Ann Inst Stat Math (2009) 61:995–1017
DOI 10.1007/s10463-008-0174-4

Asymptotic expansions in the singular value decomposition for cross covariance and correlation under nonnormality

Haruhiko Ogasawara

Received: 4 April 2007 / Revised: 22 November 2007 / Published online: 25 April 2008

Abstract Asymptotic cumulants of the distributions of the sample singular vectors and values of cross covariance and correlation matrices are obtained under nonnormality. The asymptotic cumulants are used to have the approximations of the distributions of the estimators by the Edgeworth expansions up to order $O(1/n)$ and Hall's method with variable transformation. The cases of Studentized estimators are also considered. As an application of the method, the distributions of the parameter estimators in the model of inter-battery factor analysis are expanded. Interpreting the singular vectors and values in the context of the factor model with distributional conditions, the asymptotic robustness of some lower-order normal-theory cumulants of the distributions of the sample singular vectors and values under nonnormality is shown.

Keywords Singular value decomposition · Edgeworth expansion · Studentized estimators · Asymptotic robustness · Nonnormality · Inter-battery factor analysis

1 Introduction

The singular value decomposition (SVD) is a basic tool in multivariate data analysis to extract dominant information from the associations of sets of variables. A typical example using the SVD is found in canonical correlation analysis. Let $\mathbf{x}$ and $\mathbf{y}$ be $p \times 1$ and $q \times 1$ vector variables, respectively, with $p \leq q$ without loss of generality. Let $\mathbf{v}^* = (\mathbf{x}', \mathbf{y}')'$ and

$$\mathrm{Cov}(\mathbf{v}^*) = \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix}. \tag{1}$$

H. Ogasawara (✉)
Department of Information and Management Science,
Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan
e-mail: hogasa@res.otaru-uc.ac.jp

Then, it is known that the population canonical correlation coefficients $\phi_1^* \geq \cdots \geq \phi_p^* \geq 0$ are the singular values of $\boldsymbol{\Sigma}_{XX}^{-1/2}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_{YY}^{-1/2}$, where $\boldsymbol{\Sigma}_{XX}^{-1/2}$ and $\boldsymbol{\Sigma}_{YY}^{-1/2}$ are the symmetric square roots of $\boldsymbol{\Sigma}_{XX}^{-1}$ and $\boldsymbol{\Sigma}_{YY}^{-1}$, respectively, with the assumption of their existence. The estimators $\hat{\phi}_1^*, \ldots, \hat{\phi}_p^*$ are given from the singular values of $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$ with

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix}, \tag{2}$$

which is a usual $(p+q) \times (p+q)$ unbiased sample covariance matrix based on $N = n + 1$ independent observations.

The exact and approximate distributions of $\hat{\phi}_1^*(\widehat{\phi_1^{*2}})$ and associated estimators have been extensively investigated. In earlier stages, the distributions were given under normality (Hotelling 1936; Hsu 1941; Anderson 1958; Lawley 1959; Constantine 1963; Sugiura 1976; Fujikoshi 1978; Konishi 1981; Anderson 1999). Subsequently, the corresponding results have been given under arbitrary or nonnormal distributions (Muirhead and Waternaux 1980; Fang and Krishnaiah 1982; Steiger and Browne 1984; Hayakawa 1987; Eaton and Tyler 1994; Boik 1998; Ogasawara 2007b). The asymptotic distribution of the likelihood ratio statistic for uncorrelatedness between sets of variables has also been derived under normality (Bartlett 1938; Fujikoshi 1977) and under nonnormality (Muirhead and Waternaux 1980; Bai and He 2004; Ogasawara 2007b).

It is known that the canonical correlation coefficients are scale-free in that they are invariant with respect to the multiplication of a possibly different nonzero constant to each element of $\mathbf{v}^*$. Consequently, the same canonical correlation coefficients are obtained when (1) and (2) are replaced by the population and sample correlation matrices:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{XX} & \mathbf{P}_{XY} \\ \mathbf{P}_{YX} & \mathbf{P}_{YY} \end{bmatrix}, \ \mathbf{R} = \begin{bmatrix} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{bmatrix}, \tag{3}$$

respectively. In this paper, the SVDs of $\boldsymbol{\Sigma}_{XY}$ and $\mathbf{P}_{XY}$ with their sample counterparts $\mathbf{S}_{XY}$ and $\mathbf{R}_{XY}$ are dealt with. Note that the scale-freeness of the singular values in canonical correlation analysis does not hold in this case. The SVDs of the asymmetric matrices are closely related to the model of inter-battery factor analysis first given by Tucker (1958). The inter-battery factor model is also functionally related to the covariance structure model given by canonical correlation analysis (Rao 1973, Sect. 8f.3; 1979; Browne 1979; Wegelin et al. 2002; see also Wegelin et al. 2001, 2006). A similar problem for asymmetric matrices is the SVDs of say, $N \times p$ data matrices with or without centering for columns, which are often used in bioinformatics (e.g., Liu et al. 2003; Hu and He 2007). However, the problem is not dealt with in this paper.

In the following sections the asymptotic expansions of the distributions of the estimators in the SVDs of $\boldsymbol{\Sigma}_{XY}$ and $\mathbf{P}_{XY}$, and the associated parameter estimators in the inter-battery factor model will be derived under nonnormality. Some of the results are given in the references addressed earlier. Eaton and Tyler (1994) derived the limiting (normal) distributions of the sample singular values of asymmetric matrices under nonnormality. Boik (1998) provided general models and the method to have the asymptotic variances and biases of the estimators in the SVDs of asymmetric matrices with adaptation to some of the estimators in inter-battery factor analysis under nonnormality. In this paper, these results will be extended by using the two-term Edgeworth expansion up to order $O(n^{-1})$ and associated methods.

It will be shown that the normal-theory (NT) asymptotic variances and biases of the elements of the unit-norm sample singular vectors for $\mathbf{S}_{XY}$ are robust against the violation of the normality assumption under some distributional conditions with the inter-batter factor model. The similar robustness of the NT asymptotic biases of the sample singular values for $\mathbf{S}_{XY}$ and the estimators of the inter-battery and battery-specific factor variances and covariances will also be shown.

2 The asymptotic expansions using the least squares discrepancy function

In this section, the SVDs of $\mathbf{\Sigma}_{XY}$ and $\mathbf{S}_{XY}$ are dealt with. The corresponding results for $\mathbf{P}_{XY}$ and $\mathbf{R}_{XY}$ will be given later. The SVD of an asymmetric matrix is usually obtained by the spectral resolution of the matrix post or premultiplied by its transpose. The equivalent result is given by using the following least squares (LS) discrepancy function in the case of $\mathbf{S}_{XY}$:

$$F_{\mathrm{LS}} = \frac{1}{2}\mathrm{tr}\{(\mathbf{A\Phi B}' - \mathbf{S}_{XY})(\mathbf{A\Phi B}' - \mathbf{S}_{XY})'\}, \quad \mathbf{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_p) \tag{4}$$

with the restrictions

$$\mathbf{A}'\mathbf{A} = \mathbf{B}'\mathbf{B} = \mathbf{I}_p, \tag{5}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix. The matrices $\mathbf{A}(p \times p)$, $\mathbf{\Phi}(p \times p)$ and $\mathbf{B}(q \times p)$ are mathematical variables in (4). They are also used as population values for simplicity of notation. The population singular values ϕ_i $(i = 1, \ldots, p)$ are assumed to be $\phi_1 > \phi_2 > \cdots > \phi_p > 0$. The cases of multiple roots will be treated later in the discussion section. The estimators of the singular vectors (the columns of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$) and the singular values $(\hat{\phi}_i)$ are given, in form, by minimizing (4) with respect to $\mathbf{A}$, $\mathbf{\Phi}$ and $\mathbf{B}$ under the restrictions (5) with the minor indeterminacy of the sign in each pair of the singular vectors of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$. Under the condition that $\mathbf{v}^*$ has a continuous distribution, we have $\hat{\phi}_1 > \hat{\phi}_2 > \cdots > \hat{\phi}_p > 0$ with probability 1. It is possible to use other discrepancy functions, e.g., the scaled negative log-likelihood including $\mathbf{S}_{XX}$ and $\mathbf{S}_{YY}$ with unstructured $\mathbf{\Sigma}_{XX}$ and $\mathbf{\Sigma}_{YY}$. This gives the same results since $\mathbf{A\Phi B}'$ is a saturated model for the asymmetric matrix. Arguably, the discrepancy function of (4) is the simplest one.

Let $\boldsymbol{\theta}$ be the $Q \times 1$ vector of parameters with $Q = p^2 + pq + p$ as

$$\boldsymbol{\theta} = (\mathrm{vec}'\mathbf{A}, \mathrm{vec}'\mathbf{B}, \mathbf{1}_p'\mathbf{\Phi})', \tag{6}$$

where vec $\cdot$ is the vectorizing operator stacking the columns of an argument matrix with $\text{vec}'\cdot = \{\text{vec}(\cdot)\}'$ and $\mathbf{1}_p$ is the $p \times 1$ vector of 1's. The restrictions (5) are described by using the following $(p^2 + p) \times 1$ vector set equal to $\mathbf{0}$:

$$\mathbf{h}(\boldsymbol{\theta}) = (\text{v}'(\mathbf{A}'\mathbf{A} - \mathbf{I}_p), \text{v}'(\mathbf{B}'\mathbf{B} - \mathbf{I}_p))' = \mathbf{0}, \tag{7}$$

where $\text{v}(\cdot)$ is the vectorizing operator taking the nonduplicated elements of a symmetric matrix with $\text{v}'(\cdot) = \{\text{v}(\cdot)\}'$. Since the restrictions in (5) or (7) can be seen as those for model identification, the first-order conditions for the vector of the estimators $\hat{\boldsymbol{\theta}}$ are given by

$$\left(\left.\frac{\partial F_{\text{LS}}}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \hat{\mathbf{h}}' \right)' = \mathbf{0} \quad \text{with } \hat{\mathbf{h}} = \mathbf{h}(\hat{\boldsymbol{\theta}}). \tag{8}$$

