CORRELATIONS AMONG MAXIMUM LIKELIHOOD AND WEIGHTED/UNWEIGHTED LEAST SQUARES ESTIMATORS IN FACTOR ANALYSIS

Haruhiko Ogasawara*

The asymptotic correlations among maximum likelihood (ML) and various least squares (LS) estimators in factor analysis are derived. The LS estimators include the unweighted (ULS) and weighted estimators for unstandardized variables and the ULS estimators for standardized variables. The derived formulas cover the cases with restrictions on parameters. Numerical examples with simulations are provided to confirm the accuracy of the formulas and the influence of scales on the asymptotic correlations.

1. Introduction

We have various methods of estimation in factor analysis (FA) or more generally in structural equation modeling (SEM). Among others are maximum likelihood (ML), generalized least squares (GLS) and unweighted least squares (ULS) estimation, which are widely used and employed by the familiar programs e.g., LISREL 8 (Jöreskog & Sörbom, 1996; see also Jöreskog, Sörbom, du Toit & du Toit, 1999), EQS (Bentler, 1989), Amos 4.0 (Arbuckle & Wothke, 1999) and RAMONA (Browne & Mels, 2000). The so-called ML estimator is based on the assumption of multivariate normality for observed variables or equivalently the Wishart distribution of a sample covariance matrix. However, it is known that the Wishart ML (WML) estimators have robustness against the violation of multivariate normality under fairly mild conditions for various models including the factor analysis model (see Anderson, 1987, 1989; Anderson & Amemiya, 1988; Anderson & Rubin, 1956; Browne, 1987; Browne & Shapiro, 1988; Mooijart & Bentler, 1991; Satorra, 1990; Satorra & Bentler, 1990).

The GLS estimator using the appropriate weight matrix with the assumption of multivariate normality is asymptotically equivalent to the WML estimator (see e.g., Browne, 1974/1977). From this equivalence, the GLS estimation also enjoys the robustness against the violation of normality mentioned above. The GLS estimation can be adapted to the situation with unknown distributions (Browne, 1982, 1984), though we need large sample sizes for stable estimation.

On the other hand, the statistical properties of the ULS and weighted LS (except for GLS) estimators in SEM have not been so well investigated. ULS estimation is included in the programs probably because it is simple and the form of the ULS discrepancy function is free from distribution types with some loss of

Key Words and Phrases: asymptotic correlations, efficiency, maximum likelihood estimation, least squares estimation, restrictions, factor analysis, structural equation modeling.

* Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 Japan. E-mail address: hogasa@res.otaru-uc.ac.jp

efficiency. The standard errors of the LS estimators (Browne, 1982) are given by the programs (see e.g., Jöreskog et al., 1999, p.195). It is known that the ULS estimator is not scale free (see Krane & McDonald, 1978), while the ML and GLS estimators are generally scale free.

The purpose of this study is to derive the asymptotic correlations among the ML and various weighted/unweighted LS estimators in FA. The asymptotic correlations are given as indices to show the degrees of similarity between different estimators in FA or SEM. Note that even if two estimators have the same asymptotic distributions with the same standard errors, generally the correlation of the two estimators may not be the same. In the behavioral sciences, observed variables are frequently standardized. That is, sample correlation matrices are often used as input data in FA. The asymptotic correlations between the ML estimators constructed appropriately for standardized variables and the LS estimators using sample correlations will also be given. Numerical examples with simulations will be provided to see the accuracy of our formulas and illustrate the asymptotic correlations for typical FA models in situations with normally or nonnormally distributed observed variables using finite sample sizes. Note that the computation of the asymptotic correlation is free from the sample sizes though the correlation is given when sample sizes are large. It will be illustrated that the ULS estimator gives almost the same behavior as the ML estimator when the variances of observed variables are approximately the same with normally or nonnormally distributed variables. The influence of scale differences in observed variables on the LS estimators will also be shown. In situations with substantially different scales for observed variables, the simple or scale free LS estimator (SLS, see e.g., Arbuckle & Wothke, 1999, p.392) will be shown to be promising.

2. Asymptotic Correlations Between Two Estimators

Let f_1 and f_2 be two possibly different discrepancy functions with necessary conditions for such functions (see Browne, 1982, p.81), which include those for the ML and various LS estimators. Let θ be a $q \times 1$ parameter vector in the $p \times p$ covariance matrix $\Sigma(\theta)$ for p observed variables. It is assumed that θ has r restrictions such that the $r \times 1$ vector $\mathbf{h}(\theta) = \mathbf{0}$. Let

$$\mathbf{g}_{(i)}(\hat{\theta}_{(i)}) = \frac{\partial f_i}{\partial \theta} \Big|_{\theta=\hat{\theta}_{(i)}} \equiv \frac{\partial f_i}{\partial \hat{\theta}_{(i)}}, \quad (i = 1, 2), \quad (1)$$

where $\hat{\theta}_{(i)}$ is the estimator of θ_0 (the population value of θ), which minimizes f_i , ($i = 1, 2$ or e.g., $i \in (\text{ML}, \text{ULS})$ will also be used to identify the estimation method). Then, the necessary conditions for the estimates of θ_0 , which minimize f_1 or f_2 , are

$$(\mathbf{g}'_{(i)}(\hat{\theta}_{(i)}) + \hat{\zeta}'_{(i)} \frac{\partial \mathbf{h}(\hat{\theta}_{(i)})}{\partial \hat{\theta}'_{(i)}}), \mathbf{h}'(\hat{\theta}_{(i)})' = \mathbf{0}, \quad (i = 1, 2), \quad (2)$$

where $\hat{\zeta}_{(i)}$ is the estimate of the $r \times 1$ Lagrange multiplier vector ζ_0 by the i -th estimation method.

Note that f_i is a function of $\Sigma(\theta)$ and S , where S is an unbiased sample covariance matrix based on $n + 1 = N$ observations. Let $\mathbf{s}$ be $\mathbf{v}(S)$ or $\mathbf{v}S$, where $\mathbf{v}(X)$ or $\mathbf{v}X$ is the $p(p + 1)/2 \times 1$ vector consisting of the nonduplicated elements of a $p \times p$ symmetric matrix X . Noting that $\hat{\theta}_{(i)}$ and $\hat{\zeta}_{(i)}$ are implicit functions of S or equivalently $\mathbf{s}$ with the assumption of the existence of such functions, it follows from (1) and (2) that

$$\begin{aligned} & \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \hat{\theta}_{(i)}'} \frac{\partial \hat{\theta}_{(i)}}{\partial \mathbf{s}'} + \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \mathbf{s}'} + \frac{\partial \mathbf{h}'(\hat{\theta}_{(i)})}{\partial \hat{\theta}_{(i)}} \frac{\partial \hat{\zeta}_{(i)}}{\partial \mathbf{s}'} \\ & + \hat{\zeta}_{(i)}' \frac{\partial^2 \mathbf{h}(\hat{\theta}_{(i)})}{\partial \hat{\theta}_{(i)} \partial \hat{\theta}_{(i)}'} \frac{\partial \hat{\theta}_{(i)}}{\partial \mathbf{s}'} = O, \\ & \frac{\partial \mathbf{h}(\hat{\theta}_{(i)})}{\partial \hat{\theta}_{(i)}'} \frac{\partial \hat{\theta}_{(i)}}{\partial \mathbf{s}'} = O, \quad (i = 1, 2), \end{aligned} \quad (3)$$

where $\mathbf{s}$ and its functions are temporarily used as mathematical variables for simplicity of notation followed by evaluation at their corresponding sample values as in (1). Since the fourth term in the first equation of (3) is of order smaller than the third term (see e.g., Silvey, 1975, p.80), we have

$$\begin{aligned} \begin{bmatrix} \frac{\partial \hat{\theta}_{(i)}}{\partial \mathbf{s}'} \\ \frac{\partial \hat{\zeta}_{(i)}}{\partial \mathbf{s}'} \end{bmatrix} &= - \begin{bmatrix} \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \hat{\theta}_{(i)}'} & \frac{\partial \mathbf{h}'(\hat{\theta}_{(i)})}{\partial \hat{\theta}_{(i)}} \\ \frac{\partial \mathbf{h}(\hat{\theta}_{(i)})}{\partial \hat{\theta}_{(i)}'} & O \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \mathbf{s}'} \\ O \end{bmatrix} + o_p(1) \\ &= - \begin{bmatrix} H_i^*(\hat{\theta}_{(i)}) & \# \\ \# & \# \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \mathbf{s}'} \\ O \end{bmatrix} + o_p(1), \quad (i = 1, 2), \end{aligned} \quad (4)$$

where the existence of the inverted matrix is assumed, and $\#$ s denote submatrices which will not directly be used here. From (4), we have

$$\frac{\partial \hat{\theta}_{(i)}}{\partial \mathbf{s}'} = -H_i^*(\hat{\theta}_{(i)}) \frac{\partial^2 f_i}{\partial \hat{\theta}_{(i)} \partial \mathbf{s}'} + o_p(1), \quad (i = 1, 2). \quad (5)$$

Equation (5) represents a matrix of the coefficients or constants in the

asymptotically linear function $\hat{\theta}_{(i)}$ with respect to $\mathbf{s}$ when evaluated at their corresponding population values omitting higher order terms (i.e., $\hat{\theta}_{(i)} - \theta_0 = (\partial \hat{\theta}_{(i)} / \partial \mathbf{s}')|_{\mathbf{s}=\sigma_0, \hat{\theta}_{(i)}=\theta_0}(\mathbf{s} - \sigma_0) + o_p(1/\sqrt{n})$ with $\sigma_0 = \mathbf{v}\Sigma_0 = \mathbf{v}(\Sigma(\theta_0))$).

