigma_{aef} - \sigma_{abc}\sigma_{def} - \sigma_{abd}\sigma_{cef} \\ &- \sigma_{abe}\sigma_{cdf} - \sigma_{abf}\sigma_{cde} + 2\sigma_{ab}\sigma_{cd}\sigma_{ef}) \\ &+ \sum_{a \ge b} \sum_{c \ge d} \sum_{e \ge f} \sum_{g \ge h} \left(\frac{1}{2} \frac{\partial^2 \theta_i}{\partial \sigma_{ab} \partial \sigma_{ef}} \frac{\partial^2 \theta_i}{\partial \sigma_{cd} \partial \sigma_{gh}} + \frac{\partial \theta_i}{\partial \sigma_{gh}} \frac{\partial^3 \theta_i}{\partial \sigma_{ab} \partial \sigma_{cd} \partial \sigma_{ef}} \right) \\ &\times (\sigma_{abcd} - \sigma_{ab}\sigma_{cd})(\sigma_{efgh} - \sigma_{ef}\sigma_{gh}) \Bigg\}, \\ &(i = 1, \ldots, q),\end{aligned} \quad (9)$$

where σ_{abcdef} and σ_{abc} are the multivariate sixth-, and third-order central moments of the variables corresponding to the subscripts, respectively.

Proof. Using Lemmas 1 and 3, considering the symmetric properties of the three terms in parentheses on the right-hand side of (A19) and removing the term for the squared bias in (4), the result follows. Q.E.D.

An equivalent expression of the third term in braces on the right-hand side of (9) is $n^2 \mathrm{tr}\left\{\frac{1}{2}\mathrm{acov}\left(\frac{\partial\hat{\theta}_i}{\partial\mathbf{s}}\right)\mathrm{acov}(\mathbf{s}) + \mathrm{acov}\left(\hat{\theta}_i, \frac{\partial^2\hat{\theta}_i}{\partial\mathbf{s}\partial\mathbf{s}'}\right)\mathrm{acov}(\mathbf{s})\right\}$. The asymptotic bias is similarly written as

$$\mathrm{abis}(\hat{\theta}_i; n^{-1}) = \frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2\theta_i}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\mathrm{acov}(\mathbf{s})\right\}, \quad (i = 1, \ldots, q). \tag{10}$$

When the multivariate normality holds, we have the following result using elementwise expression.

Corollary 1. The bias-corrected higher-order asymptotic variance of $\hat{\theta}_i$ with the assumption of multivariate normality in addition to that in Theorem 1 is

$$\begin{aligned}\Delta\mathrm{avar}(\hat{\theta}_i; n^{-2}) = n^{-2}\Bigg[&\sum_{a\ge b}\sum_{c\ge d}\sum_{e\ge f}\frac{\partial\theta_i}{\partial\sigma_{ab}}\frac{\partial^2\theta_i}{\partial\sigma_{cd}\partial\sigma_{ef}} \\ &\times\{(\sigma_{fa}\sigma_{bc} + \sigma_{fb}\sigma_{ac})\sigma_{de} + (\sigma_{fa}\sigma_{bd} + \sigma_{fb}\sigma_{ad})\sigma_{ce} \\ &+ (\sigma_{ea}\sigma_{bc} + \sigma_{eb}\sigma_{ac})\sigma_{df} + (\sigma_{ea}\sigma_{bd} + \sigma_{eb}\sigma_{ad})\sigma_{cf}\} \\ &+\sum_{a\ge b}\sum_{c\ge d}\sum_{e\ge f}\sum_{g\ge h}\left(\frac{1}{2}\frac{\partial^2\theta_i}{\partial\sigma_{ab}\partial\sigma_{ef}}\frac{\partial^2\theta_i}{\partial\sigma_{cd}\partial\sigma_{gh}} + \frac{\partial\theta_i}{\partial\sigma_{gh}}\frac{\partial^3\theta_i}{\partial\sigma_{ab}\partial\sigma_{cd}\partial\sigma_{ef}}\right) \\ &\times(\sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc})(\sigma_{eg}\sigma_{fh} + \sigma_{eh}\sigma_{fg})\Bigg],\end{aligned} \tag{11}$$

$(i = 1, \ldots, q)$.

Proof. Using Lemmas 2 and 3 with $n\,\mathrm{acov}(s_{ab}, s_{cd}) = \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}$ in the appendix and the result of Theorem 1, (11) follows. Q.E.D.

3. The third partial derivatives of an estimator with respect to sample variances and covariances

In this section, we derive the third partial derivatives of an estimator $\hat{\theta}_i$

with respect to sample variances and covariances with the assumption that the functions for parameter constraints are four times differentiable with respect to parameters. The first partial derivative is required for the usual asymptotic variances (see (5)). The second partial derivative is required for the asymptotic biases (see (6) and (10)), which was given by Shapiro (1983, Theorem 4.3) in a general expression without specifying discrepancy functions for estimators. The actual expressions of the second partial derivatives for the MLEs in structural equation modeling were derived by Ogasawara (2004a) and the results for the generalized, scale-free and unweighted least squares estimators by Ogasawara (2004b). The general result of the third partial derivatives shown below is an extension of these results.

Suppose that the discrepancy function F in (1) is minimized with an r-dimensional vector of restrictions set equal to 0 as $\mathbf{h} = \mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$. Let $G = F + \boldsymbol{\xi}'\mathbf{h}$ with $\boldsymbol{\eta} = (\boldsymbol{\theta}', \boldsymbol{\xi}')'$, where $\boldsymbol{\xi}$ is a $r \times 1$ Lagrange multiplier vector. Then, the first-order condition of $\hat{\boldsymbol{\theta}}$ minimizing F is given as

$$\frac{\partial \hat{G}}{\partial \hat{\boldsymbol{\eta}}} = \left(\frac{\partial \hat{F}}{\partial \hat{\boldsymbol{\theta}}'} + \hat{\boldsymbol{\xi}}' \frac{\partial \hat{\mathbf{h}}}{\partial \hat{\boldsymbol{\theta}}'}, \ \hat{\mathbf{h}}' \right)' = \mathbf{0}, \tag{12}$$

where

$$\hat{G} = \hat{F} + \hat{\boldsymbol{\xi}}'\hat{\mathbf{h}} = F(\mathbf{S}, \hat{\boldsymbol{\Sigma}}) + \hat{\boldsymbol{\xi}}'\hat{\mathbf{h}} = F(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) + \hat{\boldsymbol{\xi}}'\mathbf{h}(\hat{\boldsymbol{\theta}}). \tag{13}$$

Equation (12) represents an (implicit) function for $\hat{\boldsymbol{\theta}}$ in terms of sample variances and covariances.

Differentiating (12) with respect to s_{ab}, we have

$$\frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \frac{\partial \hat{\boldsymbol{\eta}}}{\partial s_{ab}} + \frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial s_{ab}} = \mathbf{0}, \tag{14}$$

which gives the first partial derivative of $\hat{\boldsymbol{\eta}}$ with respect to s_{ab},

$$\frac{\partial \hat{\boldsymbol{\eta}}}{\partial s_{ab}} = -\left(\frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \right)^{-1} \frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial s_{ab}} \equiv -\left(\frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \right)^{-1} \hat{\mathbf{a}}_{1,ab}, \ \ (p \ge a \ge b \ge 1). \tag{15}$$

Again, differentiating (14) with respect to s_{cd},

$$\sum_i \sum_j \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial \hat{\eta}_j} \frac{\partial \hat{\eta}_i}{\partial s_{ab}} \frac{\partial \hat{\eta}_j}{\partial s_{cd}} + \sum_i \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial s_{cd}} \frac{\partial \hat{\eta}_i}{\partial s_{ab}}$$

$$+ \sum_i \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial s_{ab}} \frac{\partial \hat{\eta}_i}{\partial s_{cd}} + \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial s_{ab} \partial s_{cd}} = - \frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \frac{\partial^2 \hat{\boldsymbol{\eta}}}{\partial s_{ab} \partial s_{cd}}, \quad (16)$$

$$(p \geq a \geq b \geq 1; p \geq c \geq d \geq 1).$$

Let $\hat{\mathbf{a}}_{2,ab,cd}$, similarly to $\hat{\mathbf{a}}_{1,ab}$ in (15), denote the $(q+r) \times 1$ vector corresponding to the left-hand side of (16). Then, we have the second partial derivative of $\hat{\boldsymbol{\eta}}$ with respect to s_{ab} and s_{cd}:

$$\frac{\partial^2 \hat{\boldsymbol{\eta}}}{\partial s_{ab} \partial s_{cd}} = - \left(\frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \right)^{-1} \hat{\mathbf{a}}_{2,ab,cd}, \ (p \geq a \geq b \geq 1; p \geq c \geq d \geq 1). \quad (17)$$

Equation (17) is equivalent to the result first given by Shapiro (1983).