The equations in (8) represent implicit functions $\hat{\boldsymbol{\theta}} (= \boldsymbol{\theta}(\mathbf{s}))$ in terms of $\mathbf{s} = \text{vec}\,\mathbf{S}_{XY}$. Let θ denote an element of $\boldsymbol{\theta}$. Assume that the following Taylor series expansion of $\hat{\theta}$ about its true value θ_0 holds with the existence of the moments of the associated variables up to a required order:

$$\begin{aligned} \hat{\theta} = \theta_0 &+ \left.\frac{\partial \theta}{\partial \mathbf{s}'}\right|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s} - \boldsymbol{\sigma}) + \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{s}'} \right)^{<2>} \theta \Big|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s} - \boldsymbol{\sigma})^{<2>} \\ &+ \frac{1}{6} \left(\frac{\partial}{\partial \mathbf{s}'} \right)^{<3>} \theta \Big|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s} - \boldsymbol{\sigma})^{<3>} + o_p(n^{-3/2}), \end{aligned} \tag{9}$$

where $\boldsymbol{\sigma} = \text{vec}\,\boldsymbol{\Sigma}_{XY}$; $\mathbf{s}$ is also used as a mathematical vector variable; $\mathbf{X}^{<k>} = \mathbf{X} \otimes \cdots \otimes \mathbf{X}$ (k times); $\otimes$ denotes the Kronecker product. Let $w = n^{1/2}(\hat{\theta} - \theta_0)$. It is assumed that the asymptotic cumulants of w up to the fourth order can be written as follows:

$$\begin{aligned} \kappa_1(w) &= \text{E}(w) = n^{-1/2}\alpha_1 + o(n^{-1/2}), \\ \kappa_2(w) &= \text{E}[\{w - \text{E}(w)\}^2] = \alpha_2 + n^{-1}\Delta\alpha_2 + o(n^{-1}), \\ \kappa_3(w) &= \text{E}[\{w - \text{E}(w)\}^3] = n^{-1/2}\alpha_3 + o(n^{-1/2}), \\ \kappa_4(w) &= \text{E}[\{w - \text{E}(w)\}^4] - 3\{\kappa_2(w)\}^2 = n^{-1}\alpha_4 + o(n^{-1}). \end{aligned} \tag{10}$$

Then, from (9) it is known that

$$\alpha_1 = \frac{1}{2}\text{tr}\left(\frac{\partial^2 \theta}{\partial \boldsymbol{\sigma}\, \partial \boldsymbol{\sigma}'} \boldsymbol{\Omega} \right), \quad \alpha_2 = \frac{\partial \theta}{\partial \boldsymbol{\sigma}'} \boldsymbol{\Omega} \frac{\partial \theta}{\partial \boldsymbol{\sigma}}, \tag{11}$$

where $n^{-1}\boldsymbol{\Omega}$ is the asymptotic covariance matrix of $\mathbf{s}$, i.e., $n\,\text{acov}(\mathbf{s}) = \boldsymbol{\Omega}$; $\text{acov}(\cdot)$ denotes the asymptotic covariance matrix of order $O(n^{-1})$ for the argument vector; $\partial\theta/\partial\boldsymbol{\sigma} = \partial\theta/\partial\mathbf{s}|_{\mathbf{s}=\boldsymbol{\sigma}}$ with the similar expressions for partial derivatives for simplicity of notation. Ogasawara (2006) gave expressions for $\Delta\alpha_2$, α_3 and α_4 under the condition that eighth-order moments of the observable variables exist. These expressions require partial derivatives of θ with respect to $\boldsymbol{\sigma}$ up to the third order. However, the results are somewhat involved and are not repeated here.

Using the asymptotic cumulants in (11) and employing Cramér's condition for the distribution of the associated variables (see, e.g., Hall 1992a, Theorem 2.2), the Edgeworth expansion of the distribution function of standardized w or $\hat{\theta}$ is given as follows:

$$\begin{aligned}\Pr\left(\frac{w}{\alpha_2^{1/2}} \leq z\right) &= \boldsymbol{\Phi}(z) - n^{-1/2}\left\{\frac{\alpha_1}{\alpha_2^{1/2}} + \frac{\alpha_3}{6\alpha_2^{3/2}}(z^2-1)\right\} \\ &\quad \times\phi(z) - n^{-1}\left\{\frac{1}{2}(\Delta\alpha_2 + \alpha_1^2)\frac{z}{\alpha_2} + \left(\frac{\alpha_4}{24} + \frac{\alpha_1\alpha_3}{6}\right)\right. \\ &\quad \left.\times\frac{z^3-3z}{\alpha_2^2} + \frac{\alpha_3^2(z^5-10z^3+15z)}{72\alpha_2^3}\right\}\phi(z) + o(n^{-1}), \quad (12)\end{aligned}$$

where $\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$ and $\boldsymbol{\Phi}(z) = \int_{-\infty}^{z}\phi(t)\mathrm{d}t$.

It is known that the approximations to the distribution functions given by Edgeworth expansions are not necessarily non-decreasing in finite samples. These anomalous phenomena can be avoided by using Hall's (1992b) method removing asymptotic skewness with monotone transformation.

The asymptotic expansions shown above were derived by using population asymptotic cumulants, which are not given in practice. On the other hand, we have the Studentized estimator

$$t = \frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\alpha}_2^{1/2}} = \frac{w}{\hat{\alpha}_2^{1/2}}, \quad (13)$$

whose asymptotic cumulants are

$$\begin{aligned}&\kappa_1(t) = n^{-1/2}\alpha_1' + o(n^{-1/2}), \quad \kappa_2(t) = 1 + o(1), \\ &\kappa_3(t) = n^{-1/2}\alpha_3' + o(n^{-1/2}),\end{aligned} \quad (14)$$

where the expressions for α_1' and α_3' were given by Ogasawara (2005a, 2007a).

Let $z_{\tilde{\alpha}} = \boldsymbol{\Phi}^{-1}(1-\tilde{\alpha})$ (e.g., $\tilde{\alpha} = 0.05$). Then, the confidence interval for θ with the asymptotic confidence coefficient $1-\tilde{\alpha}$ accurate up to order $O(n^{-1/2})$ by the usual Cornish–Fisher expansion is

$$\hat{\theta} + \left[\pm z_{\tilde{\alpha}/2} - n^{-1/2}\left\{\hat{\alpha}_1' + (\hat{\alpha}_3'/6)(z_{\tilde{\alpha}/2}^2 - 1)\right\}\right] n^{-1/2}\hat{\alpha}_2^{1/2}. \quad (15)$$

The corresponding confidence interval given by Hall's method is

$$\hat{\theta} - n^{-1}\hat{\alpha}_2^{1/2}\hat{\alpha}_1' + 6\hat{\alpha}_2^{1/2}(\hat{\alpha}_3')^{-1}\Bigg[\ \Big\{1 - (1/2)\hat{\alpha}_3'(\pm n^{-1/2}z_{\tilde{\alpha}/2} - (n^{-1}/6)\hat{\alpha}_3')\Big\}^{1/3} - 1\Bigg]. \tag{16}$$

In practice, $\hat{\alpha}_1'$ and $\hat{\alpha}_3'$ tend to be unstable since the quantities $\hat{\alpha}_1'$ and $\hat{\alpha}_3'$ involve sample moments up to the sixth order under nonnormality. On the other hand, the NT estimators or those under normality, denoted by $\hat{\alpha}_{\mathrm{NT1}}'$ and $\hat{\alpha}_{\mathrm{NT3}}'$, are relatively stable. A typical situation encountered in practice is to use the NT Studentized estimator under nonnormality. The asymptotic cumulants of the NT Studentized estimator under such a condition are known (Ogasawara 2005a, 2007a). They are denoted by α_{NT1}'', $\alpha_{\mathrm{NT2}}''(\neq 1)$ and α_{NT3}''.

3 The partial derivatives

As addressed earlier, the formulas of the asymptotic cumulants of w in (11) involve the partial derivatives of the parameter estimators with respect to $\mathbf{s}$ up to the third order. They are derived from the partial derivatives in implicit functions given by (8). Differentiating (8) with respect to $\mathbf{s}$ using the chain rule when necessary, we have

$$\begin{bmatrix} \dfrac{\partial^2 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} & \dfrac{\partial\,\mathbf{h}'}{\partial\boldsymbol{\theta}} \\ \dfrac{\partial\,\mathbf{h}}{\partial\boldsymbol{\theta}'} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \dfrac{\partial\boldsymbol{\theta}}{\partial\,\mathbf{s}'} \\ \mathbf{O} \end{bmatrix} + \begin{bmatrix} \dfrac{\partial^2 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\,\partial\,\mathbf{s}'} \\ \mathbf{O} \end{bmatrix} = \mathbf{O}, \tag{17}$$

which gives the first partial derivatives of $\boldsymbol{\theta}$ with respect to $\mathbf{s}$ as

$$\begin{bmatrix} \dfrac{\partial\boldsymbol{\theta}}{\partial\,\mathbf{s}'} \\ \mathbf{O} \end{bmatrix} = -\mathbf{J}_{\mathrm{LS}}^{-1} \begin{bmatrix} \dfrac{\partial^2 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\,\partial\,\mathbf{s}'} \\ \mathbf{O} \end{bmatrix} \quad \text{with } \mathbf{J}_{\mathrm{LS}} = \begin{bmatrix} \dfrac{\partial^2 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} & \dfrac{\partial\,\mathbf{h}'}{\partial\boldsymbol{\theta}} \\ \dfrac{\partial\,\mathbf{h}}{\partial\boldsymbol{\theta}'} & \mathbf{O} \end{bmatrix}. \tag{18}$$

In (17) and (18), the actual expressions of the partial derivatives of F_{LS} evaluated at $\mathbf{s} = \boldsymbol{\sigma}$ with the parameter vector being at its population value (from now on stated simply as "evaluated at the population values") are

$$\frac{\partial^2 F_{\mathrm{LS}}}{(\partial\boldsymbol{\theta}')^{<2>}} = \mathrm{vec}'\,\mathbf{I}_{pq}\left(\frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\theta}'}\right)^{<2>}, \quad \frac{\partial^2 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}'\otimes\partial\,\mathbf{s}'} = -\mathrm{vec}'\,\frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\theta}'}, \tag{19}$$

where

$$\frac{\partial\boldsymbol{\sigma}}{\partial\boldsymbol{\theta}'} = [(\mathbf{B\Phi})\otimes\mathbf{I}_p, \{\mathbf{I}_q\otimes(\mathbf{A\Phi})\}\mathbf{K}_{qp}, (\mathbf{B}\otimes\mathbf{A})\mathbf{M}_p], \tag{20}$$

$\mathbf{K}_{qp}$ is the commutation matrix with $\mathbf{K}_{qp}\mathrm{vec}\,\mathbf{B} = \mathrm{vec}\,\mathbf{B}'$ (see Magnus and Neudecker 1999, p. 47), and $\mathbf{M}_p$ is the $p^2 \times p$ matrix with $\mathbf{M}_p\boldsymbol{\Phi}\mathbf{1}_p = \mathrm{vec}\,\boldsymbol{\Phi}$. For the nonzero partial derivatives of $\mathbf{h}$,

$$\begin{aligned}\frac{\partial\, \mathrm{v}(\mathbf{A}'\mathbf{A} - \mathbf{I}_p)}{\partial\, \mathrm{vec}'\mathbf{A}} &= \mathbf{L}_p\{\mathbf{I}_p \otimes \mathbf{A}' + (\mathbf{A}' \otimes \mathbf{I}_p)\mathbf{K}_{pp}\},\\ \frac{\partial\, \mathrm{v}(\mathbf{B}'\mathbf{B} - \mathbf{I}_p)}{\partial\, \mathrm{vec}'\mathbf{B}} &= \mathbf{L}_p\{\mathbf{I}_p \otimes \mathbf{B}' + (\mathbf{B}' \otimes \mathbf{I}_p)\mathbf{K}_{qp}\},\end{aligned} \tag{21}$$

where $\mathbf{L}_p$ is the $(p^2 + p)/2 \times p^2$ elimination matrix with $\mathrm{v}(\mathbf{A}'\mathbf{A}) = \mathbf{L}_p\mathrm{vec}(\mathbf{A}'\mathbf{A})$.