By using the delta method, the asymptotic covariance matrix between $\hat{\theta}_{(i)}$ and $\hat{\theta}_{(j)}$, $(i, j = 1, 2)$ with the assumption of the existence of the finite fourth-order moments for the observed variables becomes

$$\begin{aligned} \text{acov}(\hat{\theta}_{(i)}, \hat{\theta}_{(j)}) &= H_i^*(\theta_0) \frac{\partial^2 f_i}{\partial \theta_0 \partial \sigma'_0} E\{(\mathbf{s} - \sigma_0)(\mathbf{s} - \sigma_0)'\} \\ &\quad \times \frac{\partial^2 f_j}{\partial \sigma_0 \partial \theta'_0} H_j^*(\theta_0) \\ &= H_i^*(\theta_0) \frac{\partial^2 f_i}{\partial \theta_0 \partial \sigma'_0} \text{acov}(\mathbf{s}) \frac{\partial^2 f_j}{\partial \sigma_0 \partial \theta'_0} H_j^*(\theta_0), \quad (i = 1, 2), \end{aligned} \quad (6)$$

where the expectation is taken in large samples and $\partial^2 f_i / \partial \theta_0 \partial \sigma'_0 = \partial^2 f_i / \partial \theta \partial \sigma' |_{\theta=\theta_0, \mathbf{s}=\sigma_0}$, $(i = 1, 2)$.

It is to be noted that (6) holds irrespective of the types of distributions for observed variables. An element of $\text{acov}(\mathbf{s})$ in (6) is

$$\text{acov}(s_{ij}, s_{kl}) = \frac{\sigma_{0ijkl} - \sigma_{0ij} \sigma_{0kl}}{n} \quad (7)$$

(see Browne, 1982, 1984), where σ_{0ijkl} is the fourth-order central moment with respect to the (i, j, k, l) th variables and σ_{0ij} is the (i, j) th element of Σ_0 , $(i, j, k, l = 1, \dots, p)$. Under the assumption of multivariate normality, (7) becomes

$$\text{acov}(s_{ij}, s_{kl}) = \frac{\sigma_{0ik} \sigma_{0jl} + \sigma_{0il} \sigma_{0jk}}{n} \quad (8)$$

and (6) can further be simplified for the ML and GLS estimators.

3. ML and Weighted/Unweighted LS Estimators

The discrepancy function for ML estimation is

$$f_{\text{ML}} = \ln |\Sigma(\theta)| + \text{tr}(\Sigma(\theta)^{-1} S) + \text{constant}. \quad (9)$$

From (9),

$$\begin{aligned} \frac{\partial^2 f_{\text{ML}}}{\partial \theta \partial \sigma'} \Big|_{\theta=\theta_0, \mathbf{s}=\sigma_0} &= - \frac{\partial \text{vec}' \Sigma}{\partial \theta_0} \frac{\partial \text{vec} (\Sigma_0^{-1} S \Sigma_0^{-1})}{\partial \sigma'} \\ &= - \frac{\partial \sigma'}{\partial \theta_0} D'_p (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) D_p, \end{aligned} \quad (10)$$

where $\text{vec}(X)$ or $\text{vec}X$ is the vectorizing operator which stacks the columns of the matrix X sequentially; $\otimes$ denotes the Kronecker product; D_p is the matrix satisfying $\text{vec} \Sigma = D_p \sigma$; and the partial derivatives with respect to θ_0 are those of θ set at $\theta = \theta_0$. Let I_{ML} be the information matrix per observation. Then,

$$I_{\text{ML}} = \frac{1}{2} H_{\text{ML}}(\theta_0) \equiv \frac{1}{2} \frac{\partial^2 f_i}{\partial \theta \partial \theta'} \Big|_{\theta=\theta_0, \mathbf{s}=\sigma_0} \quad (11)$$

and the augmented information matrix is

$$I_{\text{MLa}} = \frac{1}{2} \begin{bmatrix} H_{\text{ML}}(\theta_0) & \partial \mathbf{h}'(\theta)/\partial \theta_0 \\ \partial \mathbf{h}(\theta)/\partial \theta_0' & O \end{bmatrix} \quad (12)$$

where $\partial \mathbf{h}(\theta)/\partial \theta_0' = \partial \mathbf{h}(\theta)/\partial \theta'|_{\theta=\theta_0}$. The inverse of (12), assuming its existence, is denoted by

$$I_{\text{MLa}}^{-1} = \begin{bmatrix} I_{\text{ML}}^* & \# \\ \# & \# \end{bmatrix}. \quad (13)$$

From (4), (10), (12) and (13),

$$\frac{\partial \hat{\theta}_{\text{ML}}}{\partial \mathbf{s}'} \Big|_{\hat{\theta}_{\text{ML}}=\theta_0, \mathbf{s}=\sigma_0} = \frac{1}{2} I_{\text{ML}}^* \frac{\partial \sigma'}{\partial \theta_0} D_p' (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) D_p. \quad (14)$$

Using (14), under the assumption of multivariate normality, it can be shown that

$$\text{acov}(\hat{\theta}_{\text{ML}}) = I_{\text{ML}}^* / n \quad (15)$$

(see e.g., Browne, 1974/1977, 1982; Silvey, 1975).

On the other hand, the discrepancy function for the weighted/unweighted LS estimation with the $p \times p$ weight matrix W is

$$f_{\text{WLS}} = \frac{1}{2} \text{tr}\{(W(\Sigma(\theta) - S))^2\}, \quad (16)$$

where $W = S^{-1}$, $(\text{diag} S)^{-1}$ and $I_{(p)}$ for the GLS, SLS and ULS estimators, respectively with $\text{diag } S$ being the diagonal matrix whose diagonal elements are those of S and $I_{(p)}$ a $p \times p$ identity matrix.

We will also consider the weighted LS estimation using $W = S^{-k}$ with $0 \leq k \leq 1$, e.g., $k = 0.5$, where S^{-k} is defined as a matrix with the same eigenvectors of S or S^{-1} and the eigenvalues being the k -th powers of those of S^{-1} . When $k=0, 1$, the corresponding estimators become the ULS and GLS estimators, respectively. So, the LS estimator with the weight matrix S^{-k} , ($0 < k < 1$) is a compromise between the GLS and ULS estimators. The estimation method is called the LS method with a powered covariance matrix (LSPC). It is known that the GLS method gives biased estimates in FA (see Jöreskog & Goldberger, 1972). The LSPC method is expected to reduce the biases without increasing the instability of estimates to a large extent.

It is to be noted that though the SLS method uses a diagonal weight matrix $(\text{diag} S)^{-1}$, it is different from the so-called diagonally weighted least squares method (DWLS, Jöreskog & Sörbom, 1996, pp.23-24). The latter uses the estimates of the diagonal elements of $\text{acov}(\mathbf{s})$, i.e., $(s_{ii}s_{jj} + s_{ij}^2)/n$, ($i, j = 1, \dots, p$) with the assumption of multivariate normality, which are different from the diagonal elements of S . Since the discrepancy function of the DWLS cannot be

described as (16) though the DWLS method as well as the SLS is scale free, the DWLS method will not be used in this paper

It is known that the ML method can be classified as a weighted least squares method using (16) when the weight matrix W is iteratively updated by the inverse of the estimated covariance matrix ($\hat{\Sigma}^{-1}$) of observed variables using the estimated parameters in an iterative computation (Lee & Jennrich, 1979). The asymptotic covariances between the estimators by possibly different estimation methods i and j using W_i and W_j , respectively in (16) are given by the following theorem.

Theorem 1. Let W_{i0} and W_{j0} be the population counterparts of W_i and W_j in (16) for two weighted LS methods i and j including the ML method; and $\hat{\theta}_{W_iLS}$ and $\hat{\theta}_{W_jLS}$ be the corresponding estimators of parameters with restrictions $\mathbf{h}(\hat{\theta}_{W_iLS}) = \mathbf{h}(\hat{\theta}_{W_jLS}) = \mathbf{0}$. Then, the asymptotic covariance matrix of $\hat{\theta}_{W_iLS}$ and $\hat{\theta}_{W_jLS}$ is

$$\begin{aligned} \text{acov}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) &= \frac{1}{4} I_{W_iLS}^* \frac{\partial \sigma'}{\partial \theta_0} D_p'(W_{i0} \otimes W_{i0}) D_p \\ &\quad \times \text{acov}(\mathbf{s}) D_p'(W_{j0} \otimes W_{j0}) D_p \frac{\partial \sigma}{\partial \theta_0'} I_{W_jLS}^*, \quad (i, j = 1, 2). \end{aligned} \quad (17)$$

Under the assumption of multivariate normality for observed variables, (17) reduces to

$$\text{acov}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) = \frac{1}{n} I_{W_iLS}^* I_{W_iS W_jLS} I_{W_jLS}^*, \quad (i, j = 1, 2). \quad (18)$$

Proof. The definition of $I_{W_iLS}^*$ similar to I_{ML}^* and the proof of the theorem are provided in Appendix 1. ■

From Theorem 1 the asymptotic correlation matrix is immediately given as

$$\begin{aligned} \text{acor}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) &= \{\text{diag}(\text{acov}(\hat{\theta}_{W_iLS}))\}^{-1/2} \\ &\quad \times \text{acov}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) \{\text{diag}(\text{acov}(\hat{\theta}_{W_jLS}))\}^{-1/2}. \end{aligned} \quad (19)$$

Corollary 1. When $\theta_{W_iLS} = \theta_{ML}$, (18) becomes

$$\text{acov}(\hat{\theta}_{ML}, \hat{\theta}_{W_jLS}) = \frac{1}{n} I_{ML}^*. \quad (20)$$