Differentiating (16) further with respect to s_{ef},

$$\sum_i \sum_j \sum_k \frac{\partial^4 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial \hat{\eta}_j \partial \hat{\eta}_k} \frac{\partial \hat{\eta}_i}{\partial s_{ab}} \frac{\partial \hat{\eta}_j}{\partial s_{cd}} \frac{\partial \hat{\eta}_k}{\partial s_{ef}}$$

$$+ \sum_{(U,V,W)}^{{}_3C_1} \sum_i \sum_j \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial \hat{\eta}_j} \frac{\partial \hat{\eta}_i}{\partial s_U} \frac{\partial^2 \hat{\eta}_j}{\partial s_V \partial s_W}$$

$$+ \sum_{(U,V,W)}^{{}_3C_1} \sum_i \sum_j \frac{\partial^4 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial \hat{\eta}_j \partial s_U} \frac{\partial \hat{\eta}_i}{\partial s_V} \frac{\partial \hat{\eta}_j}{\partial s_W}$$

$$+ \sum_{(U,V,W)}^{{}_3C_1} \sum_i \frac{\partial^3 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial s_U} \frac{\partial^2 \hat{\eta}_i}{\partial s_V \partial s_W}$$

$$+ \sum_{(U,V,W)}^{{}_3C_1} \sum_i \frac{\partial^4 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\eta}_i \partial s_U \partial s_V} \frac{\partial \hat{\eta}_i}{\partial s_W} + \frac{\partial^4 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial s_{ab} \partial s_{cd} \partial s_{ef}} \quad (18)$$

$$= - \frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \frac{\partial^3 \hat{\boldsymbol{\eta}}}{\partial s_{ab} \partial s_{cd} \partial s_{ef}},$$

$$(p \geq a \geq b \geq 1; p \geq c \geq d \geq 1; p \geq e \geq f \geq 1).$$

where $\sum_{(U,V,W)}^{{}_3C_1}$ denotes a summation over the range:

$$(U,V,W) \in \{(ab,cd,ef),(ef,ab,cd),(cd,ef,ab)\}.$$

The range for $\sum_i, \sum_j$ and $\sum_k$ in (16) and (18) should be that for the associated parameters with functional relationships though usually $\sum_{i=1}^{q}, \sum_{j=1}^{q}$ and $\sum_{k=1}^{q}$ suffice as in factor analysis models. From (18), we have

Theorem 2. The third partial derivatives of the estimators $\hat{\boldsymbol{\theta}}$, and estimated Lagrange multipliers $\hat{\boldsymbol{\xi}}$ for restrictions $\mathbf{h}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ with the discrepancy function F under the assumption of the existence of the partial derivatives of $\hat{G}$ with respect to $\hat{\boldsymbol{\theta}}$ and sample variances and covariances up to the required number of times is given as

$$\frac{\partial^3 \hat{\boldsymbol{\eta}}}{\partial s_{ab} \partial s_{cd} \partial s_{ef}} = -\left(\frac{\partial^2 \hat{G}}{\partial \hat{\boldsymbol{\eta}} \partial \hat{\boldsymbol{\eta}}'} \right)^{-1} \hat{\mathbf{a}}_{3,ab,cd,ef}, \quad (p \geq a \geq b \geq 1; p \geq c \geq d \geq 1; p \geq e \geq f \geq 1). \tag{19}$$

where $\hat{\mathbf{a}}_{3,ab,cd,ef}$ is the left-hand side of (18).

It is obvious that Theorem 2 can be simplified when no restrictions are imposed on $\boldsymbol{\theta}$. In this case, G and $\boldsymbol{\eta}$ can be replaced by F and $\boldsymbol{\theta}$, respectively. When the restrictions, if any, are only for model identification, the Lagrange multipliers can be omitted because in this case $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}} = \mathbf{0}$, which yields

$$\left(\frac{\partial \hat{F}}{\partial \hat{\boldsymbol{\theta}}'}, \hat{\mathbf{h}}' \right)' = \mathbf{0} \tag{20}$$

in place of (12), and $\mathbf{0}$ for $\partial \hat{\boldsymbol{\xi}} / \partial s_{ab}, (p \geq a \geq b \geq 1)$. Though the result of Theorem 2 is a general one, in actual computation we require different explications for various discrepancy functions of estimators even in a stage without specifying structural models. The results in the case of the MLE are shown in the appendix.

4. A numerical example under normality and nonnormality

The purpose of this section is to illustrate the accuracy of the proposed formulas and the relative size of HASE to ASE using a typical structural model with moderate sample size under normality and nonnormality. The model employed is an oblique confirmatory factor analysis model for six unstandardized observed variables with the following population values,

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi},$$

$$\mathbf{\Lambda} = \begin{bmatrix} 3 & 5 & 7^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 4 & 6 & 8^* \end{bmatrix}', \mathbf{\Phi} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}, \tag{21}$$

$$\mathbf{\Psi} = \text{diag}(18, 19, 20, 21, 22, 23),$$

where the loadings with asterisks are fixed parameters for model identification and the nonduplicated elements of $\mathbf{\Phi}$ are free ones. It is known that the normal theory asymptotic standard errors up to order $O(n^{-1/2})$ of the loading estimators in the model have asymptotic robustness when the common factors and unique factors are independently nonnormally distributed. The remaining parameters do not enjoy such property (see e.g., Browne & Shapiro, 1988, Proposition 3.1). Simulations were performed under normality and nonnormality. Nonnormal data were generated using independently chi-square distributed variables with a common value of degrees of freedom followed by affine transformation to have zero means and unit variances for the first common factor and the unique factors. The second common factor was generated by the weighted sum of the first factor, an independently chi-square distributed variable and a constant to have the population pattern of $\mathbf{\Phi}$. The strength of nonnormality was controlled by the degrees of freedom for the chi-square distributions. The values *df*=3 and 1 were used to represent substantial and strong nonnormality, respectively.

In the simulation, the moderate sample size *N*=300 was used with the number of replications 1,000,000. No non-convergent cases occurred. The numbers of the Heywood cases included are 21, 50 and 152 for the normal and nonnormal data with *df*=3 and 1, respectively. The simulated or empirical standard errors (denoted by SDs) corresponding to $\text{HASE}(\hat{\theta}_i)$'s (see(8)) were given from the standard deviations or the square roots of the usual unbiased sample variances based on 1,000,000 estimates for each parameter. Note that the SDs are empirically bias corrected.

Table 1 shows the results. The ASEs are the population asymptotic standard errors of order $O(n^{-1/2})$. The HASEs denote bias-corrected population $\mathrm{HASE}(\hat{\theta}_i)$'s. The values of SD/ASE are the ratios of the true standard errors to the corresponding asymptotic ones. The corresponding theoretical values are given by HASE/ASE. In the table, it is seen that the ASEs and HASEs for factor variance-covariance estimators become larger when nonnormality becomes stronger. On the other hand, the ASEs for loading estimators are unchanged irrespective of nonnormality, which stems from their asymptotic robustness. Unfortunately, the corresponding HASEs of the loading estimators increase mildly but systematically with the strength of nonnormality, which shows that the HASEs are not robust against the violation of normality.