The second partial derivatives of $\boldsymbol{\theta}$ with respect to $\mathbf{s}$ are given by differentiating (17) with respect to $\mathbf{s}'$ and solving the equation:

$$\begin{bmatrix} \dfrac{\partial^2\boldsymbol{\theta}}{(\partial\,\mathbf{s}')^2} \\ \mathbf{O} \end{bmatrix} = -\mathbf{J}_{\mathrm{LS}}^{-1} \begin{bmatrix} \dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}(\partial\,\boldsymbol{\theta}')^{<2>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right)^{<2>} + \dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'@}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'} \otimes \dfrac{@}{\partial\,\mathbf{s}'}\right) \\ + \dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'@}\left(\dfrac{@}{\partial\,\mathbf{s}'} \otimes \dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right) + \dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}(\partial\,\mathbf{s}')^{<2>}} \\ \dfrac{\partial^2\mathbf{h}}{(\partial\,\boldsymbol{\theta}')^{<2>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right)^{<2>} \end{bmatrix}, \tag{22}$$

where @ denotes the correspondence of variables in differentiation, that is, the column for s_{ab} and s_{cd} ($a, c = p+1, \ldots, p+q;\ b, d = 1, \ldots, p$) in $\dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'@}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'} \otimes \dfrac{@}{\partial\,\mathbf{s}'}\right)$ is $\dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\partial s_{cd}}\dfrac{\partial\boldsymbol{\theta}}{\partial s_{ab}}$ with $\dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'@}\left(\dfrac{@}{\partial\mathbf{s}'} \otimes \dfrac{\partial\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right) = \dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'@}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'} \otimes \dfrac{@}{\partial\,\mathbf{s}'}\right)\mathbf{K}_{pq,pq}$. In (22), the nonzero partial derivatives evaluated at the population values are

$$\begin{aligned}\frac{\partial^3 F_{\mathrm{LS}}}{(\partial\,\boldsymbol{\theta}')^{<3>}} &= \sum^{3} \mathrm{vec}'\mathbf{I}_{pq}\left(\frac{\partial\,\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'} \otimes \frac{\partial^2\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<2>}}\right),\\ \frac{\partial^3 F_{\mathrm{LS}}}{(\partial\,\boldsymbol{\theta}')^{<2>} \otimes \partial\,\mathbf{s}'} &= -\mathrm{vec}'\frac{\partial^2\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<2>}}\end{aligned} \tag{23}$$

where Σ^3 denotes the sum of three similar terms considering the combinations of the vector variables, i.e.,

$$\mathrm{vec}'\mathbf{I}_{pq}\left(\frac{\partial\,\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'} \otimes \frac{\partial^2\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<2>}} + \frac{\partial^2\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<2>}} \otimes \frac{\partial\,\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'} + \frac{\partial^2\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'@} \otimes \frac{\partial\,\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'} \otimes \frac{@}{\partial\,\boldsymbol{\theta}'}\right)$$

and the nonzero elements of $\partial^2\boldsymbol{\sigma}/(\partial\,\boldsymbol{\theta}')^{<2>}$ are

$$\frac{\partial^2\boldsymbol{\sigma}}{\partial a_{ij}\partial b_{kj}} = \phi_j \mathrm{vec}\mathbf{E}_{ik}, \quad \frac{\partial^2\boldsymbol{\sigma}}{\partial a_{ij}\partial \phi_j} = \mathrm{vec}(\mathbf{E}_{ij}\mathbf{B}') = (\mathbf{B}\otimes\mathbf{I}_p)\,\mathrm{vec}\mathbf{E}_{ij},$$
$$\frac{\partial^2\boldsymbol{\sigma}}{\partial b_{ij}\partial \phi_j} = \mathrm{vec}(\mathbf{A}\mathbf{E}_{ji}) = (\mathbf{I}_q\otimes\mathbf{A})\,\mathrm{vec}\mathbf{E}_{ji}, \tag{24}$$

where a_{ij} and b_{kj} $(i, j = 1, \ldots, p; k = 1, \ldots, q)$ are the elements of $\mathbf{A}$ and $\mathbf{B}$, respectively, and $\mathbf{E}_{ik}$ is the matrix of an appropriate size whose (i, k)th element is 1 with the remaining ones being 0. For the nonzero partial derivatives of $\mathbf{h}$,

$$\frac{\partial^2 \mathrm{v}(\mathbf{A}'\mathbf{A}-\mathbf{I}_p)}{\partial\,\mathrm{vec}'\mathbf{A}\,\partial\,a_{ij}} = \mathbf{L}_p\{\mathbf{I}_p\otimes\mathbf{E}_{ji} + (\mathbf{E}_{ji}\otimes\mathbf{I}_p)\mathbf{K}_{pp}\},$$
$$\frac{\partial^2 \mathrm{v}(\mathbf{B}'\mathbf{B}-\mathbf{I}_p)}{\partial\,\mathrm{vec}'\mathbf{B}\,\partial\,b_{ij}} = \mathbf{L}_p\{\mathbf{I}_p\otimes\mathbf{E}_{ji} + (\mathbf{E}_{ji}\otimes\mathbf{I}_p)\mathbf{K}_{qp}\}. \tag{25}$$

Lastly, the third partial derivatives of $\boldsymbol{\theta}$ with respect to $\mathbf{s}'$ are given as follows:

$$\begin{bmatrix} \dfrac{\partial^3\boldsymbol{\theta}}{(\partial\,\mathbf{s}')^3} \\ \mathbf{O} \end{bmatrix} = -\mathbf{J}_{\mathrm{LS}}^{-1}\left[\begin{array}{l} \dfrac{\partial^4 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}(\partial\,\boldsymbol{\theta}')^{<3>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right)^{<3>} + \displaystyle\sum^3\left\{\dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,(\partial\,\boldsymbol{\theta}')^{<2>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\otimes\dfrac{\partial^2\boldsymbol{\theta}}{(\partial\,\mathbf{s}')^2}\right)\right. \\ +\dfrac{\partial^4 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,(\partial\,\boldsymbol{\theta}')^{<2>}\,@}\left(\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right)^{<2>}\otimes\dfrac{@}{\partial\,\mathbf{s}'}\right) \\ +\dfrac{\partial^3 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'\,@}\left(\dfrac{\partial^2\boldsymbol{\theta}}{(\partial\,\mathbf{s}')^{<2>}}\otimes\dfrac{@}{\partial\,\mathbf{s}'}\right) \\ \left.+\dfrac{\partial^4 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}\,\partial\,\boldsymbol{\theta}'\,@}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\otimes\dfrac{@}{(\partial\,\mathbf{s}')^{<2>}}\right)\right\} + \dfrac{\partial^4 F_{\mathrm{LS}}}{\partial\,\boldsymbol{\theta}(\partial\,\mathbf{s}')^{<3>}} \\ \displaystyle\sum^3\dfrac{\partial^2\mathbf{h}}{(\partial\,\boldsymbol{\theta}')^{<2>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\otimes\dfrac{\partial^2\boldsymbol{\theta}}{(\partial\,\mathbf{s}')^{<2>}}\right) + \dfrac{\partial^3\mathbf{h}}{(\partial\,\boldsymbol{\theta}')^{<3>}}\left(\dfrac{\partial\,\boldsymbol{\theta}}{\partial\,\mathbf{s}'}\right)^{<3>} \end{array}\right]. \tag{26}$$

In (26), the nonzero partial derivatives of F_{LS} evaluated at the population values are

$$\frac{\partial^4 F_{\mathrm{LS}}}{(\partial\,\boldsymbol{\theta}')^{<4>}} = \mathrm{vec}'\mathbf{I}_{pq}\left\{\sum^4 \frac{\partial\,\boldsymbol{\sigma}}{\partial\,\boldsymbol{\theta}'}\otimes\frac{\partial^3\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<3>}} + \sum^3\left(\frac{\partial^2\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<2>}}\right)^{<2>}\right\},$$
$$\frac{\partial^4 F_{\mathrm{LS}}}{(\partial\,\boldsymbol{\theta}')^{<3>}\otimes\partial\,\mathbf{s}'} = -\mathrm{vec}'\frac{\partial^3\boldsymbol{\sigma}}{(\partial\,\boldsymbol{\theta}')^{<3>}}, \tag{27}$$

where the nonzero third partial derivatives of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ are

$$\frac{\partial^3 \boldsymbol{\sigma}}{\partial a_{ij} \partial b_{kj} \partial \phi_j} = \mathrm{vec}\mathbf{E}_{ik} \tag{28}$$

In (22) and (26), the vanishing partial derivatives, when evaluated at the population values, are included for completeness or for later use in similar problems.