Proof. When θ has no restriction, it is easy to derive (20) because in this case $I_{W_iLS}^* = I_{ML}^{-1}$, $I_{W_iS W_jLS} = I_{W_jLS}$ (note that asymptotically $W_i = S^{-1}$ for the ML method) and $I_{W_jLS}^* = I_{W_jLS}^{-1}$. For the derivation under the condition with some restriction(s) on θ , see Appendix 2. ■

Equation (20) is equivalent to $\text{acov}(\hat{\theta}_{ML})$ with the assumption of multivariate normality. Because the asymptotic variances of $\hat{\theta}_{W_jLS}$ are not less than the corresponding ones of $\hat{\theta}_{ML}$ under the normality assumption, (20) shows that

$$\hat{\theta}_{W_jLS} = \hat{\theta}_{ML} + \mathbf{e}_j \quad (21)$$

where $\text{acov}(\hat{\theta}_{\text{ML}}, \mathbf{e}_j) = O$. That is, $\hat{\theta}_{W_j\text{LS}}$ is asymptotically not more accurate than $\hat{\theta}_{\text{ML}}$ in the sense that the random term $\mathbf{e}_j$ is added to $\hat{\theta}_{\text{ML}}$. The asymptotic correlations between $\mathbf{e}_j$ and $\mathbf{e}_k$ given by the weighted LS estimators with W_j and W_k , respectively can be calculated as the asymptotic partial correlations between $\hat{\theta}_{W_j\text{LS}}$ and $\hat{\theta}_{W_k\text{LS}}$ partialing out $\hat{\theta}_{\text{ML}}$. Though this requires only elementary calculation, the partial correlations are unstable when one or both of the estimators is highly correlated with the ML estimator. We conjecture that (21) holds in more general situations than the case of structural equation models (the derivation in the case of a single parameter is given by Cox & Hinkley, 1978, p.139, and Stuart, Ord, & Arnold, 1999, p.21). Note that while (20) holds, $\text{acor}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{W_j\text{LS}})$ is not necessarily equal to $\text{acor}(\hat{\theta}_{\text{ML}})$ or $\text{acor}(\hat{\theta}_{W_j\text{LS}})$ because generally $\text{acov}(\hat{\theta}_{\text{ML}})$ is not equal to $\text{acov}(\hat{\theta}_{W_j\text{LS}})$.

Corollary 2. Under the same conditions as in Corollary 1

$$\begin{aligned} \text{acor}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{W_j\text{LS}}) &= \text{acor}(\hat{\theta}_{\text{ML}}) \{ \text{diag}(\text{acov}(\hat{\theta}_{\text{ML}})) \}^{1/2} \\ &\quad \times \{ \text{diag}(\text{acov}(\hat{\theta}_{W_j\text{LS}})) \}^{-1/2}. \end{aligned} \quad (22)$$

Proof. Using (20) in (19), we have (22). ■

Corollary 2 shows that $\text{acor}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{W_j\text{LS}})$ is given by postmultiplying $\text{acor}(\hat{\theta}_{\text{ML}})$ by the ratios of the asymptotic standard errors of $\hat{\theta}_{\text{ML}}$ to those of $\hat{\theta}_{W_j\text{LS}}$. Since these ratios are not greater than one, the absolute values of the elements of $\text{acor}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{W_j\text{LS}})$ tend to be smaller than $\text{acor}(\hat{\theta}_{\text{ML}})$. From (22), it is easy to see that the asymptotic correlations of $\hat{\theta}_{\text{ML}}$ and the corresponding elements of $\hat{\theta}_{W_j\text{LS}}$ are equal to the ratios of the corresponding standard errors.

For the LS estimators with $W_i = S^{-k_i}$ and $W_j = S^{-k_j}$, we have the following result.

Corollary 3. For the LS estimators with $W_i = S^{-k_i}$ and $W_j = S^{-k_j}$, ($0 \leq k_i \leq 1$ and $0 \leq k_j \leq 1$) (18) becomes

$$\begin{aligned} \text{acov}(\hat{\theta}_{S^{-k_i}\text{LS}}, \hat{\theta}_{S^{-k_j}\text{LS}}) &= \frac{1}{n} I_{S^{-k_i}\text{LS}}^* I_{S^{1-k_i-k_j}\text{LS}} I_{S^{-k_j}\text{LS}}^*, \quad (23) \\ (i, j &= 1, 2). \end{aligned}$$

Proof. From $W_i S W_j = S^{-k_i} S S^{-k_j} = S^{1-k_i-k_j}$, (23) follows. ■

Recalling that the LSPC method with $W_i = S^{-k_i}$ includes the ML (asymptotically $k_i = 1$), GLS ($k_i = 1$) and ULS ($k_i = 0$) as special cases, (23) may be convenient for computation for the family of the LSPC method. When $k_i + k_j = 1$ as for the pair of the ULS and ML (GLS) or the pair of the LSPC methods with $k_i = k_j = 0.5$, (23) becomes

$$\frac{1}{n} I_{S^{-k_i}\text{LS}}^* I_{S^{0}\text{LS}} I_{S^{-k_j}\text{LS}}^* = \frac{1}{n} I_{S^{-k_i}\text{LS}}^* I_{\text{ULS}} I_{S^{-k_j}\text{LS}}^*, \quad (k_i + k_j = 1), \quad (24)$$

where

$$I_{\text{ULS}} = \frac{1}{2} \frac{\partial \sigma'}{\partial \theta_0} D_p' D_p \frac{\partial \sigma}{\partial \theta_0'}. \quad (25)$$

4. LS Estimator Using a Sample Correlation Matrix

As was mentioned in the introductory section, it is a common practice in the behavioral sciences to use sample correlations in place of sample covariances for fitting FA models. The discrepancy function of the unweighted LS estimation in such situation is

$$\begin{aligned} f_{\rho\text{LS}} &= \frac{1}{2} \text{tr}\{(\mathbf{P}(\theta_{\rho*}) - \mathbf{R} - \text{diag}(\mathbf{P}(\theta_{\rho*}) - \mathbf{R}))^2\} \\ &= (\boldsymbol{\rho} - \mathbf{r})'(\boldsymbol{\rho} - \mathbf{r}), \end{aligned} \quad (26)$$

where $\mathbf{P}(\theta_{\rho*})$ with $\text{diag } \mathbf{P}(\theta_{\rho*}) = \mathbf{I}_{(p)}$ and $\theta_{\rho*} = (\theta'_{\rho}, \theta'_{d\rho})'$ is a structured correlation matrix; θ_{ρ} with its population value $\theta_{\rho 0}$ is the $q^{\#} \times 1$ vector of parameters of interest in the correlation structure such as factor loadings in FA; $\theta_{d\rho}$ with its population value $\theta_{d\rho 0}$ is the $(q - q^{\#}) \times 1$ vector of parameters specific to the diagonal elements of $\mathbf{P}(\theta_{\rho*})$ such as the variances of unique factors; $\mathbf{R}$ is a sample correlation matrix; $\boldsymbol{\rho} = \rho(\theta_{\rho}) = \text{vb}(\mathbf{P}(\theta_{\rho*}))$ and $\mathbf{r} = \text{vb}(\mathbf{R})$, where $\text{vb}(\mathbf{R})$ is the $(p^2 - p)/2 \times 1$ vector consisting of the elements below the main diagonal of $\mathbf{R}$. Note that $\theta_{\rho*}$ is different from θ in $\Sigma(\theta) = (\text{diag } \Sigma)^{1/2} \mathbf{P}(\theta_{\rho*}) (\text{diag } \Sigma)^{1/2}$ since θ includes scale parameters as $\theta = (\theta'_{\rho*}, \mathbf{1}'_p (\text{diag } \Sigma)^{1/2})'$ with $\mathbf{1}_p$ being the $p \times 1$ vector of unities. The discrepancy function (26) does not include the parameters $\theta_{d\rho}$ and is equivalent to that of the minres method (Harman & Jones, 1966; see also Harman, 1976, Chapter 9) when $\mathbf{P}(\theta_{\rho*})$ has a form of the exploratory FA model.

Let $\hat{\theta}_{\rho\text{LS}}$ be the estimator of θ_{ρ} which minimizes (26) with restriction $\mathbf{h}(\hat{\theta}_{\rho\text{LS}}) = \mathbf{0}$, and

$$I_{\rho\text{LS}} = \frac{1}{2} \frac{\partial^2 f_{\rho\text{LS}}}{\partial \theta_{\rho} \partial \theta'_{\rho}} \Big|_{\theta_{\rho} = \theta_{\rho 0}, \mathbf{r} = \mathbf{r}_{\rho 0}}, \quad (27)$$

where ρ_0 is the population value of ρ with $I_{\rho\text{LS}}^*$ defined similarly to $I_{W_i\text{LS}}^*$. Then, from

$$\frac{\partial^2 f_{\rho\text{LS}}}{\partial \theta_{\rho} \partial \mathbf{r}'} \Big|_{\theta_{\rho} = \theta_{\rho 0}, \mathbf{r} = \mathbf{r}_{\rho 0}} = -2 \frac{\partial \rho'}{\partial \theta_{\rho 0}} \frac{\partial \mathbf{r}}{\partial \mathbf{r}'} = -2 \frac{\partial \rho'}{\partial \theta_{\rho 0}}, \quad (28)$$

where $\partial \rho' / \partial \theta_{\rho 0} = \partial \rho' / \partial \theta_{\rho} \Big|_{\theta_{\rho} = \theta_{\rho 0}}$, it follows that

$$\frac{\partial \hat{\theta}_{\rho\text{LS}}}{\partial \mathbf{r}'} \Big|_{\hat{\theta}_{\rho\text{LS}} = \theta_{\rho 0}, \mathbf{r} = \mathbf{r}_{\rho 0}} = I_{\rho\text{LS}}^* \frac{\partial \rho'}{\partial \theta_{\rho 0}}. \quad (29)$$