We find that the ratios HASE/ASE are reasonably similar to their corresponding simulated ones SD/ASE and that these ratios in nonnormal samples are different from those in normal samples. Especially for the loading estimators, the ratios in nonnormal samples are relatively larger than those in other parameters due to the unchanged robust ASEs. The largest value of the fractional part of AHSE/ASE is 6% for λ_{21} with the corresponding simulated value being 7%. Similar results were obtained using similar models e.g., the one-factor model and the two-factor exploratory model with factor rotation though they are not shown here. From the limited result of the table, the accuracy of the formulas of HASE is illustrated, while the relative size of HASE to ASE may not be substantial except for some parameters, which is due to the moderate sample size to have stable simulated data. In the next section, we deal with a case with small sample sizes.

5. An illustration with small sample sizes

An advantage of the higher-order asymptotic standard errors over the usual asymptotic ones are seen more clearly in data sets with small sample sizes than those with large ones (note that $\mathrm{HASE/ASE}=1+O(n^{-1})$). In this section, we illustrate such cases. The theoretical values with small sample sizes are easily calculated by changing the values of *n*. On the other hand, we often have difficulty in obtaining the corresponding true or simulated values for usual parameter estimators in small samples due to anomalous cases such as non-convergence. To avoid this instability, we use the following saturated model and population parameter values as an illustration

$$\mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}' \tag{22}$$

with $\mathbf{\Lambda} = \begin{bmatrix} 1^* & 0^* & 0^* \\ 0.2 & 1^* & 0^* \\ 0.1 & 0.2 & 1^* \end{bmatrix}$ and $\mathbf{\Phi} = \mathrm{diag}(\phi_1, \phi_2, \phi_3) = \begin{bmatrix} 1 & 0^* & 0^* \\ 0^* & 1 & 0^* \\ 0^* & 0^* & 1 \end{bmatrix}$,

where the values with asterisks are fixed parameters for model identification and $\lambda_{21}, \lambda_{31}, \lambda_{32}, \phi_1, \phi_2$ and ϕ_3 are free ones to be estimated. The model is a just identified component analysis model including *p* full components with unconstrained component variances. Note that the normal theory asymptotic standard errors of the free loading estimators have the asymptotic robustness under nonnormality when the *p* full components are independently distributed. The estimates of the parameters are given, without iterative computation, by the Cholesky-decomposition of a sample covariance matrix including scaling to satisfy (22).

Simulations were performed using $n = N - 1 = 25$ under normality and nonnormality with the same number of replications as before. Nonnormal samples were given by independently uniform, *t* (*df*=9) and chi-square (*df*=10, 3 and 1) distributed components including standardization for the components. It is known that the skewnesses or standardized third cumulants of the uniform, *t* (*df*=9) and chi-square (*df*=10) distributions are 0, 0 and $2/\sqrt{5}$ while the corresponding kurtoses or standardized fourth cumulants are $-6/5, 6/5$ and 6/5, respectively (see e.g., Stuart & Ort, 1994, Sections 16.3 and 16.11). These values give convenient comparison of the results.

Table 2 shows the results. No observation was discarded until 1,000,000 samples were obtained in each simulation. The theoretical ratios HASE/ASE are similar to their corresponding simulated ones. The theoretical values of HASE/ASE with samples sizes other than *n*=25 can be easily obtained by noting that $(\mathrm{HASE/ASE})^2 - 1$ is inversely proportional to *n*. For instance when *n* is doubled, the ratio HASE/ASE for $\hat{\phi}_1$ under the uniform distribution becomes $\sqrt{1 + \{(1.0296^2 - 1) \times 25/50\}} = 1.0149$. The simulated ratios with increased sample sizes were found to be close to their corresponding theoretical ratios though not shown in the table.

The ASEs of the loading estimators are unchanged irrespective of nonnormality, which is due to the asymptotic robustness of the estimators. On the other hand, the ASEs of the component variance estimators under the uniform distribution are smaller than those under normality while those under the remaining nonnormal distributions are larger, which is expected from the negative or positive kurtoses of the nonnormal distributions. The same values of the ASEs of the component variance estimators under the *t* (*df*=9) and chi-square (*df*=10) distributions stem from the common kurtosis of 6/5. The HASE/ASE's for these two distributions are computationally the same. The author conjectures that they are algebraically the same. This finding suggests that kurtosis has greater influence on HASEs, as is known on ASEs, than skewness. The largest absolute value of $(\mathrm{HASE}/\mathrm{ASE})-1$ is as large as 26% for λ_{32} in the chi-squared case with *df*=1, whose corresponding simulated value is 29%. These results are encouraging in that the higher-order asymptotic standard errors can have substantial improvement of approximation to their corresponding true standard errors in relatively small samples.

The HASEs and ASEs in the numerical examples are given by using population parameters. In practice, however, only their sample counterparts are available. Considering this situation, a small simulation study under normality using the same model as in Table 2 was carried out with sample sizes *n*=25 and 50. The values of $\widehat{\mathrm{HASE}}$ s and $\widehat{\mathrm{ASE}}$ s were given by using parameter estimates in each replication. The number of replications was reduced to 100,000 due to excessive computing time required. Table 3 shows the results, where *m* and *sd* stand for the means and standard deviations of the 100,000 estimates of HASE and ASE, respectively.

From the table, we easily see that the *sd*'s of $\widehat{\mathrm{ASE}}$ s are larger than the differences of the *m*'s of $\widehat{\mathrm{ASE}}$ and $\widehat{\mathrm{HASE}}$. This is expected since the asymptotic standard errors of $\widehat{\mathrm{ASE}}$ s are of order $O(n^{-1})$, which is asymptotically larger than $O(n^{-3/2})$ for $\mathrm{HASE}-\mathrm{ASE}$. That is, by asymptotic theory the added information in $\widehat{\mathrm{HASE}}$ cannot be separated from the sampling error of $\widehat{\mathrm{ASE}}$.

Though this is mathematically or asymptotically true, we still have an advantage for $\widehat{\mathrm{HASE}}$ s. The last column of Table 3 shows the ratios of *m* of HASE to *m* of $\widehat{\mathrm{ASE}}$, which happen to be the same as the corresponding population values (see Table 2 when *n*=25). Actually, the ratios $\widehat{\mathrm{HASE}}/\widehat{\mathrm{ASE}}$ for all parameters are (computationally) the same as the corresponding population values from replication to replication in this case. The equivalence of the sample and population ratios can also be algebraically shown for component variance estimators. That is,

$$\frac{\mathrm{HASE}(\hat{\phi}_i)}{\mathrm{ASE}(\hat{\phi}_i)} = \frac{\widehat{\mathrm{HASE}}(\hat{\phi}_i)}{\widehat{\mathrm{ASE}}(\hat{\phi}_i)} = \sqrt{1-\frac{i-1}{n}}, (i=1,\ldots,p) \qquad (23)$$

(for derivation see the appendix).

We find that (23) is a decreasing function of *i*, whose smallest value among the range of *i* for fixed *p* and *n* is $\sqrt{1-\{(p-1)/n\}}$. This value has a lower bound $n^{-1/2}$, under the condition that **S** is positive definite with probability 1, which is attained when *p*=*n*. With *n*=25, the lower bound for (23) becomes as small as $25^{-1/2}=0.2$. In such cases, though it is an extreme one, $\widehat{\mathrm{HASE}}$ has an advantage over $\widehat{\mathrm{ASE}}$ (we see that from (A37) HASE is the exact standard error in this case).