4 Inter-battery factor analysis

In this section, we introduce the inter-batter factor analysis model since the model gives the asymptotic robustness of some lower-order cumulants of the NT estimators in the SVD against the violation of the normality assumption under some conditions, which will be shown later. The model of inter-battery factor analysis is written as follows:

$$\mathbf{x} = \mathbf{A}_1\mathbf{f} + \mathbf{e}_X, \ \ \mathbf{y} = \mathbf{B}_1\mathbf{f} + \mathbf{e}_Y, \tag{29}$$

where $\mathbf{A}_1(\mathbf{B}_1)$ is a $p \times K(q \times K)$ factor loading matrix; $\mathbf{f}$ is a $K \times 1$ vector of inter-battery factors; $\mathbf{e}_X(\mathbf{e}_Y)$ is a $p \times 1(q \times 1)$ vector of battery-specific factors. It is assumed that

$$\mathrm{Cov}(\mathbf{f}) = \mathbf{I}_K, \quad \mathrm{Cov}(\mathbf{e}_X) = \boldsymbol{\Psi}_X, \quad \mathrm{Cov}(\mathbf{e}_Y) = \boldsymbol{\Psi}_Y \tag{30}$$

and that $\mathbf{f}$, $\mathbf{e}_X$ and $\mathbf{e}_Y$ are mutually uncorrelated. The number of the inter-battery factors K may be 1 through p (note $p \le q$). From (29) and (30) with associated assumptions, it follows that

$$\mathrm{Cov}\{(\mathbf{x}', \mathbf{y}')'\} = \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{A}_1\mathbf{A}_1' + \boldsymbol{\Psi}_X & \mathbf{A}_1\mathbf{B}_1' \\ \mathbf{B}_1\mathbf{A}_1' & \mathbf{B}_1\mathbf{B}_1' + \boldsymbol{\Psi}_Y \end{bmatrix}. \tag{31}$$

When $K < p$, the model is not a saturated one for the cross covariance matrix. Consequently, the parameter estimators generally depend on discrepancy functions used for estimation. However, when (4) with (5) is used, the parameter estimators for $\mathbf{A}_1$ and $\mathbf{B}_1$ defined below in (32) are equal to the corresponding ones in the saturated model while the estimators for $\boldsymbol{\Psi}_X$ and $\boldsymbol{\Psi}_Y$ tend to become larger in, e.g., Löwner's sense, than those in the saturated model. In the following we consider the case $K = p$.

For simplicity, we use the orthogonal model or uncorrelated inter-battery factors with conventional unit factor variances without loss of generality. With this restriction, however, (31) is still unidentified in that $\mathbf{A}_1$ and $\mathbf{B}_1$ can be replaced by $\mathbf{A}_1\mathbf{T}$ and $\mathbf{B}_1\mathbf{T}'^{-1}$, where $\mathbf{T}$ is a $K \times K$ nonsingular matrix with $\boldsymbol{\Psi}_X = \boldsymbol{\Sigma}_{XX} - \mathbf{A}_1\mathbf{T}\mathbf{T}'\mathbf{A}_1'$ and $\boldsymbol{\Psi}_Y = \boldsymbol{\Sigma}_{YY} - \mathbf{B}_1(\mathbf{T}\mathbf{T}')^{-1}\mathbf{B}_1'$ to yield the same $\boldsymbol{\Sigma}$. It is to be noted that this indeterminacy is not restricted to the rotational one found in the usual factor analysis model (see, e.g., Ogasawara 1986). In this paper, we consider the following practical estimators of $\mathbf{A}_1$ and $\mathbf{B}_1$,

$$\hat{\mathbf{A}}_1 = (p/q)^{1/4}\hat{\mathbf{A}}\hat{\mathbf{\Phi}}^{1/2}, \quad \hat{\mathbf{B}}_1 = (q/p)^{1/4}\hat{\mathbf{B}}\hat{\mathbf{\Phi}}^{1/2}, \tag{32}$$

which give the same mean-square loadings per observed variable in each set (battery) of variables $((pq)^{-1/2}\sum_{i=1}^{K}\hat{\phi}_i)$.

The asymptotic expansions of the distributions of $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{B}}_1$ are given when their partial derivatives with respect to $\mathbf{S}_{XY}$ up to the third order are given. The partial derivatives are obtained by using the chain rule and the partial derivatives of $\mathbf{A}$, $\mathbf{\Phi}$ and $\mathbf{B}$. Similarly, the asymptotic expansions of the distributions of $\hat{\mathbf{\Psi}}_X(= \mathbf{S}_{XX} - \hat{\mathbf{A}}_1\hat{\mathbf{A}}_1')$ and $\hat{\mathbf{\Psi}}_Y(= \mathbf{S}_{YY} - \hat{\mathbf{B}}_1\hat{\mathbf{B}}_1')$ are given from those of $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{B}}_1$. The partial derivatives will be provided in the "Appendix".

5 The estimators for standardized variables

When observed variables are standardized with unit variances, the SVD is carried out to $\mathbf{R}_{XY}$ [see (3)]. In this case (4) with (5) is replaced by

$$\begin{aligned} F_{\rho\mathrm{LS}} &= \frac{1}{2}\mathrm{tr}\{(\mathbf{A}_\rho\mathbf{\Phi}_\rho\mathbf{B}_\rho' - \mathbf{R}_{XY})(\mathbf{A}_\rho\mathbf{\Phi}_\rho\mathbf{B}_\rho' - \mathbf{R}_{XY})'\},\\ \mathbf{\Phi}_\rho &= \mathrm{diag}(\phi_{\rho 1}, \ldots, \phi_{\rho p}), \quad \mathbf{A}_\rho'\mathbf{A}_\rho = \mathbf{I}_p, \quad \mathbf{B}_\rho'\mathbf{B}_\rho = \mathbf{I}_p, \end{aligned} \tag{33}$$

where $\mathbf{A}_\rho$, $\mathbf{B}_\rho$ and $\mathbf{\Phi}_\rho$ are defined similarly to $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Phi}$ in (4). Let $\boldsymbol{\theta}_\rho = (\mathrm{vec}'\mathbf{A}_\rho, \mathrm{vec}'\mathbf{B}_\rho, \mathbf{1}_p'\mathbf{\Phi}_\rho)'$. Then, the asymptotic expansions of the distributions of the LS estimators $\hat{\boldsymbol{\theta}}_\rho$ optimizing (33) are given in a manner similar to those for $\hat{\boldsymbol{\theta}}$. We should note that since $\mathbf{R}_{XY}$ is a matrix function of s_{ii} $(i = 1, \ldots, p+q)$ as well as $\mathbf{S}_{XY}$, $\hat{\boldsymbol{\theta}}_\rho$ is a function of these variances and covariances. In order to have the required partial derivatives for $\hat{\boldsymbol{\theta}}_\rho$, it is convenient to consider two steps: the first step for $\hat{\boldsymbol{\theta}}_\rho$ seen as a function of $\mathbf{R}_{XY}$ followed by the second step of $\mathbf{R}_{XY}$ considered as a function of $\mathbf{S}$ or $\mathbf{s}^*(= \mathrm{v}(\mathbf{S}))$.

In the first step, $\partial^k F_{\rho\mathrm{LS}}/\partial\theta_{\rho\, i_1}\cdots\partial\theta_{\rho\, i_k}$ $(k = 2, 3, 4)$ evaluated at the population values are given by replacing $\mathbf{\Sigma}_{XY}$ and $\boldsymbol{\theta}$ in (19)–(21), (23)–(25), (27) and (28) with $\mathbf{P}_{XY}$ and $\boldsymbol{\theta}_\rho$, respectively. The remaining partial derivatives required are

$$\frac{\partial^k F_{\rho\mathrm{LS}}}{(\partial\,\boldsymbol{\theta}_\rho')^{<j>} \otimes (\partial\,\mathbf{s}')^{<k-j>}} \quad (j = 1, \ldots, k-1; \;\; k = 2, 3, 4). \tag{34}$$

where the partial derivatives of $\mathbf{R}_{XY}$ in the second step with respect to $\mathbf{s}^*$ are required, but are omitted since they are given straightforwardly [see, e.g., Ogasawara 2006, Eq. (29)].

6 Asymptotic robustness of the normal-theory lower-order cumulants of some estimators

Let $\mathbf{u}$ be a $r \times 1$ vector of observed variables given by

$$\mathbf{u} = \mathrm{E}(\mathbf{u}) + \sum_{i=1}^{C} \mathbf{\Lambda}_i \mathbf{f}_i \quad with\ \mathrm{Cov}(\mathbf{f}_i) = \mathbf{\Phi}_i, \tag{35}$$

where $\mathbf{f}_i$ is a $c_i \times 1$ random vector; $\mathbf{\Phi}_i$ is the unconstrained covariance matrix of $\mathbf{f}_i$; $\mathbf{f}_i$ $(i = 1, \ldots, C)$ are mutually independently distributed; and $\mathbf{\Lambda}_i$ has restrictions to have an identified covariance structure of $\mathbf{u}$. Then, it is known that the NT asymptotic standard errors of order $O(n^{-1/2})$ for unweighted/weighted LS estimators of the unknown parameter estimators in $\mathbf{\Lambda}_i$ $(i = 1, \ldots, C)$ are robust against the violation of the normality assumption (Browne and Shapiro 1988, p. 197 and Proposition 3.1). Similarly, the NT asymptotic biases of order $O(n^{-1})$ for the unknown parameter estimators in $\mathbf{\Lambda}_i$ and $\mathbf{\Phi}_i$ $(i = 1, \ldots, C)$ have such robustness (Ogasawara 2005b).

Noting $\mathrm{Cov}(\mathbf{u}) = \sum_{i=1}^{C} \mathbf{\Lambda}_i \mathbf{\Phi}_i \mathbf{\Lambda}_i'$, the SVD of $\mathbf{\Sigma}_{XY}$ is reformulated under (35) with the model of inter-battery factor analysis:

$$\begin{aligned} &\mathbf{\Lambda}_i = [\mathbf{a}_i', \mathbf{b}_i']', \quad \mathbf{\Phi}_i = \phi_i \ \ (i = 1, \ldots, p), \\ &\mathbf{\Lambda}_{p+1} = [\mathbf{I}_p \ \ \mathbf{O}]', \quad \mathbf{\Phi}_{p+1} = \mathbf{\Psi}_X, \quad \mathbf{\Lambda}_{p+2} = [\mathbf{O}, \ \ \mathbf{I}_q]', \ \ \mathbf{\Phi}_{p+2} = \mathbf{\Psi}_Y, \end{aligned} \tag{36}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the ith singular vectors or the ith columns of $\mathbf{A}$ and $\mathbf{B}$ corresponding to the ith largest singular value ϕ_i. Note that ϕ_i corresponds to the variance of the ith inter-battery factor, whose restriction with conventional unit value used in Sect. 4 has been relaxed by regarding ϕ_i's as free parameters to satisfy (35). From (36) with Browne and Shapiro (1988) and Ogasawara (2005b, Theorem 1), we have

Theorem *The NT asymptotic standard errors of order $O(n^{-1/2})$ for the sample singular vectors in the SVD of $\mathbf{S}_{XY}$ hold under nonnormality irrespective of the violation of the normality assumption when the corresponding inter-battery factor model holds with $p + 2$ sets of factors being mutually independently distributed, where the first p sets are p inter-battery factors each with unconstrained distinct variance and the last two are two sets of battery-specific factors with unconstrained covariance matrices. Under the same conditions, the NT asymptotic biases of order $O(n^{-1})$ for the sample singular vectors and the sample singular values are robust under nonnormality.*

For $\hat{\mathbf{\Psi}}_X (= \mathbf{S}_{XX} - \hat{\mathbf{A}}_1 \hat{\mathbf{A}}_1')$ and $\hat{\mathbf{\Psi}}_Y (= \mathbf{S}_{YY} - \hat{\mathbf{B}}_1 \hat{\mathbf{B}}_1')$, we have from Theorem and (11) (see also the proof of Corollary 4, Ogasawara 2005a, 2007a)