Using (29), we have

$$\text{acov}(\hat{\theta}_{\rho\text{LS}}) = I_{\rho\text{LS}}^* \frac{\partial \rho'}{\partial \theta_{\rho 0}} \text{acov}(\mathbf{r}) \frac{\partial \rho}{\partial \theta'_{\rho 0}} I_{\rho\text{LS}}^*. \quad (30)$$

Under the assumption of multivariate normality, the elements of $\text{acov}(\mathbf{r})$ in (30) become

$$\begin{aligned}
\text{acov}(r_{ij}, r_{kl}) = & \frac{1}{n} \left\{ \frac{1}{2} (\rho_{0ij} \rho_{0kl} (\rho_{0ik}^2 + \rho_{0il}^2 + \rho_{0jk}^2 + \rho_{0jl}^2) \right. \\
& + \rho_{0ik} \rho_{0jl} + \rho_{0il} \rho_{0jk} - \rho_{0ij} (\rho_{0jk} \rho_{0jl} + \rho_{0ik} \rho_{0il}) \\
& \left. - \rho_{0kl} (\rho_{0jk} \rho_{0ik} + \rho_{0jl} \rho_{0il}) \right\}, \\
& (p \geq i > j \geq 1; p \geq k > l \geq 1),
\end{aligned} \tag{31}$$

(Girshick, 1939, (3.23); Steiger & Hakstian, 1982, (4.2)), where r_{ij} and ρ_{0ij} are the (i, j) th elements of R and $P(\theta_{\rho 0})$, respectively. For nonnormal cases, see Steiger and Hakstian (1982).

We have the following theorem for the asymptotic covariance between $\hat{\theta}_{W_i \text{LS}} = (\hat{\theta}'_{\rho^* W_i \text{LS}}, \mathbf{1}_p' (\text{diag} \hat{\Sigma})^{1/2})'$ and $\hat{\theta}_{\rho \text{LS}}$ with $\hat{\theta}_{\rho^* W_i \text{LS}}$ being the weighted LS estimate of $\theta_{\rho^* 0} = (\theta'_{\rho 0}, \theta'_{d\rho 0})'$ with the model $\Sigma(\theta) = (\text{diag} \Sigma)^{1/2} P(\theta_{\rho^*}) (\text{diag} \Sigma)^{1/2}$.

Theorem 2. The asymptotic covariance matrix of $\hat{\theta}_{W_i \text{LS}}$ and $\hat{\theta}_{\rho \text{LS}}$ is

$$\begin{aligned}
\text{acov}(\hat{\theta}_{W_i \text{LS}}, \hat{\theta}_{\rho \text{LS}}) = & \frac{1}{2} I_{W_i \text{LS}}^* \frac{\partial \sigma'}{\partial \theta_0} D_p' (W_{i0} \otimes W_{i0}) D_p \\
& \times \text{acov}(\mathbf{s}) \frac{\partial \rho'}{\partial \sigma_0} \frac{\partial \rho}{\partial \theta'_{\rho 0}} I_{\rho \text{LS}}^*.
\end{aligned} \tag{32}$$

Under the assumption of multivariate normality (32) becomes

$$\begin{aligned}
\text{acov}(\hat{\theta}_{W_i \text{LS}}, \hat{\theta}_{\rho \text{LS}}) = & \frac{1}{n} I_{W_i \text{LS}}^* \frac{\partial \sigma'}{\partial \theta_0} D_p' (W_{i0} \Sigma_0 \otimes W_{i0} \Sigma_0) K_p \\
& \times \frac{\partial \rho'}{\partial \sigma_0} \frac{\partial \rho}{\partial \theta'_{\rho 0}} I_{\rho \text{LS}}^*.
\end{aligned} \tag{33}$$

Proof. Using (A5) and (29), we have

$$\begin{aligned}
\text{acov}(\hat{\theta}_{W_i \text{LS}}, \hat{\theta}_{\rho \text{LS}}) = & \frac{1}{2} I_{W_i \text{LS}}^* \frac{\partial \sigma'}{\partial \theta_0} D_p' (W_{i0} \otimes W_{i0}) D_p \\
& \times \text{acov}(\mathbf{s}, \mathbf{r}) \frac{\partial \rho}{\partial \theta'_{\rho 0}} I_{\rho \text{LS}}^*.
\end{aligned} \tag{34}$$

From $\text{acov}(\mathbf{s}, \mathbf{r}) = \text{acov}(\mathbf{s}) \frac{\partial \rho'}{\partial \sigma_0}$, (32) follows. For (33), under the normality assumption, using (A7) for $\text{acov}(\mathbf{s})$ and noting (A9), the result follows. ■

Corollary 4. When $\hat{\theta}_{W_i \text{LS}} = \hat{\theta}_{\text{ML}}$, (33) becomes

$$\text{acov}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{\rho \text{LS}}) = \frac{1}{n} I_{\text{ML}}^* \begin{bmatrix} I_{(q^\#)} \\ O \end{bmatrix}. \tag{35}$$

Proof. See Appendix 3. ■

In (32) and (34), the elements of $\text{acov}(\mathbf{s}, \mathbf{r}) = \text{acov}(\mathbf{s}) \partial \rho' / \partial \sigma_0$ are obtained by using the delta method:

$$\begin{aligned}
\text{acov}(s_{ij}, r_{kl}) &= \text{acov}(s_{ij}, (s_{kk}, s_{ll}, s_{kl})) \frac{\partial \rho_{kl}}{\partial (\sigma_{0kk}, \sigma_{0ll}, \sigma_{0kl})'} \\
&= \text{acov}(s_{ij}, (s_{kk}, s_{ll}, s_{kl})) \left(-\frac{1}{2\sigma_{0kk}}, -\frac{1}{2\sigma_{0ll}}, \frac{1}{\sigma_{0kl}} \right)' \rho_{0kl}, \quad (36) \\
&(p \geq i \geq j \geq 1; p \geq k > l \geq 1).
\end{aligned}$$

where ρ_{kl} is the (k, l) th element of $P(\theta_{\rho^*})$.

5. Numerical Examples with Equal Scales

In this section, two artificial numerical examples with exploratory FA models are presented for unstandardized or standardized observed variables. The theoretical values are given under the assumption of multivariate normality, while simulated observations are randomly generated by normally or nonnormally distributed latent variables with the same population covariance matrices given by the models employed. The nonnormal observations are generated using chi-square distributed variables with $df=10$ or 1 followed by the standardization as $(\chi^2 - df)/\sqrt{2df}$ to have mean zero and unit variance for common and unique factors. The cases with $df=10$ and 1 are used to represent moderate and strong nonnormality, respectively. Note that the variances of the factors are set at unities as is usually done in practice. However, this does not satisfy a necessary condition for the asymptotic robustness against nonnormality (see the references cited in the first paragraph of Section 1). That is, our nonnormal cases in this section represent somewhat “bad” conditions for normal theory estimation. The sample sizes are set at $N=300$ or 1200.

The first example is assumed to have the following one-factor model:

$$\Sigma = \lambda \lambda' + \Psi, \quad (37)$$

whose population values are

$$\begin{aligned}
\lambda_0 &= (.2 \ .3 \ .4 \ .5 \ .6 \ .7 \ .8 \ .9)' \text{ and} \\
\Psi_0 &= \text{diag}(.96 \ .91 \ .84 \ .75 \ .64 \ .51 \ .36 \ .19).
\end{aligned}$$

That is, the eight observed variables have the same unit population variances. The estimates are obtained by minimizing f_{ML} , f_{GLS} , f_{LSPC} with $k = 0.5$, f_{SLS} , f_{ULS} or f_{PLS} . When observed variables are standardized, (37) becomes somewhat different, i.e.,

$$\begin{aligned}
\Sigma &= C(\lambda_{\rho} \lambda'_{\rho} + \Psi_{\rho})C, \\
\text{diag}(\lambda_{\rho} \lambda'_{\rho} + \Psi_{\rho}) &= I_{(p)}, \quad (38)
\end{aligned}$$

where $C = \{\text{diag} \Sigma\}^{1/2}$ is the diagonal matrix whose diagonal elements are the standard deviations of the observed variables (their population values are unities in this case). For standardized variables, only f_{ML} and f_{PLS} are minimized

with the restriction $\text{diag}(\lambda_\rho \lambda'_\rho + \Psi_\rho) = I_{(p)}$. Actually, estimation is performed by regarding R as S which gives proper parameter estimates with required restrictions automatically satisfied. That is, for ML estimation $\ln |\lambda_\rho \lambda'_\rho + \Psi_\rho| + \text{tr}\{(\lambda_\rho \lambda'_\rho + \Psi_\rho)^{-1} R\}$ is minimized while for ULS estimation $\text{tr}\{(\lambda_\rho \lambda'_\rho + \Psi_\rho - R)^2\}$ is minimized. The estimates given by this ULS method for standardized variables are equivalent to those of SLS estimation using a sample covariance matrix with the model and restrictions of (38). Tables 1-4 give the results for the first example.