6. Discussion

In the result of Section 4, the accuracy of the formulas of HASE is well illustrated. Though the amount of $(\mathrm{HASE/ASE})-1$ was not necessarily substantial for the parameter estimators due to the moderate sample size used, the example of Section 5 shows that the usual ASEs can be misleading especially when sample sizes are small under strong nonnormality in situations similar to those in Table 2. However, we should note that the example in Section 5 may be exceptional in that the HASEs are exact ones, which does not always hold.

We have difficulty when we estimate HASEs from a random sample with unknown distribution for observed variables since a HASE generally depends on the population moments of the observed variables up to the sixth order (recall (9)). This difficulty is also shared by the usual ASEs under nonnormality though an ASE generally depends on the population moments up to the fourth order. For $\widehat{\mathrm{HASE}}$s and $\widehat{\mathrm{ASE}}$s, we can use, at least in principle, sample moments. However, it is known that higher-order sample moments tend to be unstable.

A practical method is to choose independent normal/nonnormal distribution(s) with known cumulants or moments for the latent variables e.g., in the case of factor analysis models to give marginal (preferably joint) distributions similar to the sample ones of observed variables with respect to skewness and kurtosis. By this method we can estimate the required cumulants of the observed variables without using the unstable sample higher-order moments though the estimates of the cumulants may be crude ones. The method of Mattson (1997) (see also Reinartz, Echambadi & Chin, 2002) may be useful to have the required skewnesses and kurtosis in the observed variables. For the nonnormal distributions, the generalized gamma distribution (Ramberg & Schmeiser, 1974) has a convenient property in that the three parameters in addition to the location parameter give various combinations of variances, skewnesses and kurtoses. For a similar purpose, a method using the Pearson distribution system (Nagahara, 2004) is available.

When raw observations are not available other than the corresponding sample covariance matrix, it is impossible to make inference under nonnormality since there is no information about the population moments higher than the second order. Unfortunately, this sometimes happens in practice. In such situation, when nonnormality is suspected, it is recommended to make inference using the nonnormal distributions strong enough with positive kurtosis among possible alternatives e.g., the chi-square distribution with *df*=1 whose standardized kurtosis is as large as 12 (note that the negative kurtosis as in the uniform distribution gives ASEs (and possibly HASEs) smaller than those under normality, which is not a conservative or safe choice).

For further research, when we require the asymptotic distributions of the parameter estimators beyond the usual normal approximation, the standard method is to use asymptotic expansion e.g., Edgeworth expansion (see Hall, 1992), where the HASEs or the higher-order variances are required when the two-term Edgeworth expansion is used. Recently, the method of asymptotic expansion adapted to the situations with small sample sizes have been developed by using the saddlepoint approximation (see e.g., Goutis & Casella, 1999; Stuart & Ort, 1994, Sections 11.13-11.17). In the possible application of the saddlepoint method, the HASEs are required.

Appendix

The third-order central moments of sample variances and covariances

The expectation of $(\mathbf{s}-\boldsymbol{\sigma})^{<3>}$ or the third-order central moments of **s** are given by the following formula derived by Kaplan (1952, Equation (6); see also Stuart & Ort, 1994, Chapters 12 and 13) with the assumption of its existence:

$$K(ab,cd,ef) \equiv \mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})\}$$

$$= \frac{1}{N^2}K_{abcdef} + \frac{1}{N(N-1)}\sum^{12} K_{abce}K_{df}$$

$$+ \frac{N-2}{N(N-1)^2}\sum^{4} K_{ace}K_{bdf} + \frac{1}{(N-1)^2}\sum^{8} K_{ac}K_{be}K_{df}, \qquad \text{(A1)}$$

$$(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1),$$

where K_{abcdef} is the sixth-order multivariate cumulant for the variables $X_a, X_b, ..., X_f$; K_{abce}, K_{ace} and K_{df} are similarly defined; and $\sum^{12}, \sum^{4}$ and

$\sum^{8}$ denote the sums of the associated terms, whose numbers are given over the symbol, where the symmetric property of the cumulants with respect to the variables concerned is to be considered. The cumulants on the right-hand side of (A1) are expressed by moments as

$$\begin{aligned} &K_{abcdef} = \sigma_{abcdef} - \sum^{15} \sigma_{abcd}\sigma_{ef} - \sum^{10} \sigma_{abc}\sigma_{def} + 2\sum^{15} \sigma_{ab}\sigma_{cd}\sigma_{ef}, \\ &K_{abcd} = \sigma_{abcd} - (\sigma_{ab}\sigma_{cd} + \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}), \\ &K_{abc} = \sigma_{abc},\ K_{ab} = \sigma_{ab}, \\ &(p \ge (a,b,c,d,e,f) \ge 1). \end{aligned} \tag{A2}$$

The exact third-order central moments of sample variances and covariances using moment expression are given by inserting (A2) into (A1). However, the result is involved. An asymptotically equivalent result accurate up to order $O(n^{-2})$ is given as follows. Substituting (A2) for (A1), we have

$$\begin{aligned} &K(ab,cd,ef) = n^{-2}\{\sigma_{abcdef} - \sum^{15} \sigma_{abcd}\sigma_{ef} - \sum^{10} \sigma_{abc}\sigma_{def} + 2\sum^{15} \sigma_{ab}\sigma_{cd}\sigma_{ef}, \\ &+\sum^{12} (\sigma_{abce} - \sigma_{ab}\sigma_{ce} - \sigma_{ac}\sigma_{be} - \sigma_{ae}\sigma_{bc})\sigma_{df} + \sum^{4} \sigma_{ace}\sigma_{bdf} + \sum^{8} \sigma_{ac}\sigma_{be}\sigma_{df}\} \\ &+O(n^{-3}), \qquad (p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1). \end{aligned} \tag{A3}$$

In (A3),

$$\begin{aligned} &-\sum^{15} \sigma_{abcd}\sigma_{ef} + \sum^{12} \sigma_{abce}\sigma_{df} = -(\sigma_{abcd}\sigma_{ef} + \sigma_{abef}\sigma_{cd} + \sigma_{cdef}\sigma_{ab}), \\ &-\sum^{10} \sigma_{abc}\sigma_{def} + \sum^{4} \sigma_{ace}\sigma_{bdf} \\ &= -(\sigma_{acd}\sigma_{bef} + \sigma_{bcd}\sigma_{aef} + \sigma_{abc}\sigma_{def} + \sigma_{abd}\sigma_{cef} + \sigma_{abe}\sigma_{cdf} + \sigma_{abf}\sigma_{cde}). \end{aligned} \tag{A4}$$

The remaining terms except σ_{abcdef} are summed as

$$2\sum^{15} \sigma_{ab}\sigma_{cd}\sigma_{ef} - \sum^{12} (\sigma_{ab}\sigma_{ce} + \sigma_{ac}\sigma_{be} + \sigma_{ae}\sigma_{bc})\sigma_{df} + \sum^{8} \sigma_{ac}\sigma_{be}\sigma_{df}, \tag{A5}$$

which is written as $2(A+B+C)-(2A+3B)+B=2C$, where A is the sum of the six products whose subscripts for two factors among the three factors in each product are chosen from different two pairs in (a, b), (c, d) and (e, f) (e.g., $\sigma_{ab}\sigma_{ce}\sigma_{df}$) without repetition; B is the sum of the eight products whose subscripts for the three factors in each product are given from the different three pairs in (*a*, *b*), (*c*, *d*) and (*e*, *f*) (e.g., $\sigma_{ac}\sigma_{be}\sigma_{df}$); and *C* is equal to $\sigma_{ab}\sigma_{cd}\sigma_{ef}$, which is the single product whose three factors are chosen from the same pairs in (*a*, *b*), (*c*, *d*) and (*e*, *f*). That is, (A5) is equal to $2\sigma_{ab}\sigma_{cd}\sigma_{ef}$. From (A3)-(A5), we have

Lemma 1. The asymptotic third-order central moment of s_{ab}, s_{cd} and s_{ef} with the assumption of its existence is

$$\begin{aligned}
&\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})\} \\
&= n^{-2}(\sigma_{abcdef}-\sigma_{abcd}\sigma_{ef}-\sigma_{abef}\sigma_{cd}-\sigma_{cdef}\sigma_{ab} \\
&\quad -\sigma_{acd}\sigma_{bef}-\sigma_{bcd}\sigma_{aef}-\sigma_{abc}\sigma_{def}-\sigma_{abd}\sigma_{cef} \\
&\quad -\sigma_{abe}\sigma_{cdf}-\sigma_{abf}\sigma_{cde}+2\sigma_{ab}\sigma_{cd}\sigma_{ef})+O(n^{-3}), \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A6}$$

When the multivariate normality for observed variables holds, we have the following result.