Corollary 1 *When $\mathbf{A}_1$ and $\mathbf{B}_1$ with $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ are given from the transformation such that*

$$\begin{aligned} &\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B}_1 \end{bmatrix} \left[\mathbf{A}_1' \ \mathbf{B}_1'\right] = \left[\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_p\right] diag(\phi_1, \ldots, \phi_p) \left[\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_p\right]', \\ &\mathbf{\Lambda}_{p+1} = [\mathbf{I}_p \ \ \mathbf{O}]', \quad \mathbf{\Lambda}_{p+2} = [\mathbf{O} \ \ \mathbf{I}_q]', \\ &\mathbf{\Psi}_X = \mathbf{\Phi}_{p+1} = \mathbf{\Sigma}_{XX} - \mathbf{A}_1 \mathbf{A}_1', \quad \mathbf{\Psi}_Y = \mathbf{\Phi}_{p+2} = \mathbf{\Sigma}_{YY} - \mathbf{A}_2 \mathbf{A}_2', \end{aligned} \tag{37}$$

where $\mathbf{\Lambda}_i$ *and* $\mathbf{\Phi}_i$ $(i = 1, \ldots, p+2)$ *satisfy* (35) *with the associated conditions, the NT asymptotic biases of* $\hat{\mathbf{\Psi}}_X$ *and* $\hat{\mathbf{\Psi}}_Y$ *are robust under nonnormality.*

It is to be noted that while $\hat{\mathbf{A}}_1\hat{\mathbf{B}}_1' = \hat{\mathbf{A}}\hat{\mathbf{\Phi}}\hat{\mathbf{B}}'$, $\hat{\mathbf{\Psi}}_X(\hat{\mathbf{\Psi}}_Y)$ is generally different from $\mathbf{S}_{XX} - \hat{\mathbf{A}}\hat{\mathbf{\Phi}}\hat{\mathbf{A}}'$ $(\mathbf{S}_{YY} - \hat{\mathbf{B}}\hat{\mathbf{\Phi}}\hat{\mathbf{B}}')$, which also enjoy the robustness in Corollary 1 under the conditions in Theorem. Unfortunately, $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{B}}_1$ do not have the robust property shown in Corollary 1 since the corresponding variances (ϕ_i) are set to be 1 [see the first equation of (37)].

Using Ogasawara (2005a, 2007a, Corollary 4),

Corollary 2 *The asymptotic standard errors of order* $O(1)$ *and the asymptotic biases of order* $O(n^{-1/2})$ *of the NT Studentized estimators of the singular vectors in Theorem with the associated conditions are robust under nonnormality.*

The obvious results of the unit asymptotic standard errors in Corollary 2 are included for completeness. More general results are also obtained:

Corollary 3 *Let* $g(\mathbf{A}, \mathbf{B})$ *be a differentiable function of* $\mathbf{A}$ *and* $\mathbf{B}$ *with* $\hat{g} = g(\hat{\mathbf{A}}, \hat{\mathbf{B}})$. *Then, the NT asymptotic standard error of order* $O(n^{-1/2})$ *and the NT asymptotic bias of order* $O(n^{-1})$ *for* $\hat{g}$ *are robust against the normality assumption under the conditions of Theorem.*

Proof The robustness of the asymptotic standard error is given by the delta method with the corresponding robustness of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$. For the remaining result, let $\boldsymbol{\theta}^*(Q^* \times 1) = (\mathrm{vec}'\mathbf{A}, \mathrm{vec}'\mathbf{B})'$ with $Q^* = p(p+q)$, then n times the asymptotic bias of $\hat{g}$ is given by the formula of (11):

$$\begin{aligned}\frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 g(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\sigma}\,\partial\boldsymbol{\sigma}'}\mathbf{\Omega}\right) &= \frac{1}{2}\mathrm{tr}\left\{\left(\sum_{i=1}^{Q^*}\frac{\partial\, g(\boldsymbol{\theta}^*)}{\partial\,\theta_i^*}\frac{\partial^2\theta_i^*}{\partial\boldsymbol{\sigma}\,\partial\boldsymbol{\sigma}'} + \frac{\partial^2 g(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\sigma}\,\partial\boldsymbol{\theta}^{*\prime}}\frac{\partial\,\boldsymbol{\theta}^*}{\partial\boldsymbol{\sigma}'}\right)\mathbf{\Omega}\right\} \\ &= \sum_{i=1}^{Q^*}\left\{\frac{\partial\, g(\boldsymbol{\theta}^*)}{\partial\,\theta_i^*}\, n\ \mathrm{abis}(\hat{\theta}_i^*) + \frac{n}{2}\mathrm{acov}\left(\left.\frac{\partial\, g(\boldsymbol{\theta}^*)}{\partial\,\theta_i^*}\right|_{\boldsymbol{\theta}^*=\hat{\boldsymbol{\theta}}^*}, \hat{\theta}_i^*\right)\right\},\end{aligned} \tag{38}$$

where abis$(\cdot)$ and acov$(\ \cdot,\ \cdot)$ are the asymptotic bias and covariance of order $O(n^{-1})$ for the argument estimators, respectively. Since the abis$(\cdot)$ and acov$(\ \cdot,\ \cdot)$ in (38) are robust , the whole result of (38) has the robust property. □

If $g(\cdot)$ is a differentiable function of $\mathbf{\Phi}$ as well as $\mathbf{A}$ and $\mathbf{B}$, Corollary 3 does not generally hold (an exception is the linear function of the elements of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Phi}$) since some of the acov$(\ \cdot,\ \cdot)$'s in (38) have not the robust property although abis$(\cdot)$'s in (38) are robust. None of the parameter estimators $\hat{\boldsymbol{\theta}}_\rho$ have the robust properties shown above since standardization of observed variables violates the necessary condition of unconstrained $\mathbf{\Phi}_i$.

7 A numerical example with simulations

A small numerical example with $p = 2$ and $q = 2$ is given for illustration:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & & & \text{sym.} \\ 0.4 & 1 & & \\ 0.3 & 0.2 & 1 & \\ 0.2 & 0.3 & 0.4 & 1 \end{bmatrix}, \quad \mathbf{A} = \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Big/ \sqrt{2}, \quad \boldsymbol{\Phi} = \mathrm{diag}(0.5,\ 0.1), \tag{39}$$
$$\mathbf{A}_1 = \mathbf{B}_1 = \mathbf{A}\boldsymbol{\Phi}^{1/2} = \begin{bmatrix} 5 & \sqrt{5} \\ 5 & -\sqrt{5} \end{bmatrix} /10, \quad \boldsymbol{\Psi}_X = \boldsymbol{\Psi}_Y = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix},$$

where the transformation of (32) is used. Simulations were performed to see the accuracy of the asymptotic results in finite samples under normality and nonnormality. Nonnormal observations were generated to satisfy (35), where the two inter-battery factors were independently chi-square distributed with 1 *df* followed by scaling. The two sets of battery-specific factors were given by $\mathbf{e}_X = \boldsymbol{\Psi}_X^* \mathbf{f}_X$ and $\mathbf{e}_Y = \boldsymbol{\Psi}_Y^* \mathbf{f}_Y$, where $\boldsymbol{\Psi}_X^*$ and $\boldsymbol{\Psi}_Y^*$ are the Cholesky-decomposed lower-triangular matrices of $\boldsymbol{\Psi}_X$ and $\boldsymbol{\Psi}_Y$, respectively, and each element of $\mathbf{f}_X$ and $\mathbf{f}_Y$ is independently chi-square distributed with 1 *df* with scaling to have unit variance.

In the first half of the simulations sample sizes were varied from $n = N - 1 =$ 100–800. From each set of generated observations, the parameters were estimated using the SVD of $\mathbf{S}_{XY}$ with $\mathbf{S}_{XX}$ and $\mathbf{S}_{YY}$ for $\boldsymbol{\Psi}_X$ and $\boldsymbol{\Psi}_Y$. NT Studentized parameter estimates were also obtained using estimated NT standard errors. This was replicated 1,000,000 times. Tables 1 and 2 show the simulated and theoretical values of the cumulants for selected parameter estimators. The simulated cumulants were given by k-statistics (unbiased estimators of cumulants) based on 1,000,000 estimates for each parameter with multiplications of appropriate powers of n for ease of comparison. In Table 2, the theoretical ratio HASE/ASE $= \sqrt{(\alpha_2/n) + (\Delta\alpha_2/n^2)}/\sqrt{\alpha_2/n}$ (asymptotic standard error, ASE; higher-order ASE, HASE) depends on n. The corresponding simulated ratio is SD/ASE, where SD is the square root of the usual unbiased variance based on 1,000,000 estimates.

In the tables an "a" indicates that the corresponding NT value holds due to the asymptotic robustness. The results of the tables show that the asymptotic values are reasonably similar to their corresponding simulated ones except for some results with relatively small sample sizes. The simulated values corresponding to the asymptotically robust NT cumulants well show the robustness. The results for a_{11} in Table 2 when n is less than 400 are unstable. The sample size more than 400 may be needed to have stable results for the Studentized estimators of the singular vectors. It is of interest to find that Studentization tends to give the reversal of the sign of skewness in these data. The absolute values of the cumulants in the nonnormal case are mostly larger than the corresponding NT values, which stems primarily from the large kurtosis of the nonnormal distribution used in the example.