Table 1: Efficiencies and biases for the one-factor model of unstandardized normally/nonnormally distributed variables

Unstandardized variables		ML T (S)	GLS (S)	LSPC($k = .5$) T (S)	SLS T (S)	ULS (S)
$N=300$						
Efficiency	Normal	1(1)	(1.003)	.990(.995)	.970(.971)	(.971)
Nonnormal	$df=10$	1(1)	(1.014)	(1.001)	(.980)	(.0979)
	$df=1$	1(1)	(1.053)	(1.017)	(.993)	(.993)
ARMB	Normal	(.0030)	(.0213)	(.0115)	(.0033)	(.0033)
Nonnormal	$df=10$	(.0023)	(.0202)	(.0105)	(.0023)	(.0023)
	$df=1$	(.0033)	(.0196)	(.0100)	(.0033)	(.0033)
$N=1,200$						
Efficiency	Normal	1(1)	(1.001)	.990(.990)	.970 (.970)	(.970)
Nonnormal	$df=10$	1(1)	(1.006)	(.997)	(.980)	(.980)
	$df=1$	1(1)	(1.012)	(1.000)	(.991)	(.991)
ARMB	Normal	(.0012)	(.0053)	(.0029)	(.0012)	(.0012)
Nonnormal	$df=10$	(.0009)	(.0049)	(.0026)	(.0009)	(.0009)
	$df=1$	(.0014)	(.0049)	(.0026)	(.0014)	(.0014)

Note. Efficiency = $\text{tr}\{\text{acov}(\hat{\theta}_{\text{ML}})\}/\text{tr}\{\text{acov}(\hat{\theta}_{(i)})\}$, ($i \in (\text{ML}, \dots, \text{ULS})$).

ARMB=overall root mean squared biases, T(S)=theoretical (simulated) values.

The summary efficiency index of the estimators is shown in Table 1. The definition of the efficiency index for the i -th method is

$$\text{Efficiency} = \text{tr}\{\text{acov}(\hat{\theta}_{\text{ML}})\}/\text{tr}\{\text{acov}(\hat{\theta}_{(i)})\}. \quad (39)$$

In $\hat{\theta}_{(i)}$ of (39), the nuisance parameters (the standard deviations) of observed variables are temporarily excluded in the case of standardized variables because they are of little interest.

The simulated values of the efficiency index are shown in parentheses in the tables. These were empirically obtained by the following method. First, random observations with $N=300$ or 1200 following normal/nonnormal distributions were generated. From these observations, the parameter estimates were obtained by various estimation methods. The sign of the factor loading vector was changed when the sum of the loading estimates was negative. This was replicated 1,000

times, which gave empirical covariance matrices for the sets of parameter estimates. The simulated values of the efficiency index were given by these empirical covariance matrices in place of the asymptotic ones. While the theoretical efficiency index is free from sample sizes, the simulated ones are slightly varied from $N=300$ to $N=1,200$. Table 1 also shows the index of overall root mean squared bias (ARMB), which was computed as $\sqrt{(\hat{\theta}_{(i)} - \theta_0)'(\hat{\theta}_{(i)} - \theta_0)/q^*}$ where $\hat{\theta}_{(i)}$ is a mean vector of the parameter estimates of interest over 1,000 replications given by the i -th estimation method and q^* is the number of the parameters of interest.

Table 2: Efficiencies and biases for the one-factor model of standardized normally/nonnormally distributed variables

Standardized variables		ML T (S)	ULS T (S)
$N=300$			
Efficiency	Normal	1(1)	.946(.947)
Nonnormal	$df=10$	1(1)	(.961)
	$df=1$	1(1)	(.975)
ARMB	Normal	(.0018)	(.0021)
Nonnormal	$df=10$	(.0018)	(.0019)
	$df=1$	(.0035)	(.0039)
$N=1,200$			
Efficiency	Normal	1(1)	.946(.941)
Nonnormal	$df=10$	1(1)	(.957)
	$df=1$	1(1)	(.972)
ARMB	Normal	(.00088)	(.00086)
Nonnormal	$df=10$	(.00067)	(.00075)
	$df=1$	(.00079)	(.00091)

Note. Efficiency = $\text{tr}\{\text{acov}(\hat{\theta}_{\text{ML}})\}/\text{tr}\{\text{acov}(\hat{\theta}_i)\}$, ($i \in (\text{ML}, \text{ULS})$).

Note that the nuisance parameters (the SDs of observed variables) are not included in $\hat{\theta}$. ARMB=overall root mean squared biases. T(S)=theoretical (simulated) values.

From Table 1, we see that the efficiencies of the SLS and ULS estimators are slightly but consistently smaller than those by the ML method both in theoretical and simulated values (note that the theoretical values are common to the SLS and ULS estimators in this case). The simulated efficiencies seem to increase as nonnormality becomes strong. This is expected because in such situation, the ML estimation does not necessarily show asymptotic efficiency. The GLS estimator shows simulated efficiencies slightly more than one especially in nonnormal cases (note that theoretical efficiencies are one for the GLS estimator). However, the GLS estimator gives large biases which were reduced by the LSPC method to about halves though they are still large. It is to be noticed that the SLS and ULS

methods give biases as small as those by the ML method in all situations in the table. The substantial influence of the difference of sample sizes is shown only for the absolute levels of biases. Table 2 shows the results for standardized variables. In the table the tendencies similar to those in Table 1 are observed.

Table 3: Theoretical and simulated correlations for the one-factor model for unstandardized/standardized variables with $N=300$

Unstandardized variables									
Variable		$\hat{\lambda}_{ML}$				$\hat{\lambda}_{ULS}$			
No.		2	4	6	8	2	4	6	8
$\hat{\lambda}_{ML}$	2	1.00	.06	.14	.21	.98	.07	.14	.21
	4	.07	1.00	.13	.25	.07	.98	.13	.25
	6	.11	.19	1.00	.36	.13	.13	.98	.35
	8	.15	.26	.38	1.00	.20	.25	.35	.99
$\hat{\lambda}_{ULS}$	2	.98	.07	.11	.15	1.00	.07	.14	.19
	4	.07	.98	.19	.25	.07	1.00	.13	.24
	6	.11	.19	.98	.38	.11	.19	1.00	.34
	8	.15	.26	.38	.99	.14	.24	.37	1.00
Standardized variables									
Variable		$\hat{\lambda}_{\rho ML}$				$\hat{\lambda}_{\rho LS}$			
No.		2	4	6	8	2	4	6	8
$\hat{\lambda}_{\rho ML}$	2	1.00	.05	.09	.03	.98	.05	.09	.04
	4	.07	1.00	.09	.09	.05	.98	.09	.10
	6	.09	.14	1.00	.02	.09	.09	.95	.02
	8	.04	.06	.02	1.00	.03	.10	.03	.97
$\hat{\lambda}_{\rho LS}$	2	.98	.06	.08	.04	1.00	.06	.09	.01
	4	.06	.98	.14	.06	.07	1.00	.09	.08
	6	.08	.14	.96	.02	.08	.13	1.00	.01
	8	.04	.06	.02	.97	.02	.03	.00	1.00

Note. The lower and upper triangular halves of the matrices indicate theoretical and simulated correlations, respectively.

Table 3 shows the theoretical asymptotic correlations between the ML and ULS estimators for factor loadings (the lower triangular halves of the two matrices in the table), and the corresponding empirical correlations (the upper triangular halves) in the case with $N=300$ given by the simulation mentioned above. They are shown for every other observed variables to save space. The theoretical values are close to the empirical ones, which indicates accuracy of our formulas. Overall, in the case of $\hat{\lambda}$ the correlations between not corresponding parameter estimates are small for standardized variables while somewhat correlated for unstandardized variables. On the other hand, the correlations between the corresponding

parameter estimates are .95-1.00, which is consistent with the high values of the summary efficiency index in Tables 1 and 2.

Table 4: Asymptotic/simulated correlations between parameter estimates for the one-factor model of unstandardized/standardized normally/nonnormally distributed variables

		[ML, GLS] (Cor)	[ML,LSPC ($k = .5$)] Acor(Cor)	[ML, SLS] Acor(Cor)	[ML, ULS] Acor(Cor)	[GLS, ULS] (Cor)	[SLS, ULS] Acor(Cor)
Unstandardized variables		$N = 300$					
$\hat{\lambda}_4$	Normal	(1.00)	.99(.99)	.98(.98)	.98(.98)	(.98)	1.00(1.00)
Nonnormal	$df=10$	(1.00)	(.99)	(.99)	(.99)	(.98)	(1.00)
	$df=1$	(1.00)	(.99)	(.99)	(.99)	(.99)	(1.00)
$\hat{\psi}_{44}$	Normal	(.96)	1.00(.99)	.99(.99)	.99(.99)	(.95)	1.00(1.00)
Nonnormal	$df=10$	(.98)	(.99)	(.99)	(.99)	(.97)	(1.00)
	$df=1$	(.99)	(1.00)	(1.00)	(1.00)	(.99)	(1.00)
		$N = 1,200$					
$\hat{\lambda}_4$	Normal	(1.00)	.99(.99)	.98(.98)	.98(.98)	(.98)	1.00(1.00)
Nonnormal	$df=10$	(1.00)	(.99)	(.99)	(.99)	(.98)	(1.00)
	$df=1$	(1.00)	(1.00)	(.99)	(.99)	(.99)	(1.00)
$\hat{\psi}_{44}$	Normal	(.99)	1.00(.99)	.99(.99)	.99(.99)	(.98)	1.00(1.00)
Nonnormal	$df=10$	(.99)	(1.00)	(.99)	(.99)	(.99)	(1.00)
	$df=1$	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
Standardized variables		$N = 300$		[ML,ULS]	[ML,ULS]		
Nonnormal	Normal	$\hat{\lambda}_4$		.98(.98)	$\hat{\psi}_{44}$	.98(.98)	
	$df=10$			(.98)		(98)	
	$df=1$			(.99)		(99)	
		$N = 1,200$					
Nonnormal	Normal	$\hat{\lambda}_4$		.98(.98)	$\hat{\psi}_{44}$	.98(.98)	
	$df=10$			(.98)		(98)	
	$df=1$			(.99)		(99)	

Note. Acor= asymptotic correlations, Cor=simulated correlations.