Lemma 2. Under the assumption of multivariate normality, the third-order central moment of s_{ab}, s_{cd} and s_{ef} is

$$\begin{aligned}
&\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})\} \\
&= n^{-2}\{(\sigma_{fa}\sigma_{bc}+\sigma_{fb}\sigma_{ac})\sigma_{de}+(\sigma_{fa}\sigma_{bd}+\sigma_{fb}\sigma_{ad})\sigma_{ce} \\
&\qquad +(\sigma_{ea}\sigma_{bc}+\sigma_{eb}\sigma_{ac})\sigma_{df}+(\sigma_{ea}\sigma_{bd}+\sigma_{eb}\sigma_{ad})\sigma_{cf}\}, \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A7}$$

Proof 1. When the multivariate normality holds, the cumulants higher than the second-order in (A1) vanish, which gives (A7) with $K_{ab} = \sigma_{ab}$. Q.E.D.

Note that the eight terms in (A7) comprise of the combinations of different three covariances in which each covariance should be for the variables from different pairs of (*a*, *b*), (*c*, *d*) or (*e*, *f*).

Lemma 2 is also proved without using Kaplan's (1952) result as follows.

Proof 2. For a typical element of the third-order central moment, we have

$$\begin{aligned}
&\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})\} \\
&= \mathrm{E}(s_{ab}s_{cd}s_{ef}) - \sigma_{ab}\mathrm{E}(s_{cd}s_{ef}) - \sigma_{cd}\mathrm{E}(s_{ab}s_{ef}) - \sigma_{ef}\mathrm{E}(s_{ab}s_{cd}) \\
&\quad + 2\sigma_{ab}\sigma_{cd}\sigma_{ef}, \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A8}$$

Noting that $n s_{ab}$ is written as the sum of n independently distributed products of the deviations from means for the associated variables X_a and X_b when multivariate normality holds (see e.g., Anderson, 1984, Theorem 3.3.2), we have

$$n^2\mathrm{E}(s_{ab}s_{cd}) = n\sigma_{abcd} + n(n-1)\sigma_{ab}\sigma_{cd} \tag{A9}$$

which gives

$$\begin{aligned}
&\mathrm{E}(s_{ab}s_{cd}) = n^{-1}\sigma_{abcd} + (1-n^{-1})\sigma_{ab}\sigma_{cd}, \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1).
\end{aligned} \tag{A10}$$

The remaining expected term in (A8) is given in a similar manner as

$$\begin{aligned}
&n^3\mathrm{E}(s_{ab}s_{cd}s_{ef}) = n\sigma_{abcdef} + n(n-1)(\sigma_{ab}\sigma_{cdef} + \sigma_{cd}\sigma_{abef} + \sigma_{ef}\sigma_{abcd}) \\
&\qquad + n(n-1)(n-2)\sigma_{ab}\sigma_{cd}\sigma_{ef}, \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A11}$$

From (A11),

$$\begin{aligned}
&\mathrm{E}(s_{ab}s_{cd}s_{ef}) = n^{-2}\sigma_{abcdef} + n^{-1}(1-n^{-1})(\sigma_{ab}\sigma_{cdef} + \sigma_{cd}\sigma_{abef} + \sigma_{ef}\sigma_{abcd}) \\
&\qquad + (1-n^{-1})(1-2n^{-1})\sigma_{ab}\sigma_{cd}\sigma_{ef}, \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A12}$$

Substituting (A10) and (A12) for (A8), we have

$$\begin{aligned}
&\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})\} \\
&= n^{-2}\sigma_{abcdef} + \{n^{-1}(1-n^{-1}) - n^{-1}\}(\sigma_{ab}\sigma_{cdef} + \sigma_{cd}\sigma_{abef} + \sigma_{ef}\sigma_{abcd}) \\
&\qquad + \{(1-n^{-1})(1-2n^{-1}) - 3(1-n^{-1}) + 2\}\sigma_{ab}\sigma_{cd}\sigma_{ef} \\
&= n^{-2}(\sigma_{abcdef} - \sigma_{ab}\sigma_{cdef} - \sigma_{cd}\sigma_{abef} - \sigma_{ef}\sigma_{abcd} + 2\sigma_{ab}\sigma_{cd}\sigma_{ef}), \\
&(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).
\end{aligned} \tag{A13}$$

For σ_{abcd} in (A13), under the normality assumption, it is well known that

$$\sigma_{abcd} = \sigma_{ab}\sigma_{cd} + \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}. \tag{A14}$$

For σ_{abcdef} in (A13), we can use the moment generating function of the multivariate normal distribution $\exp(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2)$ with $\mathbf{t}=(t_a,t_b,t_c,t_d,t_e,t_f)'$ and the relationship:

$$\sigma_{abcdef} = \frac{\partial^6 \exp(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2)}{\partial t_a \partial t_b \partial t_c \partial t_d \partial t_e \partial t_f}\Big|_{\mathbf{t}=\mathbf{0}}$$

$$= \frac{\partial^6}{\partial t_a \partial t_b \partial t_c \partial t_d \partial t_e \partial t_f}\left\{1+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}+\frac{1}{2}\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right)^2+\frac{1}{6}\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right)^3+\ldots\right\}\Big|_{\mathbf{t}=\mathbf{0}}$$

$$\begin{aligned}&=\sigma_{ab}\sigma_{cd}\sigma_{ef}+\sigma_{ab}\sigma_{ce}\sigma_{df}+\sigma_{ab}\sigma_{cf}\sigma_{de}+\sigma_{ac}\sigma_{bd}\sigma_{ef}+\sigma_{ac}\sigma_{be}\sigma_{df}+\sigma_{ac}\sigma_{bf}\sigma_{de}\\&+\sigma_{ad}\sigma_{bc}\sigma_{ef}+\sigma_{ad}\sigma_{be}\sigma_{cf}+\sigma_{ad}\sigma_{bf}\sigma_{ce}+\sigma_{ae}\sigma_{bc}\sigma_{df}+\sigma_{ae}\sigma_{bd}\sigma_{cf}+\sigma_{ae}\sigma_{bf}\sigma_{cd}\\&+\sigma_{af}\sigma_{bc}\sigma_{de}+\sigma_{af}\sigma_{bd}\sigma_{ce}+\sigma_{af}\sigma_{be}\sigma_{cd}.\end{aligned} \tag{A15}$$

Note that the above $15\,(={}_6C_2\,{}_4C_2\,{}_2C_2/3!=5\times3\times1)$ terms comprise of the combinations of different three covariances of different (in notation) variables in the six variables. Since the result of (A15) may seem intractable, the corresponding matrix expression, if necessary, is available using the symmetrizer or symmetric tensor (see Holmquist, 1988; Kano, 1997). Replacing $\sigma_{abcdef}, \sigma_{cdef}, \sigma_{abef}$ and σ_{abcd} in (A13) by the corresponding results of (A14) and (A15) yields (A7). Q.E.D.

A similar result of Lemma 2 is given by Siotani, Hayakawa & Fujikoshi (1985, Problem 4.3.4, p.183) with an added term $O(n^{-3/2})$ without proof. We find that the added term can be omitted.