Table 3 gives the results for standardized observed variables. The simulations were similarly performed as in Tables 1 and 2 with the SVD of $\mathbf{R}_{XY}$ instead of $\mathbf{S}_{XY}$ with $\hat{\boldsymbol{\Psi}}_{X\rho} = \mathbf{R}_{XX} - \hat{\mathbf{A}}_{1\rho}\hat{\mathbf{A}}'_{1\rho}$ and $\hat{\boldsymbol{\Psi}}_{Y\rho} = \mathbf{R}_{YY} - \hat{\mathbf{B}}_{1\rho}\hat{\mathbf{B}}'_{1\rho}$. In the table, the simulated values are presented only with $n = 800$. While the asymptotic robustness found in

Table 1 Simulated and theoretical cumulants of the non-Studentized estimators in unstandardized-variable data

Parameter	n	$\alpha_2^{1/2}$(dispersion)		α_1(bias)		α_3(skewness)		α_4(kurtosis)	
		Nml	C1	Nml	C1	Nml	C1	Nml	C1
a_{11}	100	1.54	1.77	−1.68	−2.24	−36.0	−79.5	1,533	5,368
	200	1.47	1.62	−1.48	−1.86	−25.0	−53.8	976	4,090
	400	1.42	1.49	−1.40	−1.60	−19.2	−29.9	578	2,112
	800	1.41	1.44	−1.34	−1.41	−17.1	−21.7	453	1,189
	Th.	1.39	[a]	−1.37	[a]	−16.0	−17.3	402	838
ϕ_1	100	1.46	2.26	1.82	1.87	6.0	106	75	5,363
	200	1.47	2.26	1.83	1.91	6.6	110	46	5,404
	400	1.48	2.27	1.80	1.88	6.4	111	44	5,334
	800	1.48	2.28	1.82	1.88	6.6	110	64	5,390
	Th.	1.49	2.28	1.77	[a]	6.1	111	41	5,471
ϕ_2	100	0.56	0.63	−0.19	−0.07	0.9	2.3	0	23
	200	0.60	0.68	−0.42	−0.40	0.5	1.7	−4	11
	400	0.61	0.69	−0.46	−0.49	0.2	1.3	−1	11
	800	0.61	0.70	−0.44	−0.48	0.2	1.2	0	18
	Th.	0.61	0.70	−0.45	[a]	0.2	1.3	1	18
ψ_{X11}	100	1.38	2.76	−0.81	−0.88	5.8	235	32	15,828
	200	1.39	2.78	−0.70	−0.72	6.1	244	39	16,892
	400	1.39	2.79	−0.68	−0.65	5.9	244	32	16,531
	800	1.40	2.80	−0.66	−0.54	6.1	241	43	15,909
	Th.	1.40	2.80	−0.66	[a]	6.3	246	37	16,927

$n + 1$, The sample size in the simulation; Th., theoretical or asymptotic values; Nml, normally distributed data; C1, chi-square distributed data with $df = 1$

[a] Corresponding normal-theory values

unstandardized observed variables is lost in Table 3, some of the large α_3 and α_4 have been reduced.

In Table 4, the results of the simulations of confidence intervals for selected parameters are shown under normality with $n = 200$ for unstandardized observed variables. The one-sided confidence intervals were constructed in three ways: the usual normal approximation, the Cornish–Fisher expansion and Hall's (1992b) method by variable transformation each based on a data set randomly generated as before. The number of replications was reduced to 100,000 due to the excessive computation time required. Table 4 shows the proportions of true values below the endpoints of the 100,000 confidence intervals for each parameter. We find that the proportions given by the Cornish–Fisher expansion and Hall's method give values more similar to nominal ones than the usual normal approximation. The results of Table 4 are encouraging in that in Table 2, the simulated third cumulant under normality corresponding to α'_{NT3} was unstable with $n = 200$ while Table 4 gives reasonable values, which suggests that

Table 2 Simulated and theoretical HASEs of the non-Studentized estimators and cumulants of the Studentized estimators in unstandardized-variable data

Parameter	n	$\frac{\text{SD}}{\text{ASE}}$	$\frac{\text{HASE}}{\text{ASE}}$	$\frac{\text{SD}}{\text{ASE}}$	$\frac{\text{HASE}}{\text{ASE}}$	$\alpha_{\text{NT2}}^{1/2\prime}$	$\alpha_{\text{NT2}}^{1/2\prime\prime}$	α'_{NT1}	α''_{NT1}	α'_{NT3}	α''_{NT3}
		Nml		C1		Nml	C1	Nml	C1	Nml	C1
a_{11}	100	1.107	1.081	1.269	1.219	55.2	588.5	3.49	11.59	1.3×10^9	2.0×10^{12}
	200	1.052	1.041	1.163	1.115	3.5	6.6	1.33	1.83	4.2×10^5	2.1×10^6
	400	1.022	1.021	1.071	1.059	1.026	1.189	1.08	1.12	8.0	3.9×10^3
	800	1.010	1.010	1.033	1.030	1.011	1.015	1.06	1.07	6.6	6.2
	Th.					1	[a]	0.98	[a]	5.9	5.4
ϕ_1	100	0.980	0.977	0.990	0.982	1.009	1.531	0.19	−1.61	−5.7	−16.8
	200	0.988	0.988	0.991	0.991	0.995	1.506	0.26	−1.45	−4.0	−10.1
	400	0.994	0.994	0.995	0.996	0.996	1.512	0.26	−1.41	−3.7	−6.0
	800	0.998	0.997	0.997	0.998	0.999	1.520	0.28	−1.39	−3.6	−4.9
	Th.					1	1.535	0.26	−1.48	−3.7	−4.0
ϕ_2	100	0.927	0.994	0.906	0.945	0.886	0.986	−0.73	−0.92	1.3	2.6
	200	0.982	0.997	0.954	0.973	0.965	1.075	−1.14	−1.47	−0.3	1.2
	400	0.999	0.999	0.985	0.987	0.991	1.122	−1.22	−1.63	−1.7	−0.8
	800	1.000	0.999	0.994	0.993	0.996	1.138	−1.20	−1.64	−1.9	−1.3
	Th.					1	1.151	−1.23	−1.60	−1.9	−1.2
ψ_{X11}	100	0.984	0.989	0.985	0.987	1.029	2.013	−1.76	−5.57	−5.0	−41.6
	200	0.993	0.995	0.994	0.994	1.014	1.999	−1.66	−5.57	−4.8	−40.3
	400	0.996	0.997	0.997	0.997	1.006	1.996	−1.64	−5.61	−4.8	−38.6
	800	0.997	0.999	0.999	0.998	1.002	1.998	−1.63	−5.58	−4.7	−39.3
	Th.					1	2.000	−1.62	−5.72	−4.6	−36.1

$n+1$, The sample size in the simulation and the theoretical ratio (HASE/ASE); Th., theoretical or asymptotic values; SD, standard deviation from the simulation; ASE $= \sqrt{\alpha_2/n}$; HASE $= \sqrt{(\alpha_2/n) + (\Delta\alpha_2/n^2)}$; Nml, normally distributed data; C1, chi-square distributed data with $df = 1$

[a] Corresponding normal-theory values

Table 3 Simulated and theoretical cumulants of the non-Studentized and Studentized estimators, and HASEs of the non-Studentized estimators in standardized-variable data

Parameter	N	$\alpha_2^{1/2}$(dispersion)		α_1(bias)		α_3(skewness)		α_4(kurtosis)	
		Nml	C1	Nml	C1	Nml	C1	Nml	C1
a_{11}	800	1.29	1.47	−1.13	−1.40	−12.6	−23.0	431	1, 341
	Th.	1.28	1.43	−1.15	−1.33	−11.3	−17.8	359	961
ϕ_1	800	1.29	1.91	1.31	2.59	−2.1	19.8	7	190
	Th.	1.29	1.93	1.28	2.57	−2.3	22.0	−4	287
ϕ_2	800	0.60	0.70	−0.29	−0.11	0.1	1.1	−1	14
	Th.	0.60	0.70	−0.29	−0.06	0.2	1.2	1	14
ψ_{X11}	800	1.00	1.28	−0.54	−1.35	0.9	−4.0	−5	−46
	Th.	1.00	1.29	−0.49	−1.35	0.9	−4.2	−4	−38

Parameter	N	SE ratio		$\alpha_{\mathrm{NT2}}^{1/2\prime}$	$\alpha_{\mathrm{NT2}}^{1/2\prime\prime}$	α'_{NT1}	α''_{NT1}	α'_{NT3}	α''_{NT3}
		Nml	C1	Nml	C1	Nml	C1	Nml	C1
a_{11}	800	1.014	1.029	1.001	1.124	0.96	1.21	5.6	8.3
	Th.	1.014	1.026	1	1.121	0.90	1.15	5.4	7.9
ϕ_1	800	0.996	0.988	1.003	1.496	1.25	2.34	0.4	14.1
	Th.	0.995	0.989	1	1.495	1.24	2.33	0.4	14.8
ϕ_2	800	1.002	0.997	0.997	1.154	−0.81	−0.84	−1.4	−0.2
	Th.	1.001	0.996	1	1.165	−0.84	−0.78	−1.4	−0.1
ψ_{X11}	800	0.997	0.992	1.004	1.301	−0.84	−1.91	−0.9	−9.9
	Th.	0.997	0.992	1	1.293	−0.80	−1.90	−0.9	−9.7

$n+1$, The sample size in the simulation and the theoretical SE ratio; HASE/ASE$=\sqrt{(\alpha_2/n)+(\Delta\alpha_2/n^2)}/\sqrt{\alpha_2/n}$

The simulated SE ratio is SD/ASE, where SD is the standard deviation from the simulation. Th., theoretical or asymptotic values; Nml, normally distributed data; C1, chi-square distributed data with $df = 1$

most of the estimated α'_{NT3} used in the confidence intervals with $n = 200$ are stable in practical sense.

Table 5 gives the simulated and theoretical cumulants of the selected estimators Studentized by the asymptotically distribution (ADF) theory. In the simulations the sample cumulants were given from the corresponding sample moments up to the sixth order with the number of replications 100,000, where the results of the normally distributed data were included for comparison to those of the nonnormal data generated as before. Note that for the normal data Studentized estimators were given by the ADF theory. The sample sizes $n = 200-1{,}600$ were used considering the relatively unstable results expected. As in Table 2 given by the normal-theory Studentized estimators, the simulated results of $\alpha_2^{1/2\prime}$ and α'_3 for a_{11} in Table 5 are unstable while the remaining simulated results are reasonably similar to the corresponding asymptotic values.

In Table 6, the results of the confidence intervals based on the ADF theory using the same nonnormal data in Table 5 are represented, where the number of replications

Table 4 Simulated proportions below the lower endpoints of the confidence intervals based on the normal-theory Studentized estimators under normality in unstandardized-variable data ($n = 200$)

Parameter	Method	Nominal values						
		0.0050	0.0250	0.1000	0.5000	0.9000	0.9750	0.9950
a_{11}	N*	0.0202	0.0490	0.1257	0.5023	0.9182	0.9858	0.9983
	C-F	0.0067	0.0290	0.1057	0.5020	0.8930	0.9679	0.9899
	Hall	0.0028	0.0217	0.1040	0.5017	0.8941	0.9712	0.9929
ϕ_1	N*	0.0018	0.0168	0.0952	0.5265	0.8986	0.9693	0.9912
	C-F	0.0066	0.0283	0.1044	0.5025	0.8928	0.9699	0.9922
	Hall	0.0056	0.0266	0.1040	0.5025	0.8931	0.9707	0.9928
ϕ_2	N*	0.0017	0.0143	0.0779	0.4736	0.8861	0.9718	0.9990
	C-F	0.0073	0.0298	0.1064	0.4990	0.8944	0.9735	0.9991
	Hall	0.0068	0.0290	0.1062	0.4990	0.8944	0.9735	0.9991
ψ_{X11}	N*	0.0012	0.0114	0.0731	0.4768	0.8720	0.9558	0.9868
	C-F	0.0071	0.0291	0.1039	0.5016	0.8946	0.9712	0.9934
	Hall	0.0058	0.0274	0.1034	0.5016	0.8951	0.9727	0.9941

$n+1$, The sample size in the simulation; N*, normal approximation; C–F, Cornish–Fisher expansion; Hall, Hall's method by variable transformation

was 100,000 with $n = 200$ and 800. The results by the Cornish–Fisher expansion are omitted since they are slightly poorer than those by Hall's method. The results by Hall are better than those by the usual normal approximation especially when the sample size is relatively large with some exceptions (e.g., $n = 200$, a_{11}, the nominal value $=$ 0.9950).