Table 4 shows the asymptotic correlations for $\hat{\lambda}_4$ and those for $\hat{\psi}_{44}$. The results of the remaining parameters are similar and omitted. The values of the correlations are very large in all cases. However, we find the tendency that the values increase as nonnormality becomes strong. This is in line with the increases of the efficiencies of the estimators other than the ML estimator observed in the previous tables.

The second example is based on the following two-factor model: $\Sigma = \Lambda \Lambda' + \Psi$ with their population values being

$$\Lambda_0 = \begin{bmatrix} .67 & .64 & .74 & .33 & .21 & .30 \\ .20 & .30 & .33 & .62 & .69 & .72 \end{bmatrix}',$$

$$\Psi_0 = \text{diag}(.51 \ .50 \ .35 \ .50 \ .48 \ .39), \quad (40)$$

Table 5: Efficiencies, biases and correlations for the rotated model of standardized normally/nonnormally distributed variables

Standardized variables		ML T(S)	ULS T(S)	$\hat{\lambda}_{31}$ [ML,ULS] Acor(Cor)	$\hat{\psi}_{33}$ [ML,ULS] Acor(Cor)
<i>N</i> = 300					
Efficiency	Normal	1(1)	.998(1.039)	1.00(.99)	1.00(.99)
Nonnormal	<i>df</i> =10	1(1)	(1.036)	(1.00)	(1.00)
	<i>df</i> =1	1(1)	(1.026)	(1.00)	(1.00)
ARMB	Normal	(.0035)	(.0035)		
Nonnormal	<i>df</i> =10	(.0043)	(.0042)		
	<i>df</i> =1	(.0032)	(.0032)		
<i>N</i> = 1,200					
Efficiency	Normal	1(1)	.998(1.003)	1.00(1.00)	1.00(1.00)
Nonnormal	<i>df</i> =10	1(1)	(1.002)	(1.00)	(1.00)
	<i>df</i> =1	1(1)	(1.003)	(1.00)	(1.00)
ARMB	Normal	(.0010)	(.0010)		
Nonnormal	<i>df</i> =10	(.0013)	(.0014)		
	<i>df</i> =1	(.0010)	(.0010)		

Note. Efficiency = $\text{tr}\{\text{acov}(\hat{\theta}_{\text{ML}})\}/\text{tr}\{\text{acov}(\hat{\theta}_{(i)})\}$, ($i \in (\text{ML}, \text{ULS})$).

Note that the nuisance parameters (the SDs of observed variables) are not included in $\hat{\theta}$. ARMB=overall root mean squared biases, T(S)=theoretical (simulated) values, Acor=asymptotic correlations, Cor=simulated correlations.

which gives unit population variances for the six observed variables. The loading matrix is an orthogonally rotated one (shown up to two decimal places) by the normalized varimax method. When standardized variables are used, the model is modified as was done in the first example though the population parameter values are unchanged since $\text{diag } \Sigma_0 = I_{(6)}$. Because orthogonal rotation is employed, Λ has a restriction $h(\text{vec } \Lambda) = 0$ even in the case of unstandardized variables (for the restriction in the normalized varimax solution, see e.g., Ogasawara, 2000, and Yuan & Bentler, 2000). The results for standardized variables are shown in Table 5.

The simulated values in Table 5 are given by a method similar to that in the first example. Estimation in each replication of the simulation was performed in two steps. First, the just identified loading matrix with a fixed zero loading in the same location as in the corresponding population matrix $\tilde{\Lambda}_0 = \begin{bmatrix} .7 & .7 & .8 & .5 & .4 & .5 \\ 0 & .1 & .1 & .5 & .6 & .6 \end{bmatrix}'$ is estimated, which is followed by the normalized varimax rotation. The rotated loading matrix is transformed, if necessary, by changing the signs or permutation of the columns of $\hat{\Lambda}$ to be as close as possible to the population loading matrix in order to remove the minor indeterminacy after

rotation.

The theoretical efficiencies by ULS estimation are almost the same as , but very slightly smaller than, those by ML estimation. On the other hand, the simulated efficiencies give opposite results. These results occur because the population standard errors given by ML and LS estimation are very close to each other. The asymptotic correlations between the corresponding estimates for two typical parameters ($\hat{\lambda}_{31}$ and $\hat{\psi}_{33}$) by the ML and ULS methods are .99-1.00 both in theoretical and simulated values, which is in line with the high values of the efficiency index. The magnitude of the biases of ULS estimation is also on the same level as that of ML estimation.

6. Numerical Examples with Different Scales

In the previous section, the population variances of observed variables were the same. In this section, different variances are assumed. Four similar but different one-factor models are assumed for unstandardized variables:

$$\Sigma_i = \lambda_i \lambda_i' + \Psi_i, \quad (i \in (A, B, C, D)),$$

(41)

whose population values are

$$\begin{aligned} \lambda_{i0} &= (c_{i01}, c_{i02}, c_{i03}, c_{i04}, c_{i05}, c_{i06})' \times .6, \\ \Psi_{i0} &= \text{diag}(c_{i01}^2, c_{i02}^2, c_{i03}^2, c_{i04}^2, c_{i05}^2, c_{i06}^2) \times .64, \end{aligned}$$

and the values of $c_{i01}, \dots, c_{i06}$ for Cases A-D are shown in Table 6. Note that the parameters are the six elements of λ_i and the six diagonal elements of Ψ_i in each model. From (41), we see that the relative communalities are the same among the six variables in each model. Case A shows equal variances and Cases B, C and D different variances with varied degrees of magnitude.

Table 6: Constants for the one-factor models of unstandardized variables

Cases	A	B	C	D
Variables	c_{A0j}	c_{B0j}	c_{C0j}	c_{D0j}
1	1	1.0	1	1
2	1	1.0	2	4
3	1	1.2	3	9
4	1	1.2	4	16
5	1	1.4	5	25
6	1	1.4	6	36

The results are shown in Tables 7 and 8. Considering the previous observation that the GLS and LSPC methods showed relatively large biases, the main results (Table 7 and the first and second panels of Table 8) are given only by the remaining

methods (ML, SLS and ULS) for the case with $N=300$. In Table 7, it is shown that the SLS method gives the unit asymptotic efficiencies and slightly higher simulated efficiencies than one in the normal or weakly nonnormal cases with the biases similar to those by the ML method. On the other hand, the ULS method shows substantially smaller efficiencies especially when the scales are very different (Cases C and D). The magnitude of the biases of the ULS method for Cases C and D are also greater than those by the ML method. The first and second panels of Table 8 show the asymptotic and simulated correlations of the estimators for two parameters ($\hat{\lambda}_3$ and $\hat{\psi}_{33}$) for three pairs of estimation methods: [ML, SLS], [ML, ULS] and [SLS, ULS]. The results are clear: the correlations for [ML, SLS] are perfect or almost perfect in all conditions in the table for Cases A-D. On the other hand, the correlations for [ML, ULS] and [SLS, ULS] are substantially smaller than one especially in the normal or weakly nonnormal cases. In the case of the strongly nonnormal case ($df = 1$), we find that the simulated correlations for the pairs of [ML, ULS] and [SLS, ULS] have increased especially for Cases C and D. This is due to the fact that ML estimation does not have efficiency when the distributional assumption is not satisfied. In such situation, the estimates with different minimizing functions may become relatively similar.

In the third panel of Table 8, the asymptotic/simulated correlation matrix for six estimators of $\hat{\lambda}_3$ for Case D under the condition of normal distribution is shown, where LSPD indicate the WLS estimator with powered diagonals $W = (\text{diag}S)^{-0.5}$, which was introduced by Ogasawara (2001) in a different situation with purpose similar to that of LSPC. The matrix shows a so-called simplex structure, which indicates an order of the estimators: 1(ML, GLS, SLS), 2(LSPD), 3(LSPC), 4(ULS). The order is strongly associated with the asymptotic efficiencies of the estimators. We find that LSPD is closer to ML than LSPC in this case.

It is to be noted that the model of (41) does not satisfy the condition of asymptotic robustness. However, the FA model can be transformed into an equivalent model satisfying asymptotic robustness by reparametrization such as

$$\Sigma_i = \tilde{\lambda}_i \phi_i \tilde{\lambda}_i' + \Psi_i, \quad (i \in (A, B, C, D)), \quad (42)$$

where the last element of $\tilde{\lambda}_i$ is fixed at the corresponding population value with ϕ_i being the unconstrained variance of the common factor. Table 9 shows the results by the reparametrization, where the results for unique variances are not shown since they are the same as before. As expected we find that the simulated correlations for $\hat{\lambda}_3$ when $df = 10$ or 1 are almost the same as those of the case of normal distribution. However, the results of $\hat{\phi}$ are different depending on the strength of nonnormality as for $\hat{\psi}_{33}$ in Table 8. It is known that $\hat{\phi}$ as well as $\hat{\psi}_{jj}$ has not asymptotic robustness even when the condition of asymptotic robustness for loading estimates is satisfied. The correct values of the inflated asymptotic correlations for $\hat{\phi}$ or $\hat{\psi}_{jj}$ in nonnormal data can in principle be obtained by using Theorem 1 when $\text{acov}(s)$ is given.