The fourth-order central moments of sample variances and covariances

The exact fourth-order central moments of sample variances and covariances can be given by using the result of Kaplan (1952, Equation (9)), which is summarized as

$$\begin{aligned}K(ab,cd,ef,gh) &= \mathrm{E}[(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})(s_{gh}-\sigma_{gh})]\\&\quad-\{K(ab,cd)K(ef,gh)+K(ab,ef)K(cd,gh)\\&\quad+K(ab,gh)K(cd,ef)\}\\&=O(n^{-3}),\end{aligned} \tag{A16}$$

$$(p\ge a\ge b\ge 1;\ p\ge c\ge d\ge 1;\ p\ge e\ge f\ge 1;\ p\ge g\ge h\ge 1),$$

where $K(ab,cd,ef,gh)$ is the fourth-order multivariate cumulant of

s_{ab}, s_{cd}, s_{ef} and s_{gh}; and $K(ab,cd)$ is the second-order bivariate cumulant of s_{ab} and s_{cd} i.e.,

$$\begin{aligned}K(ab,cd) &= \mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})\}\\ &= \frac{1}{N}K_{abcd} + \frac{1}{N-1}(K_{ac}K_{bd}+K_{ad}K_{bc})\\ &= \frac{1}{N}(\sigma_{abcd}-\sigma_{ab}\sigma_{cd}-\sigma_{ac}\sigma_{bd}-\sigma_{ad}\sigma_{bc}) + \frac{1}{N-1}(\sigma_{ac}\sigma_{bd}+\sigma_{ad}\sigma_{bc}),\\ &(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1),\end{aligned} \quad \text{(A17)}$$

(see e.g., Kaplan, 1952, Equation (3)). Though the explicit expression of $O(n^{-3})$ in (A16) is available, it is involved. We note, however, that the terms in braces on the right-hand side of (A16) is of order $O(n^{-2})$ with

$$K(ab,cd) = n^{-1}(\sigma_{abcd}-\sigma_{ab}\sigma_{cd}) + O(n^{-2}) = n^{-1}\omega_{ab,cd} + O(n^{-2}), \quad \text{(A18)}$$

which gives

Lemma 3. The fourth-order central moment of s_{ab}, s_{cd}, s_{ef} and s_{gh} up to order $O(n^{-2})$ with the assumption of its existence is

$$\begin{aligned}&\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})(s_{gh}-\sigma_{gh})\}\\ &= n^{-2}(\omega_{ab,cd}\omega_{ef,gh}+\omega_{ab,ef}\omega_{cd,gh}+\omega_{ab,gh}\omega_{cd,ef}) + O(n^{-3}),\\ &(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1;\ p \ge g \ge h \ge 1).\end{aligned} \quad \text{(A19)}$$

Since $\sqrt{n}(s_{ab}-\sigma_{ab})$, $\sqrt{n}(s_{cd}-\sigma_{cd})$, $\sqrt{n}(s_{ef}-\sigma_{ef})$ and $\sqrt{n}(s_{gh}-\sigma_{gh})$ are asymptotically normally distributed, the fourth-order central moments of these terms are given from the asymptotic covariances between the terms as follows:

$$\begin{aligned}&n^2\mathrm{E}\{(s_{ab}-\sigma_{ab})(s_{cd}-\sigma_{cd})(s_{ef}-\sigma_{ef})(s_{gh}-\sigma_{gh})\}\\ &= n\,\mathrm{acov}(s_{ab},s_{cd})\,n\,\mathrm{acov}(s_{ef},s_{gh})\\ &\quad + n\,\mathrm{acov}(s_{ab},s_{ef})\,n\,\mathrm{acov}(s_{cd},s_{gh})\\ &\quad\quad + n\,\mathrm{acov}(s_{ab},s_{gh})\,n\,\mathrm{acov}(s_{cd},s_{ef}) + O(n^{-1}),\end{aligned} \quad \text{(A20)}$$

which gives the equivalent expression of (A19).

The partial derivatives of the discrepancy function for the MLE

Let

$$\hat{G} = \hat{F} + \hat{\boldsymbol{\xi}}'\hat{\mathbf{h}} \quad \text{and} \quad \hat{F} = \log|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| + \mathrm{tr}\{\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{S}\} - p \qquad \text{(A21)}$$

with. $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\theta}}', \hat{\boldsymbol{\xi}}')'$

Then, the nonzero second partial derivatives evaluated at population values are

$$\frac{\partial^2 G}{\partial\theta_i \partial\theta_j} = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right),\ (i, j = 1, \ldots, q), \qquad \text{(A22)}$$

$$\frac{\partial^2 G}{\partial\theta_i \partial\xi_j} = \frac{\partial h_j}{\partial\theta_i},\ (i = 1, \ldots, q;\ j = 1, \ldots, r), \qquad \text{(A23)}$$

$$\frac{\partial^2 G}{\partial\theta_i \partial\sigma_{ab}} = -\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\right)_{ab}(2 - \delta_{ab}),\ (i = 1, \ldots, q;\ p \geq a \geq b \geq 1), \qquad \text{(A24)}$$

where δ_{ab} is the Kronecker delta.

For the nonzero third partial derivatives, noting

$$\frac{\partial^2 \hat{G}}{\partial\hat{\theta}_i \partial\hat{\theta}_j} = \mathrm{tr}\left\{\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j} + \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j}\right.$$

$$\left. - 2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i}\right\} + \hat{\boldsymbol{\xi}}'\frac{\partial^2\hat{\mathbf{h}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j},\ (i, j = 1, \ldots, q), \qquad \text{(A25)}$$

where $\hat{\boldsymbol{\Sigma}}^{-1} = \{\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\}^{-1}$, we have

$$\frac{\partial^3 G}{\partial\theta_i\partial\theta_j\partial\theta_k} = \mathrm{tr}\left(-4\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\right.$$

$$+ \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k} + \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j} \qquad \text{(A26)}$$

$$\left. + \boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\right),$$

$$\frac{\partial^3 G}{\partial\theta_i\partial\theta_j\partial\sigma_{ab}} = \left(-\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1} \right.$$

$$\left. + \boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1} \right)_{ab} (2-\delta_{ab}), \tag{A27}$$

$(i, j, k = 1,\ldots,q;\ p \geq a \geq b \geq 1)$,

$$\frac{\partial^3 G}{\partial\theta_i\partial\theta_j\partial\xi_k} = \frac{\partial^2 h_k}{\partial\theta_i\partial\theta_j},\ (i, j = 1,\ldots,q; k = 1,\ldots,r). \tag{A28}$$

For the nonzero fourth partial derivatives, we note

$$\frac{\partial^3 \hat{G}}{\partial\hat{\theta}_i\partial\hat{\theta}_j\partial\hat{\theta}_k} = \mathrm{tr}\left(-4\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k} + \hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k} \right.$$

$$+\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j} + \hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i} + \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^3\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j\partial\hat{\theta}_k}$$

$$+2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i} + 2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i}$$

$$+2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i} - 2\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j}$$

$$\left. -2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j\partial\hat{\theta}_k}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i} - 2\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_j}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}}-\mathbf{S})\hat{\boldsymbol{\Sigma}}^{-1}\frac{\partial^2\hat{\boldsymbol{\Sigma}}}{\partial\hat{\theta}_i\partial\hat{\theta}_k} \right)$$

$$+\hat{\boldsymbol{\xi}}'\frac{\partial^3\hat{\mathbf{h}}}{\partial\hat{\theta}_i\partial\hat{\theta}_j\partial\hat{\theta}_k}, \tag{A29}$$

which yields

$$\frac{\partial^4 G}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\theta_l} = \mathrm{tr}\left(6\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\right.$$

$$+6\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}+6\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}$$

$$-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}$$

$$-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}$$

$$-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}-4\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}$$

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

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

$$\left.+\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\boldsymbol{\Sigma}^{-1}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j\partial\theta_k}\right), \qquad \text{(A30)}$$

$$\frac{\partial^4 G}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\sigma_{ab}} = \left(2\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}^{-1}+2\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_k}\boldsymbol{\Sigma}^{-1}\right.$$

$$+2\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j}\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_j\partial\theta_k}\boldsymbol{\Sigma}^{-1}$$

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

$$\left.-2\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}\right)_{ab+ba}\frac{2-\delta_{ab}}{2}, \qquad \text{(A31)}$$

$(i, j, k, l = 1, ..., q, \; p \ge a \ge b \ge 1)$,

$$\frac{\partial^4 G}{\partial\theta_i\partial\theta_j\partial\theta_k\partial\xi_l} = \frac{\partial^3 h_l}{\partial\theta_i\partial\theta_j\partial\theta_k}, \; (i, j, k = 1, ..., q; l = 1, ..., r), \qquad \text{(A32)}$$

where $(\cdot)_{ab+ba}$ is the sum of the (*a*, *b*)th and (*b*, *a*)th elements of the argument matrix.