Table 7 shows the overall sizes of the errors of the approximations of the cumulative distribution functions for selected parameter estimators which are standardized with the population asymptotic standard errors. The approximations were given by four methods shown in the table using the population asymptotic cumulants: N*, E1, E2 and Hall in the table stand for the usual normal approximation, the single-term Edgeworth expansion, the two-term Edgeworth expansion and Hall's method, respectively. The true values were given from the simulations used in Tables 1, 2 and 3. The root mean square error was obtained from the square root of the mean of the squared errors over the 40 points of a standardized parameter estimator (i.e.,$-3.8, -3.6, \ldots, 4.0$), where the error was defined as an approximated value minus the corresponding true or simulated proportion. From the table, we see that E1 and Hall have considerably reduced the errors of N*, and that E2 has further reduced the errors of E1 and Hall with some exceptions (normal unstandardized-variable data, $n = 800$, ϕ_2; normal standardized-variable data, $n = 200$, ϕ_2). The sizes of the errors of E1 and Hall seem to be similar.

8 Discussion

In the SVD of $\boldsymbol{\Sigma}_{XY}$, distinct nonzero singular values have been assumed. When some of them are equal, the singular vectors become unidentified. Let $\boldsymbol{\Phi} = \mathrm{diag}(\phi_1 \mathbf{1}'_{p_1}, \ldots,$

Table 5 Simulated and theoretical cumulants of the Studentized estimators in unstandardized-variable data

Parameter	n	$\alpha_2^{1/2\prime}$		α_1'		α_3'	
		Nml	C1	Nml	C1	Nml	C1
a_{11}	200	1.254	28.3	1.37	1.88	533.5	3.2×10^4
	400	1.040	1.051	1.10	1.25	7.7	11.8
	800	1.021	1.027	1.20	1.23	6.9	7.6
	1,600	1.011	1.011	1.16	1.24	6.2	7.3
	Th.	1	1	0.98	1.23	5.9	6.9
ϕ_1	200	1.022	1.177	0.24	−3.19	−4.8	−22.2
	400	1.007	1.100	0.30	−3.35	−3.8	−18.6
	800	1.008	1.059	0.21	−3.45	−3.4	−17.8
	1,600	1.006	1.037	0.34	−3.43	−4.1	−18.7
	Th.	1	1	0.26	−3.97	−3.7	−19.2
ϕ_2	200	0.983	1.020	−1.18	−2.15	−0.3	−2.9
	400	1.001	1.028	−1.29	−2.34	−1.7	−5.4
	800	1.002	1.017	−1.25	−2.38	−2.2	−6.5
	1,600	1.004	1.010	−1.21	−2.25	−2.1	−7.1
	Th.	1	1	−1.23	−2.53	−1.9	−7.6
ψ_{X11}	200	1.040	1.280	−1.68	−5.47	−5.3	−30.5
	400	1.019	1.174	−1.62	−5.54	−5.1	−27.6
	800	1.010	1.101	−1.64	−5.56	−4.4	−24.3
	1,600	1.007	1.062	−1.67	−5.69	−4.9	−24.4
	Th.	1	1	−1.62	−5.98	−4.6	−23.2

$n+1$, The sample size in the simulation; Th., theoretical or asymptotic values; Nml = normally distributed data; C1 = chi-square distributed data with $df = 1$

$\phi_M \mathbf{1}'_{p_M}$), where $p_1, \ldots, p_M$ $\left(p = \sum_{i=1}^M p_i; \phi_1 > \cdots > \phi_M > 0\right)$ denote multiplicities of M distinct singular values. Let $\mathbf{A}_{[i]}$ and $\mathbf{B}_{[i]}$ be the $p \times p_i$ and $q \times p_i$ matrices whose columns are the unit-norm orthogonal singular vectors corresponding to the singular value ϕ_i with multiplicity p_i $(i = 1, \ldots, M)$. Then, $\mathbf{A}_{[i]}$ and $\mathbf{B}_{[i]}$ are identified up to the post-multiplication of an orthogonal matrix say $\mathbf{T}_i$. For instance appropriate $(p_i^2 - p_i)/2$ elements of $\mathbf{A}_{[i]}$ or $\mathbf{B}_{[i]}$ can be set to zero, e.g., to have an echelon form with variable reordering if necessary. A similar method is used by Boik (1998, p. 247). When the multiplicities are known, the non-fixed parameters in $\mathbf{A}_{[i]}$, $\mathbf{B}_{[i]}$ and ϕ_i $(i = 1, \ldots, M)$ are estimated by LS based on $\mathbf{S}_{XY}$ or $\mathbf{R}_{XY}$ using some numerical methods. Since the model with multiple singular values is not a saturated one, the estimators depend on discrepancy functions used. The asymptotic expansions of the distributions of these estimators can be similarly given from the first-order conditions as in (8).

In this paper, the SVDs of asymmetric matrices have been dealt with. The corresponding results for symmetric cases reduce to the eigenvalue and eigenvector problems of sample covariance/correlation matrices, which have been well investigated. It is known

Table 6 Simulated proportions below the lower endpoints of the confidence intervals based on the ADF-theory Studentized estimators under nonnormality ($\chi^2, df = 1$) in unstandardized-variable data ($n = 200$)

Parameter	Method	Nominal values						
		0.0050	0.0250	0.1000	0.5000	0.9000	0.9750	0.9950
$n = 200$								
a_{11}	N*	0.0253	0.0577	0.1368	0.5005	0.9137	0.9835	0.9970
	Hall	0.0134	0.0394	0.1207	0.4979	0.8729	0.9564	0.9851
ϕ_1	N*	0.0002	0.0053	0.0620	0.4823	0.8303	0.9176	0.9598
	Hall	0.0057	0.0321	0.1164	0.4832	0.8453	0.9390	0.9769
ϕ_2	N*	0.0010	0.0107	0.0725	0.4620	0.8548	0.9547	0.9906
	Hall	0.0125	0.0421	0.1268	0.4933	0.8624	0.9533	0.9881
ψ_{X11}	N*	0.0001	0.0033	0.0482	0.4434	0.7910	0.8855	0.9368
	Hall	0.0054	0.0312	0.1157	0.4733	0.8299	0.9288	0.9705
$n = 800$								
a_{11}	N*	0.0112	0.0374	0.1164	0.5007	0.9078	0.9816	0.9972
	Hall	0.0080	0.0318	0.1086	0.4992	0.8914	0.9693	0.9922
ϕ_1	N*	0.0006	0.0098	0.0739	0.4916	0.8614	0.9456	0.9798
	Hall	0.0060	0.0287	0.1096	0.4960	0.8782	0.9619	0.9891
ϕ_2	N*	0.0018	0.0153	0.0837	0.4829	0.8761	0.9609	0.9898
	Hall	0.0077	0.0313	0.1114	0.4987	0.8847	0.9637	0.9896
ψ_{X11}	N*	0.0004	0.0072	0.0662	0.4729	0.8417	0.9314	0.9711
	Hall	0.0058	0.0300	0.1097	0.4942	0.8712	0.9578	0.9880

$n + 1$, The sample size in the simulation; N*, normal approximation; Expansion; Hall; Hall's method by variable transformation

that the asymptotic robustness similar to that in this paper is found for unit-norm eigenvectors and eigenvalues of sample covariance matrices (see Ogasawara 2005b). It is of interest to see that in the case of (36), $2p + q$ latent variables are required to have the asymptotic robustness while in the case of the eigenvalue and eigenvector problem, independently distributed full components in principal component analysis are required whose number is equal to that of the observed variables. Note that $2p + q$ is greater than the number of observed variables by p.

Appendix A: The partial derivatives of $\mathbf{A}_1$ and $\mathbf{B}_1$ evaluated at the population values

Let $k^* = (p/q)^{1/4}$. Then,

$$\frac{1}{k^*}\frac{\partial \mathbf{A}_1}{\partial \sigma_{p+c,d}} = \frac{\partial \mathbf{A}}{\partial \sigma_{p+c,d}}\mathbf{\Phi}^{1/2} + \frac{1}{2}\mathbf{A}\mathbf{\Phi}^{-1/2}\frac{\partial\, \mathbf{\Phi}}{\partial \sigma_{p+c,d}},$$

$$\frac{1}{k^*}\frac{\partial^2 \mathbf{A}_1}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}} = \frac{\partial^2 \mathbf{A}}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}}\mathbf{\Phi}^{1/2}$$

Table 7 $10^5 \times$Root mean square errors of the asymptotic distribution functions of the standardized estimators

Parameter	Normal				Chi-square ($df = 1$)			
	N*	E1	E2	Hall	N*	E1	E2	Hall
Unstandardized-variable data								
$n = 200$								
a_{11}	1,185	338	117	241	1612	1,034	350	936
ϕ_1	1,458	181	43	190	1,478	751	330	691
ϕ_2	1,169	209	205	218	1,635	647	325	706
ψ_{X11}	983	105	50	132	2,198	680	358	846
$n = 800$								
a_{11}	553	79	33	57	634	257	46	238
ϕ_1	721	43	22	43	727	215	66	191
ϕ_2	531	33	35	32	725	139	27	154
ψ_{X11}	474	43	27	50	1,107	195	90	216
Standardized-variable data								
$n = 200$								
a_{11}	1,134	457	150	362	1,467	880	337	764
ϕ_1	1,517	214	66	262	1,695	588	210	585
ϕ_2	812	200	211	198	1,025	519	340	534
ψ_{X11}	842	146	75	159	1311	331	103	344
$n = 800$								
a_{11}	513	110	33	85	599	215	45	189
ϕ_1	735	49	21	61	789	170	37	169
ϕ_2	350	40	31	38	395	106	41	108
ψ_{X11}	415	41	31	44	614	82	13	85