Table 7: Efficiencies and biases for the one-factor model of unstandardized variables with different scales ($N=300$)

Unstandardized variables	Cases	A T (S)	B T (S)	C T (S)	D T (S)
ML					
ARMB	Normal	(.0030)	(.0034)	(.026)	(.61)
Nonnormal	$df=10$	(.0024)	(.0036)	(.041)	(1.61)
	$df=1$	(.0067)	(.0098)	(.012)	(3.87)
SLS					
Efficiency	Normal	1(1.006)	1(1.006)	1(1.008)	1(1.009)
Nonnormal	$df=10$	(1.007)	(1.007)	(1.008)	(1.008)
	$df=1$	(1.002)	(1.003)	(1.003)	(1.003)
ARMB	Normal	(.0029)	(.0034)	(.027)	(.63)
Nonnormal	$df=10$	(.0024)	(.0036)	(.041)	(1.28)
	$df=1$	(.0067)	(.0099)	(.13)	(3.96)
ULS					
Efficiency	Normal	1(1.007)	.980(.989)	.822(.837)	.715(.730)
Nonnormal	$df=10$	(1.006)	(.992)	(.864)	(.959)
	$df=1$	(1.000)	(.996)	(.963)	(.931)
ARMB	Normal	(.0030)	(.0034)	(.037)	(1.05)
Nonnormal	$df=10$	(.0024)	(.0038)	(.058)	(1.98)
	$df=1$	(.0067)	(.0098)	(.14)	(4.68)

Note. $\text{tr}\{\text{acov}(\hat{\theta}_{\text{ML}})\}/\text{tr}\{\text{acov}(\hat{\theta}_{(i)})\}$, ($i \in (\text{ML}, \text{SLS}, \text{ULS})$).

Note that the nuisance parameters (the SDs of observed variables) are not included in $\hat{\theta}$.

ARMB=overall root mean squared biases, T(S)=theoretical (simulated) values.

Finally, for illustration of the asymptotic correlations between $\mathbf{e}_j$ and $\mathbf{e}_k$ (see (21)) given by the j -th and k -th methods, respectively, the asymptotic partial correlation between $\hat{\lambda}_3$'s by the ULS and LSPC ($W = S^{-0.5}$) methods for Case D under the same condition as in Table 8 with the assumption of multivariate normality is calculated. Since $\text{acor}[\text{ULS}, \text{LSPC}]=0.99$, $\text{acor}[\text{ULS}, \text{ML}]=0.85$ and $\text{acor}[\text{LSPC}, \text{ML}]=0.90$, the partial correlation becomes 0.97 using more accurate values. It is interesting to find that the specific components in $\hat{\lambda}_3$'s by the ULS and LSPC estimators have still high correlation after partialing out the common component of the ML estimator.

Table 8: Correlations for the one-factor model of unstandardized variables with different scales ($N=300$)

Unstandardized variables		Cases	A	B	C	D
			Acor(Cor)	Acor(Cor)	Acor(Cor)	Acor(Cor)
$\hat{\lambda}_3$	[ML,SLS]	Normal	1(1.00)	1(1.00)	1(1.00)	1(1.00)
		Nonnormal	$df=10$	(1.00)	(1.00)	(1.00)
			$df=1$	(1.00)	(1.00)	(1.00)
	[ML,ULS]	Normal	1(1.00)	.99(.99)	.92(.92)	.85(.85)
		Nonnormal	$df=10$	(.99)	(.92)	(.85)
			$df=1$	(.99)	(.95)	(.90)
	[SLS,ULS]	Normal	1(1.00)	.99(.99)	.92(.92)	.85(.86)
		Nonnormal	$df=10$	(.99)	(.92)	(.85)
			$df=1$	(.99)	(.95)	(.91)
	$\hat{\psi}_{33}$ [ML,SLS]	Normal	1(1.00)	1(1.00)	1(1.00)	1(1.00)
		Nonnormal	$df=10$	(1.00)	(1.00)	(1.00)
			$df=1$	(1.00)	(1.00)	(1.00)
	[ML,ULS]	Normal	1(1.00)	.99(.99)	.90(.89)	.82(.80)
		Nonnormal	$df=10$	(.99)	(.91)	(.84)
			$df=1$	(1.00)	(.98)	(.95)
	[SLS,ULS]	Normal	1(1.00)	.99(.99)	.90(.89)	.82(.81)
		Nonnormal	$df=10$	(.99)	(.92)	(.84)
			$df=1$	(1.00)	(.98)	(.95)
ML		1(1)	Acor(Cor) for $\hat{\lambda}_3$ of Case D			
GLS		1(1.00)	1(1)	with normal distribution		
SLS		1(1.00)	1(1.00)	1(1)		
LSPD		.92(.92)	.92(.92)	.92(.92)	1(1)	
LSPC($k = .5$)		.90(.91)	.90(.91)	.90(.91)	1.00(1.00)	1(1)
ULS		.85(.85)	.85(.85)	.85(.86)	.98(.98)	.99(.99) 1(1)

Note. Acor(Cor)= asymptotic (simulated) correlations,
LSPD=WLS with $W = (\text{diag}S)^{-0.5}$.

7. Discussion

From the results in the sections of examples, we see that when the variances of observed variables are equal, the weighted/unweighted LS estimators give efficiencies very similar to that of the ML estimators and high correlations with the ML estimators. The stability of the ULS method decreases as the difference of the variances increases. This stems from the well known property of the scale dependence of the ULS estimators. It is not recommended to use ULS estimation for models for unstandardized variables when the variances of observed variables are substantially different.

The one-factor models in Cases A-D in the example section with different variances can be reasonably modified to those for standardized variables as:

Table 9: Correlations for the one-factor model of unstandardized variables with different scales when asymptotic robustness is satisfied ($N=300$)

Unstandardized variables	Cases	A Acor(Cor)	B Acor(Cor)	C Acor(Cor)	D Acor(Cor)
$\tilde{\lambda}_3$					
[ML,SLS]	Normal	1(1.00)	1(1.00)	1(1.00)	1(1.00)
Nonnormal	$df=10$	(1.00)	(1.00)	(1.00)	(1.00)
	$df=1$	(1.00)	(1.00)	(1.00)	(1.00)
[ML,ULS]	Normal	1(1.00)	.99(.98)	.88(.88)	.78(.77)
Nonnormal	$df=10$	(1.00)	(.99)	(.88)	(.76)
	$df=1$	(1.00)	(.98)	(.87)	(.76)
[SLS,ULS]	Normal	1(1.00)	.99(.99)	.88(.88)	.78(.77)
Nonnormal	$df=10$	(1.00)	(.99)	(.88)	(.77)
	$df=1$	(1.00)	(.98)	(.87)	(.77)
$\tilde{\phi}$					
[ML,SLS]	Normal	1(1.00)	1(1.00)	1(1.00)	1(1.00)
Nonnormal	$df=10$	(1.00)	(1.00)	(1.00)	(1.00)
	$df=1$	(1.00)	(1.00)	(1.00)	(1.00)
[ML,ULS]	Normal	1(1.00)	.99(.99)	.93(.93)	.88(.86)
Nonnormal	$df=10$	(1.00)	(.99)	(.94)	(.88)
	$df=1$	(1.00)	(1.00)	(.96)	(.93)
[SLS,ULS]	Normal	1(1.00)	.99(.99)	.93(.93)	.88(.86)
Nonnormal	$df=10$	(1.00)	(.99)	(.94)	(.89)
	$df=1$	(1.00)	(1.00)	(.97)	(.94)

Note. Acor= asymptotic correlations, Cor=simulated correlations.

$$\Sigma_i = C_i(\lambda_\rho \lambda'_\rho + \Psi_\rho) C_i, \quad C_i = \text{diag}(c_{i1}, \dots, c_{i6}),$$

$$(i \in (A, B, C, D)), \quad (43)$$

with $\text{diag}(\lambda_\rho \lambda'_\rho + \Psi_\rho) = I_{(6)}$ or equivalently $\lambda_\rho \lambda'_\rho + \Psi_\rho = \lambda_\rho \lambda'_\rho + \text{diag}(I_{(6)} - \lambda_\rho \lambda'_\rho)$. When such formulation with $f_{\rho\text{LS}}$ is employed and the same R is given, Cases A-D give the same parameter estimates, the same standard errors and the same asymptotic correlations with the ML estimators for $\lambda_{\rho 0}$ and $\Psi_{\rho 0}$. This is because the model for standardized variables is the same ($\lambda_\rho \lambda'_\rho + \Psi_\rho$) and the input data R is common. This equivalence shows an advantage of estimation using $f_{\rho\text{LS}}$ or f_{SLS} .

In this study, only the exploratory FA models have been dealt with. However, the formulas for the asymptotic correlations can be used for general structural equation models such as the LISREL model. It is to be expected that the main conclusions in the numerical examples will be conveyed to such general cases

depending upon their similarity to our cases of FA.

REFERENCES

- Anderson, T. W. (1987). Multivariate linear relations. In T. Pukkila & S. Puntanen (Eds.), *Proceedings of the Second International Tampere Conference in Statistics* (pp. 9-36). Tampere, Finland: Department of Mathematical Sciences, University of Tampere.
- Anderson, T. W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.
- Anderson, T. W. & Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759-771.
- Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman, (Ed.) *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V (pp. 111-150). Berkeley: University of California Press.
- Arbuckle, J. L. & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SPSS Inc.
- Bentler, P. M. (1989). *EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.
- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1-24. Reprinted in D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models* (pp. 205-226). Amsterdam: North Holland, 1977.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.) *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.
- Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika*, **74**, 375-384.
- 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.
- Cox, D. R. & Hinkley, D. V. (1978). *Problems and solutions in theoretical statistics*. London: Chapman & Hall.
- Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.
- Geršhick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, **10**, 203-224.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.
- Harman, H. H. & Jones, W. H. (1966). Factor analysis by minimizing residuals (min-res). *Psychometrika*, **31**, 351-368.
- Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8: User's reference guide* (2nd ed.). Chicago : Scientific Software International.