For zero partial derivatives, we note

$$\frac{\partial G}{\partial \boldsymbol{\eta}} = \mathbf{0}, \quad \left(\frac{\partial}{\partial \boldsymbol{\xi}}\right)^{<u>} G = \mathbf{0},\ (u \ge 2),$$
$$\left(\frac{\partial}{\partial \boldsymbol{\xi}}\right)^{<u>} \left(\frac{\partial}{\partial \boldsymbol{\theta}'}\right)^{<v>} G = \mathbf{O},\ (u \ge 2; v \ge 1), \tag{A33}$$
$$\left(\frac{\partial}{\partial \boldsymbol{\xi}}\right)^{<u>} \left(\frac{\partial}{\partial \boldsymbol{\sigma}'}\right)^{<v>} G = \mathbf{O},\ (u \ge 1; v \ge 1),$$

which hold, if they exist, irrespective of types of estimators because of the linear property of *G* in terms of the Lagrange multipliers (we do not consider the unusual case of $\mathbf{h}(\boldsymbol{\theta})$ including **s**), and similarly

$$\left(\frac{\partial}{\partial \boldsymbol{\sigma}}\right)^{<u>} G = \mathbf{0},\ (u \ge 2), \left(\frac{\partial}{\partial \boldsymbol{\sigma}}\right)^{<u>} \left(\frac{\partial}{\partial \boldsymbol{\theta}'}\right)^{<v>} G = \mathbf{O},\ (u \ge 2; v \ge 1) \tag{A34}$$

due to the linear property of *G* for the MLEs with respect to **s**.

The equivalence of the sample and population HASE/ASE ratios in a component model under normality

Let **S** be the $p \times p$ sample counterpart of (22). Then, from the property of the Cholesky decomposition, we have

$$\hat{\phi}_1 = s_{11},\ \hat{\phi}_2 = s_{22} - (s_{21}^2 / s_{11}) \equiv s_{22 \cdot 1},$$
$$\hat{\phi}_3 = s_{33} - \frac{s_{31}^2}{s_{11}} - \frac{\{s_{32} - (s_{21} s_{31} / s_{11})\}^2}{s_{22} - (s_{21}^2 / s_{11})} \equiv s_{33 \cdot 12}, \tag{A35}$$

which can be generalized as $\hat{\phi}_i = s_{ii \cdot 12 \ldots (i-1)},\ (i = 1, \ldots, p)$ with $s_{11 \cdot 0} \equiv s_{11}$, where $s_{ii \cdot 12 \ldots (i-1)}$ is the sample residual variance of the *i*-th variable after removing the variation of the variables with indexes $1, 2, \ldots, (i-1)$.

Let

$$\hat{\phi}_i = s_{ii \cdot 12 \ldots (i-1)} = \frac{\sigma_{ii \cdot 12 \ldots (i-1)}}{n} \frac{n\, s_{ii \cdot 12 \ldots (i-1)}}{\sigma_{ii \cdot 12 \ldots (i-1)}},\ (i = 1, \ldots, p), \tag{A36}$$

where $\sigma_{ii\cdot 12...(i-1)}$, $(i=1,..., p)$ are the population values of (A36). Then, it is known that under normality $n\, s_{ii\cdot 12...(i-1)} / \sigma_{ii\cdot 12...(i-1)}$ is chi-square distributed with $df = n-i+1 = N-i$ (see Anderson, 1984, Theorem 4.3.4). From this result with (A36), we have the exact variance:

$$\begin{aligned}\operatorname{var}(\hat{\phi}_i) &= \frac{\sigma^2_{ii\cdot 12...(i-1)}}{n^2} 2(n-i+1) \\ &= \left(\frac{1}{n} - \frac{i-1}{n^2}\right) 2\sigma^2_{ii\cdot 12...(i-1)},\ (i=1,..., p).\end{aligned} \tag{A37}$$

The corresponding usual asymptotic variance is given as $n^{-1}2\sigma^2_{ii\cdot 12...(i-1)}$. Note that in the numerical example with (22), $\sigma^2_{ii\cdot 12...(i-1)}$'s were set to one, which gives $\mathrm{ASE}(\hat{\phi}_i) = \sqrt{25^{-1} \times 2} \doteq .2828$ in Table 2. From (7) and (A37), the term $\Delta\mathrm{avar}(\hat{\phi}_i; n^{-2})$ is given as

$$\begin{aligned}\Delta\mathrm{avar}(\hat{\phi}_i; n^{-2}) &= \operatorname{var}(\hat{\phi}_i) - \mathrm{avar}(\hat{\phi}_i; n^{-1}) \\ &= -\frac{i-1}{n^2} 2\sigma^2_{ii\cdot 12...(i-1)},\ (i=1,..., p).\end{aligned} \tag{A38}$$

Consequently, we have

$$\begin{aligned}\frac{\mathrm{HASE}(\hat{\phi}_i)}{\mathrm{ASE}(\hat{\phi}_i)} &= \frac{\widehat{\mathrm{HASE}}(\hat{\phi}_i)}{\widehat{\mathrm{ASE}}(\hat{\phi}_i)} \\ &= \sqrt{\frac{1}{n} - \frac{i-1}{n^2}} \Big/ \sqrt{\frac{1}{n}} = \sqrt{1 - \frac{i-1}{n}}, (i=1,..., p).\end{aligned} \tag{A39}$$

It is seen that the values 1.000, .9798 and .9592 for HASE/ASE and $\widehat{\mathrm{HASE}} / \widehat{\mathrm{ASE}}$ of $\hat{\phi}_i$'s under normality in Tables 2 and 3 are given also from (A39) with n=25.

References

Amemiya, T. (1980). The n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimator. *Annals of Statistics, 8*, 488-505.

Anderson, T. W. (1984). *An introduction multivariate statistical analysis* (2nd ed.). New York: Wiley.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SPSS Inc.
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: BMDP Statistical Software.
Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp.72-141). Cambridge: Cambridge University Press.
Browne, M. W., & Mels, G. (2000). Path analysis (RAMONA). In SPSS Inc. *SYSTAT 10 statistics II*, Chapter 7, (pp. II233-291). Chicago: Author.
Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variable models. *British Journal of Mathematical and Statistical Psychology, 41*, 193-208.
Ghosh, J. K. (1994). *Higher order asymptotics*. NSF-CBMS regional conference series in probability and statistics (Vol.4). Hayward, CA: Institute of Mathematical Statistics.
Goutis, C., & Casella, G. (1999). Explaining the saddlepoint approximation. *American Statistician, 53*, 216-224.
Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer. Corrected printing, 1997.
Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications, 6*, 273-278.
Jöreskog, K. G., & Sörbom, D. (1996). *LISREL8: User's reference guide* (2nd ed.). Chicago : Scientific Software International.
Kano, Y. (1997). Beyond third-order efficiency. *Sankhya, B, 59*, 179-197.
Kano, Y. (1998). More higher-order efficiency. *Journal of Multivariate Analysis, 67*, 349-366.
Kaplan, E. L. (1952). Tensor notation and the sampling cumulants of *k*-statistics. *Biometrika, 39*, 319-323.
Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research, 32*, 355-373.
Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angels: Authors.
Nagahara, Y. (2004). A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. *Computational Statistics and Data Analysis, 47*, 1-29.
Ogasawara, H. (2004a). Asymptotic biases in exploratory factor analysis and