$n + 1$, The sample size in the simulation; N*, normal approximation, E1, the single-term Edgeworth expansion; E2, the two-term Edgeworth expansion; Hall, Hall's method by variable transformation

$$+\frac{1}{2}\left(\frac{\partial \mathbf{A}}{\partial \sigma_{p+c,d}}\frac{\partial\,\mathbf{\Phi}}{\partial \sigma_{p+e,f}}+\frac{\partial \mathbf{A}}{\partial \sigma_{p+e,f}}\frac{\partial\,\mathbf{\Phi}}{\partial \sigma_{p+c,d}}\right)\mathbf{\Phi}^{-1/2}$$
$$-\frac{1}{4}\mathbf{A}\mathbf{\Phi}^{-3/2}\frac{\partial\,\mathbf{\Phi}}{\partial \sigma_{p+c,d}}\frac{\partial\,\mathbf{\Phi}}{\partial \sigma_{p+e,f}}$$
$$+\frac{1}{2}\mathbf{A}\mathbf{\Phi}^{-1/2}\frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}},$$

$$\frac{1}{k^*}\frac{\partial^3 \mathbf{A}_1}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}} = \frac{\partial^3 \mathbf{A}}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}}\mathbf{\Phi}^{1/2}$$
$$+\sum_{(U,V,W)}^{3}\left\{\frac{1}{2}\left(\frac{\partial^2 \mathbf{A}}{\partial \sigma_U \partial \sigma_V}\frac{\partial\,\mathbf{\Phi}}{\partial \sigma_W}+\frac{\partial \mathbf{A}}{\partial \sigma_U}\frac{\partial^2 \mathbf{\Phi}}{\partial \sigma_V \partial \sigma_W}\right)\mathbf{\Phi}^{-1/2}\right.$$

$$
\begin{aligned}
&-\frac{1}{4}\left(\frac{\partial \mathbf{A}}{\partial \sigma_U}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_V}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_W}+\mathbf{A}\frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_U \partial \sigma_V}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_W}\right)\boldsymbol{\Phi}^{-3/2}\Bigg\}\\
&+\frac{3}{8}\mathbf{A}\boldsymbol{\Phi}^{-5/2}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_{p+c,d}}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_{p+e,f}}\frac{\partial\,\boldsymbol{\Phi}}{\partial \sigma_{p+g,h}}\\
&+\frac{1}{2}\mathbf{A}\boldsymbol{\Phi}^{-1/2}\frac{\partial^3 \boldsymbol{\Phi}}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}}\\
&(c,e,g=1,\ldots,q;\;\; d,f,h=1,\ldots,p),
\end{aligned}
\tag{40}
$$

where $\sum_{(U,V,W)}^{3}$ denotes a summation over the range:

$$
\begin{aligned}
(U,V,W)\in\;&\{(p+c,d;\,p+e,f;\,p+g,h),(p+e,f;\,p+g,h;\,p+c,d),\\
&(p+g,h;\,p+c,d;\,p+e,f)\}.
\end{aligned}
$$

The partial derivatives of $\mathbf{B}_1$ can be obtained by replacing $\mathbf{A}$, $\mathbf{A}_1$ and $1/k^*$ in (40) with $\mathbf{B}$, $\mathbf{B}_1$ and k^*, respectively.

Appendix B: The partial derivatives of $\boldsymbol{\Psi}_X$ and $\boldsymbol{\Psi}_Y$ evaluated at the population values

The nonzero partial derivatives are as follows:

$$
\begin{aligned}
\frac{\partial\,\boldsymbol{\Psi}_X}{\partial \sigma_{cd}}&=\frac{2-\delta_{cd}}{2}(\mathbf{E}_{cd}+\mathbf{E}_{dc})\quad (p\ge c\ge d\ge 1),\\
\frac{\partial\,\boldsymbol{\Psi}_X}{\partial \sigma_{p+c,d}}&=-\frac{\partial \mathbf{A}_1}{\partial \sigma_{p+c,d}}\mathbf{A}_1'-\mathbf{A}_1\frac{\partial \mathbf{A}_1'}{\partial \sigma_{p+c,d}}\;\;(c=1,\ldots,q;\quad d=1,\ldots,p),\\
\frac{\partial^2 \boldsymbol{\Psi}_X}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}}&=-\frac{\partial^2 \mathbf{A}_1}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}}\mathbf{A}_1'-\frac{\partial \mathbf{A}_1}{\partial \sigma_{p+c,d}}\frac{\partial \mathbf{A}_1'}{\partial \sigma_{p+e,f}}\\
&\quad-\frac{\partial \mathbf{A}_1}{\partial \sigma_{p+e,f}}\frac{\partial \mathbf{A}_1'}{\partial \sigma_{p+c,d}}-\mathbf{A}_1\frac{\partial^2 \mathbf{A}_1'}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}}\\
&\quad(c,e=1,\ldots,q;\,d,f=1,\ldots,p),\\
\frac{\partial^3 \boldsymbol{\Psi}_X}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}}&=-\frac{\partial^3 \mathbf{A}_1}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}}\mathbf{A}_1'\\
&\quad-\sum_{(U,V,W)}^{3}\left(\frac{\partial^2 \mathbf{A}_1}{\partial \sigma_U\partial \sigma_V}\frac{\partial \mathbf{A}_1'}{\partial \sigma_W}+\frac{\partial \mathbf{A}_1}{\partial \sigma_U}\frac{\partial^2 \mathbf{A}_1'}{\partial \sigma_V\partial \sigma_W}\right)\\
&\quad-\mathbf{A}_1\frac{\partial^3 \mathbf{A}_1'}{\partial \sigma_{p+c,d}\partial \sigma_{p+e,f}\partial \sigma_{p+g,h}}\\
&\quad(c,e,g=1,\ldots,q;\;\; d,f,h=1,\ldots,p).
\end{aligned}
\tag{41}
$$

The partial derivatives for $\boldsymbol{\Psi}_Y$ are given by replacing $\boldsymbol{\Psi}_X$, $\mathbf{A}_1$, σ_{cd} and $p\ge c\ge d\ge 1$ in (41) with $\boldsymbol{\Psi}_Y$, $\mathbf{B}_1$, $\sigma_{p+c,p+d}$ and $q\ge c\ge d\ge 1$, respectively.

Acknowledgments The author is indebted to the comments and suggestions of the reviewers for the improvement of the manuscripts. This work was partially supported by Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. New York: Wiley.
Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis. *Journal of Multivariate Analysis*, *70*, 1–29.
Bai, Z. D., He, X. (2004). A chi-square test for dimensionality with non-Gaussian data. *Journal of Multivariate Analysis*, *88*, 109–117.
Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society*, *34*, 33–40.
Boik, R. J. (1998). A local parameterization of orthogonal and semi-orthogonal matrices with applications. *Journal of Multivariate Analysis*, *67*, 244–276.
Browne, M. W. (1979). The maximum likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*, 75–86.
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.
Constantine, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Annals of Mathematical Statistics*, *34*, 1270–1285.
Eaton, M. L., Tyler, D. (1994). The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. *Journal of Multivariate Analysis*, *50*, 238–264.
Fang, C., Krishnaiah, P. R. (1982). Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations. *Journal of Multivariate Analysis*, *12*, 39–63.
Fujikoshi, Y. (1977). Asymptotic expansions for the distributions of some multivariate tests. In P. R. Krishnaiah (Ed.), *Multivariate analysis-IV* (pp. 55–71). New York: North-Holland.
Fujikoshi, Y. (1978). Asymptotic expansions for the distributions of some functions of the latent roots of matrices in three situations. *Journal of Multivariate Analysis*, *8*, 63–72.
Hall, P. (1992a). The bootstrap and Edgeworth expansion. New York: Springer. (corrected printing, 1997).
Hall, P. (1992b). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, B*, *54*, 221–228.
Hayakawa, T. (1987). Normalizing and variance stabilizing transformations of multivariate statistics under an elliptical population. *Annals of Institute of Statistical Mathematics*, *39, A*, 299–306.
Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 321–377.
Hsu, P. L. (1941). On the limiting distribution of the canonical correlations. *Biometrika*, *32*, 38–45.
Hu, J., He, X. (2007). Enhanced quantile normalization of microarray data to reduce loss of information in gene expression profiles. *Biometrics*, *63*, 50–59.
Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis. *Biometrika*, *68*, 647–651.
Lawley, D. N. (1959). Tests of significance in canonical analysis. *Biometrika*, *46*, 59–66.
Liu, L., Hawkins, D. M., Ghosh, S., Young, S. S. (2003). Robust singular value decomposition analysis of microarray data. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 13167–13172.
Magnus, J. R., Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics* (Rev. Ed.). New York: Wiley.
Muirhead, R. J., Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations. *Biometrika*, *67*, 31–43.
Ogasawara, H. (1986). Transformation of the factor loadings in inter-battery factor analysis. *Japanese Journal of Psychology*, *57*, 75–82 (in Japanese).
Ogasawara, H. (2005a). *Asymptotic expansion of the distributions of the parameter estimators in structural equation modeling*. Paper presented at the international conference on the future of statistical theory, practice and education, January. Hyderabad, India.
Ogasawara, H. (2005b). Asymptotic robustness of the asymptotic biases in structural equation modeling. *Computational Statistics and Data Analysis*. *49*, 771–783.

Ogasawara, H. (2006). Asymptotic expansion of the sample correlation coefficient under nonnormality. *Computational Statistics and Data Analysis*, *50*, 891–910.
Ogasawara, H. (2007a). Asymptotic expansion of the distributions of the estimators in factor analysis under nonnormality. *British Journal of Mathematical and Statistical Psychology*, *60*, 395–420.
Ogasawara, H. (2007b). Asymptotic expansions of the distributions of estimators in canonical correlation analysis under nonnormality. *Journal of Multivariate Analysis*, *98*, 1726–1750.
Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd Ed.). New York: Wiley.
Rao, C. R. (1979). Separation theorems for singular values of matrices and their applications in multivariate analysis. *Journal of Multivariate Analysis*, *9*, 362–377.
Steiger, J. H., Browne, M. W. (1984). The comparison of interdependent correlations between optimal linear composite. *Psychometrika, 49*, 11–24.
Sugiura, N. (1976). Asymptotic expansions of the distributions of the canonical correlations. In: *Commemoration volume of founding of faculty of integrated arts and sciences*. (pp. 141–163), Hiroshima: Hiroshima University
Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika, 23*, 111–136.
Wegelin, J. A., Richardson, T. S., Ragozin, D. L. (2001). Rank-one latent models for cross-covariance. *Technical report no. 391,* University of Washington, Department of Statistics, Seattle, WA.
Wegelin, J. A., Packer, A., Richardson, T. S. (2002). Rank-*r* latent models for cross-covariance. *Technical report no. 411*, University of Washington, Department of Statistics, Seattle, WA.
Wegelin, J. A., Packer, A., Richardson, T. S. (2006). Latent models for cross-covariance. *Journal of Multivariate Analysis*, *97*, 79–102.