- Jöreskog, K. G. & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, **37**, 243-260.
- Jöreskog, K. G. & Sörbom, D. du Toit & du Toit (1999). *LISREL 8: New statistical features*. Chicago: Scientific Software International.
- Krane, W., R. & McDonald, R. P. (1978). Scale invariance and the factor analysis of correlation matrices. *British Journal of Mathematical and Statistical Psychology*, **31**, 218-228.
- Lee, S. Y. & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal*, **14**, 121-136.
- Lee, S. Y. & Jennrich, R. I. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. *Psychometrika*, **44**, 99-113.
- Magnus, J. R. & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.
- Mooijart, A. & Bentler, P. M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica*, **45**, 159-171.
- Ogasawara, H. (2000). On the standard errors of rotated factor loadings with weights for observed variables. *Behaviormetrika*, **27**, 1-14.
- Ogasawara, H. (2001). Least squares estimation of item response theory linking coefficients. *Applied Psychological Measurement*, **25**, 373-383.
- Satorra, A. (1990). Robustness issues in structural equation modeling: A review of recent developments. *Quality & Quantity*, **24**, 367-386.
- Satorra, A. & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis*, **10**, 235-249.
- Silvey, S. D. (1975). *Statistical inference*. London: Chapman & Hall.
- Steiger, J. H. & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, **35**, 208-215.
- Stuart, A., Ord, J. K. & Arnold, S. (1999). *Kendall's advanced theory of statistics: Classical inference and the linear model* (6th ed., Vol.2A). London: Arnold.
- Yuan, K.-H. & Bentler, P. (2000). On equivalence and invariance of standard errors in three exploratory factor models. *Psychometrika*, **65**, 121-133.

Received May 2001. Final version accepted April 2002.

Appendix 1. Proof of Theorem 1

Noting that

$$\begin{aligned}
 g_{W_i \text{LS } k} &\equiv \frac{\partial f_{W_i \text{LS}}}{\partial \theta_k} \Big|_{\theta = \hat{\theta}_{W_i \text{LS}}} \\
 &= \text{tr} \{ W_i (\Sigma(\hat{\theta}_{W_i \text{LS}}) - S) W_i \frac{\partial \Sigma(\theta)}{\partial \theta_k} \Big|_{\theta = \hat{\theta}_{W_i \text{LS}}} \}, \quad (\text{A1}) \\
 (i = 1, 2 ; k = 1, \dots, q),
 \end{aligned}$$

where θ_k is the k -th element of θ , we have

$$\begin{aligned} \frac{\partial^2 f_{W_iLS}}{\partial \theta \partial \mathbf{s}'} \Big|_{\theta=\theta_0, \mathbf{s}=\sigma_0} &= -\frac{\partial \text{vec}' \Sigma}{\partial \theta_0} (W_{i0} \otimes W_{i0}) \frac{\partial \text{vec } S}{\partial \mathbf{s}'} \\ &= -\frac{\partial \sigma'}{\partial \theta_0} D'_p (W_{i0} \otimes W_{i0}) D_p. \end{aligned} \quad (\text{A2})$$

Let

$$I_{W_iLSa} = \frac{1}{2} \begin{bmatrix} H_{W_iLS}(\theta_0) & \partial \mathbf{h}'(\theta)/\partial \theta_0 \\ \partial \mathbf{h}(\theta)/\partial \theta_0' & O \end{bmatrix} \quad (\text{A3})$$

and

$$I_{W_iLSa}^{-1} = \begin{bmatrix} I_{W_iLS}^* & \# \\ \# & \# \end{bmatrix}, \quad (i = 1, 2), \quad (\text{A4})$$

where $H_{W_iLS}(\theta_0)$ is defined by (11) with f_i being replaced by that for the LS method, and the notation I_{W_iLSa} similar to the augmented information matrix I_{MLa} is used for ease of comparison though I_{W_iLSa} is not the augmented information matrix. Then, from (4), (A2), (A3) and (A4),

$$\frac{\partial \hat{\theta}_{W_iLS}}{\partial \mathbf{s}'} \Big|_{\hat{\theta}_{W_iLS}=\theta_0, \mathbf{s}=\sigma_0} = \frac{1}{2} I_{W_iLS}^* \frac{\partial \sigma'}{\partial \theta_0} D'_p (W_{i0} \otimes W_{i0}) D_p, \quad (i = 1, 2), \quad (\text{A5})$$

which gives

$$\begin{aligned} \text{acov}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) &= \frac{1}{4} I_{W_iLS}^* \frac{\partial \sigma'}{\partial \theta_0} D'_p (W_{i0} \otimes W_{i0}) D_p \\ &\quad \times \text{acov}(\mathbf{s}) D'_p (W_{j0} \otimes W_{j0}) D_p \frac{\partial \sigma}{\partial \theta_0'} I_{W_jLS}^*, \quad (i, j = 1, 2). \end{aligned} \quad (\text{A6})$$

Under the assumption of multivariate normality, since

$$\text{acov}(\mathbf{s}) = 2 K_p' (\Sigma_0 \otimes \Sigma_0) K_p / n \quad (\text{A7})$$

(see e.g., Browne, 1974/1977), where K_p is the matrix satisfying $\sigma = K_p' \text{vec } \Sigma$ or $D_p = K_p (K_p' K_p)^{-1}$, (A6) becomes

$$\begin{aligned} \text{acov}(\hat{\theta}_{W_iLS}, \hat{\theta}_{W_jLS}) &= \frac{1}{4n} I_{W_iLS}^* \frac{\partial \sigma'}{\partial \theta_0} D'_p (W_{i0} \otimes W_{i0}) D_p \\ &\quad \times (2 K_p' (\Sigma_0 \otimes \Sigma_0) K_p) D'_p (W_{j0} \otimes W_{j0}) D_p \frac{\partial \sigma}{\partial \theta_0'} I_{W_jLS}^* \\ &= \frac{1}{2n} I_{W_iLS}^* \frac{\partial \sigma'}{\partial \theta_0} D'_p \{ (W_{i0} \Sigma_0 W_{j0}) \otimes (W_{i0} \Sigma_0 W_{j0}) \} \\ &\quad \times D_p \frac{\partial \sigma}{\partial \theta_0'} I_{W_jLS}^* \\ &= \frac{1}{n} I_{W_iLS}^* I_{W_iS} W_{jLS} I_{W_jLS}^*, \end{aligned} \quad (\text{A8})$$

where

$$(W \otimes W) D_p K'_p = D_p K'_p (W \otimes W) \quad (\text{A9})$$

with a symmetric W is used (see Browne, 1974/1977; Fuller, 1987, pp.384-385; Magnus & Neudecker, 1988, pp.49-50).

Appendix 2. Proof of Corollary 1 under the condition with some restriction(s) on θ

Rewrite $1/\sqrt{2}$ times $\partial \mathbf{h}'(\theta)/\partial \theta_0$ in (A3) as L . Then, noting $I_{W_j \text{LS}} = \frac{1}{2} H_{W_j \text{LS}}(\theta_0)$, $I_{W_j \text{LS}}^*$ in (18) can be explicitly written as (Lee & Bentler, 1980, Lemma 3):

$$\begin{aligned} I_{W_j \text{LS}}^* &= (I_{W_j \text{LS}} + L L')^{-1} - (I_{W_j \text{LS}} + L L')^{-1} \\ &\quad \times L (L' (I_{W_j \text{LS}} + L L')^{-1} L)^{-1} L' (I_{W_j \text{LS}} + L L')^{-1}. \end{aligned} \quad (\text{A10})$$

From (A10), it follows that for any W_j including that for ML estimation

$$L' I_{W_j \text{LS}}^* = O. \quad (\text{A11})$$

Using $I_{W_i S W_j \text{LS}} = I_{W_j \text{LS}}$, (A10) and (A11), we have

$$\begin{aligned} \text{acov}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{W_j \text{LS}}) &= \frac{1}{n} I_{\text{ML}}^* I_{W_j \text{LS}} I_{W_j \text{LS}}^* \\ &= \frac{1}{n} I_{\text{ML}}^* (I_{W_j \text{LS}} + L L') I_{W_j \text{LS}}^* \\ &= \frac{1}{n} I_{\text{ML}}^* \{I_{(q)} - L (L' (I_{W_j \text{LS}} + L L')^{-1} L)^{-1} L' \\ &\quad \times (I_{W_j \text{LS}} + L L')^{-1}\} \\ &= \frac{1}{n} I_{\text{ML}}^*. \end{aligned} \quad (\text{A12})$$

Appendix 3. Proof of Corollary 4.

For ML estimation, asymptotically $W_{i0} = \Sigma_0^{-1}$. From this and noting that $D'_p K_p$ is an identity matrix, (33) becomes

$$\begin{aligned} \text{acov}(\hat{\theta}_{\text{ML}}, \hat{\theta}_{\rho \text{LS}}) &= \frac{1}{n} I_{\text{ML}}^* \frac{\partial \sigma'}{\partial \theta_0} \frac{\partial \rho'}{\partial \sigma_0} \frac{\partial \rho}{\partial \theta'_{\rho 0}} I_{\rho \text{LS}}^* \\ &= \frac{1}{n} I_{\text{ML}}^* \begin{bmatrix} I_{(q^\#)} \\ O \end{bmatrix} \left(\frac{\partial \rho'}{\partial \theta_{\rho 0}} \frac{\partial \rho}{\partial \theta'_{\rho 0}} + L_\rho L'_\rho \right) I_{\rho \text{LS}}^* \\ &= \frac{1}{n} I_{\text{ML}}^* \begin{bmatrix} I_{(q^\#)} \\ O \end{bmatrix}, \end{aligned} \quad (\text{A13})$$

where $L_\rho = (1/\sqrt{2}) \partial \mathbf{h}'(\theta_\rho)/\partial \theta_{\rho 0}$. The second and third equations of (A13) follow in a manner similar to (A12) with (A10) and (A11).