structural equation modeling. *Psychometrika*, *69*, 235-256.
Ogasawara, H. (2004b). Asymptotic biases of least squares estimators in structural equation modeling. In S. P. Shohov (Ed.), *Advances in psychology research Vol.27*. (pp. 65-94). New York: Nova.
Ogasawara, H. (2004, October). Higher-order estimation error in factor analysis and structural equation modeling under nonnormality. In Y. Kano (Organizer), *Proceedings of the 83rd Symposium of Behaviormetrics: Factor Analysis Centennial Symposium* (pp.39-53). Osaka University, Osaka, Japan.
Ramberg, J. S., & Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the Association for Computing Machinery*, *17*, 78-82.
Reinartz, W. R., Echambadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research*, *37*, 227-244.
Rothenberg, T. J. (1984). Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol.II, pp.882-935). New York: Elsevier.
Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (A unified approach). *South African Statistical Journal*, *17*, 33-81.
Siotani, M., Hayakawa, T., & Fujikoshi, Y. (1985). *Modern multivariate statistical analysis: A graduate course and handbook*. Columbus, OH: American Sciences Press.
Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics: Distribution theory* (6th ed., Vol.1). London: Arnold.

Table 1. Higher-order errors of the confirmatory oblique factor model for unstandardized variables under normality/nonnormality (N=300; Number of replications=1,000,000)

	Normal		Chi-square (df) (3)		Chi-square (df) (1)	
	ASE	HASE	ASE	HASE	ASE	HASE
ψ_1	1.673	1.675	2.671	2.664	3.975	3.960
ψ_2	2.500	2.523	3.329	3.339	4.554	4.552
ψ_3	4.114	4.183	4.720	4.771	5.743	5.770
ψ_4	2.027	2.030	3.164	3.156	4.670	4.652
ψ_5	2.886	2.905	3.848	3.852	5.268	5.257
ψ_6	4.380	4.429	5.124	5.153	6.357	6.357
ϕ_{11}	.1345	.1354	.1774	.1777	.2413	.2411
ϕ_{21}	.0799	.0801	.0986	.0988	.1281	.1283
ϕ_{22}	.1238	.1242	.1539	.1541	.2010	.2010
λ_{11}	.3364	.3410	*	.3448	*	.3523
λ_{21}	.4727	.4830	*	.4893	*	.5017
λ_{42}	.3657	.3700	*	.3722	*	.3765
λ_{52}	.4804	.4886	*	.4914	*	.4970
	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$
ψ_1	1.0013	1.0019	.9975	.9974	.9962	.9972
ψ_2	1.0093	1.0098	1.0033	1.0037	.9996	.9995
ψ_3	1.0167	1.0190	1.0108	1.0127	1.0048	1.0058
ψ_4	1.0012	1.0028	.9976	.9986	.9962	.9960
ψ_5	1.0066	1.0064	1.0012	1.0011	.9979	.9992
ψ_6	1.0113	1.0113	1.0055	1.0058	1.0000	1.0009
ϕ_{11}	1.0065	1.0074	1.0021	1.0041	.9993	1.0003
ϕ_{21}	1.0035	1.0035	1.0025	1.0031	1.0017	1.0025
ϕ_{22}	1.0034	1.0041	1.0015	1.0014	1.0001	1.0012
λ_{11}	1.0136	1.0141	1.0249	1.0269	1.0472	1.0522
λ_{21}	1.0218	1.0230	1.0351	1.0382	1.0613	1.0692
λ_{42}	1.0117	1.0136	1.0177	1.0185	1.0296	1.0324
λ_{52}	1.0172	1.0184	1.0230	1.0232	1.0346	1.0375

Note. ASE=$\sqrt{\text{avar}(\hat{\theta}_i; n^{-1})}$, HASE=$\sqrt{\text{avar}(\hat{\theta}_i; n^{-1}, n^{-2})}$, SD=Standard deviations from simulation. The asterisks denote that the corresponding normal theory asymptotic standard errors hold.

Table 2. Theoretical and simulated ratios of the higher-order asymptotic standard errors to their corresponding usual ones for the saturated model in normal/nonnormal samples (*n*=25; Number of replications=1,000,000)

	ASE	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$	ASE	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$
		Normal			Uniform	
ϕ_1	.2828	1.0000	.9997	.1789	1.0296	1.0296
ϕ_2	.2828	.9798	.9795	.1789	1.0392	1.0327
ϕ_3	.2828	.9592	.9600	.1789	1.0488	1.0363
λ_{21}	.2000	1.0392	1.0434	*	1.0159	1.0179
λ_{31}	.2040	1.0392	1.0417	*	1.0159	1.0184
λ_{32}	.2000	1.0583	1.0652	*	1.0354	1.0420
		t (*df*=9)			Chi-square (*df*=10)	
ϕ_1	.3578	.9925	.9916	.3578	.9925	.9923
ϕ_2	.3578	.9644	.9669	.3578	.9644	.9674
ϕ_3	.3578	.9354	.9394	.3578	.9354	.9399
λ_{21}	*	1.0621	1.0623	*	1.0621	1.0651
λ_{31}	*	1.0621	1.0611	*	1.0621	1.0633
λ_{32}	*	1.0807	1.0854	*	1.0807	1.0878
		Chi-square (*df*=3)			Chi-square (*df*=1)	
ϕ_1	.4899	.9866	.9862	.7483	.9827	.9803
ϕ_2	.4899	.9522	.9564	.7483	.9442	.9490
ϕ_3	.4899	.9165	.9267	.7483	.9040	.9215
λ_{21}	*	1.1136	1.1102	*	1.2490	1.2618
λ_{31}	*	1.1136	1.1113	*	1.2490	1.2585
λ_{32}	*	1.1314	1.1137	*	1.2649	1.2881

Note. ASE=$\sqrt{\text{avar}(\hat{\theta}_i;n^{-1})}$, HASE=$\sqrt{\text{avar}(\hat{\theta}_i;n^{-1},n^{-2})}$, SD= Standard deviations from simulation, $n = N - 1$. The asterisks denote that the corresponding normal theory asymptotic standard errors hold.

Table 3. Means and standard deviations of the estimates of the usual and higher-order asymptotic standard errors for the saturated model in normal samples (Number of replications=100,000)

	$\widehat{\text{ASE}}$		$\widehat{\text{HASE}}$		$\frac{m \text{ of } \widehat{\text{HASE}}}{m \text{ of } \widehat{\text{ASE}}}$
	m	*sd*	*m*	*sd*	
(*n*=25)					
ϕ_1	.2826	.0800	.2826	.0800	1.0000
ϕ_2	.2721	.0787	.2666	.0771	.9798
ϕ_3	.2601	.0766	.2494	.0735	.9592
λ_{21}	.2003	.0418	.2081	.0434	1.0392
λ_{31}	.2040	.0425	.2120	.0441	1.0392
λ_{32}	.1998	.0426	.2114	.0451	1.0583
(*n*=50)					
ϕ_1	.1997	.0399	.1997	.0399	1.0000
ϕ_2	.1961	.0395	.1942	.0391	.9899
ϕ_3	.1918	.0392	.1879	.0384	.9798
λ_{21}	.1416	.0205	.1444	.0209	1.0198
λ_{31}	.1442	.0208	.1471	.0213	1.0198
λ_{32}	.1413	.0206	.1455	.0212	1.0296

Note. *m* and *sd* of $\widehat{\text{ASE}}$ and $\widehat{\text{HASE}}$ =Means and standard deviations of the estimates of ASE and HASE from simulation, $n = N - 1$.