On Covariance Estimators of Factor Loadings in Factor Analysis

Kentaro Hayashi and Pranab Kumar Sen

University of California, Los Angeles; and University of North Carolina at Chapel Hill

Received July 24, 1996; revised July 29, 1997

We report a matrix expression for the covariance matrix of MLEs of factor loadings in factor analysis. We then derive the analytical formula for covariance matrix of the covariance estimators of MLEs of factor loadings by obtaining the matrix of partial derivatives, which maps the differential of sample covariance matrix (in vector form) into the differential of the covariance estimators. © 1998 Academic Press

AMS 1991 subject classification: primary 62H25; secondary 62F12

Key words and phrases: asymptotic normality, Kronecker product, maximum likelihood estimator, vec operator.

1. INTRODUCTION

The purpose of this paper is to report a matrix expression for the covariance matrix of MLEs of factor loadings in factor analysis and the formula for covariance matrix for the covariance estimators of MLEs of factor loadings. For the coordinatewise expression for covariance matrix of MLEs of factor loadings, see Lawley and Maxwell [6], with some corrections by Jennrich and Thayer [5].

Consider a $p \times 1$ random vector of observations $x_i$ ($i = 1, \ldots, n$) with $E(x_i) = 0$. Let $A$ be a $p \times m$ matrix of factor loadings, $f_i$ be a $m \times 1$ vector of factor scores, and $e_i$ be a $p \times 1$ vector of unique factors. Then the factor analysis model is written as $x_i = Af_i + e_i$, $i = 1, \ldots, n$, with the assumptions $E(f_i) = 0; E(e_i) = 0; \text{Cov}(f_i, e_i) = E(f_i e_i) = 0$; and $\text{Cov}(e_i) = E(e_i e_i') = \Psi$, where $\Psi$ is a positive definite diagonal matrix. Assuming the orthogonal model (where the factors are uncorrelated), the covariance matrix of $x_i$ is expressed as $\Sigma = AA' + \Psi$.

In maximum likelihood estimation, we further assume that $x_i$'s are random samples from the normal population $N_p(0, \Sigma)$. To remove the indeterminacy regarding orthogonal rotations, the additional side condition that $\Lambda' \Psi^{-1} \Lambda$ is diagonal is typically employed. By differentiating the log likelihood function with respect to $A$ and $\Psi$ and setting them to a null matrix, with further algebra we obtain the two equations from which MLE $\hat{A}_n$ of
2. COVARIANCE MATRIX OF MLEs OF FACTOR LOADINGS

Define a pm \times pm matrix \(V = (n') \text{Cov}(\hat{\lambda}_n)\), where \(n' = n - 1\), \(\hat{\lambda}_n = \text{vec}(\hat{A}_n)\), and vec(\(\hat{A}_n\)) denotes the pm-dimensional column vector listing \(m\) columns of \(\hat{A}_n\) starting from the first column. Then the formula for \(V\) is given by

\[
V = A + 2EF, \quad (1)
\]

where the matrix \(A\) is expressed as

\[
A = A_1 - A_2
\]

\[
= \{ M \otimes \Sigma + (M \otimes AM)(\text{diag}(K_{mm}\gamma^*))\}(I_m \otimes A') - \{ A_{21} \# A_{21}' \# A_{22} \}, \quad (2)
\]

with

\[
A_{21} = (1_m \otimes A \otimes 1_{p'}) - (\text{diag}(\hat{A}))\(I_m \otimes 1_p\), \quad (3)
\]

\[
A_{22} = (\Theta(\Theta^* - I_m) \otimes 1_p)\gamma^*, \quad (4)
\]

\[
\gamma^* = \text{vec}(\Theta - I_m)^2(\Theta^* - I_m) - I_m\gamma^* + (1/2)I_m, \quad (5)
\]

\[
M = \Theta(\Theta - I_m)^{-1}, \quad (6)
\]

\[
\Theta^* = \text{vec}^{-1}(I_m \otimes \Theta - \Theta \otimes I_m + \text{diag}(\text{vec}(I_m)))^{-1}1_m, \quad (7)
\]

(i.e., vec(\(\Theta^*) = (I_m \otimes \Theta - \Theta \otimes I_m + \text{diag}(\text{vec}(I_m)))^{-1}1_m\), where \(\otimes\) and \(\#\) denote the Kronecker and the Hadamard products, respectively; \(K_{mm}\) is a \(m^2 \times m^2\) matrix defined such that \(K_{mm}\text{vec}(G) = \text{vec}(G')\) for any \(m \times m\) matrix \(G\); diag(\(z\)) denotes the diagonal matrix whose diagonal elements are vector \(z\); and \(\Theta = A'\Psi^{-1/2}\) is the diagonal matrix whose elements are the first \(m\) largest eigenvalues of \(\Psi^{-1/2}S_n\Psi^{-1/2}\). Next, the matrix \(B\) is given by

\[
B = -B_1 \# B_2
\]

\[
= -\{ \Psi^{-2}A(\Theta - I_m)^{-1} \otimes 1_p' \}
\]

\[
\# \{ 1_m \otimes \Psi + (1_m' \Theta \otimes A)(\text{diag}(K_{mm}\Theta^*))(I_m \otimes A') \}, \quad (8)
\]
where
\[ \theta^* = (I_m \otimes \Theta - \Theta \otimes I_m)^{-1} \mathbb{1}_m; \]
(= \Theta^* \mathbb{1}_m). \hfill (9)

The matrix expression for \( E \) has already been given by Lawley and Maxwell [6] and is
\[ E = (\Phi \otimes \Phi)^{-1}, \hfill (10) \]
where \( \Phi = \Psi^{-1} - \Psi^{-1} \Lambda (\Theta - I_m)^{-1} A' \Psi^{-1}. \)

3. ASYMPTOTIC NORMALITY OF COVARIANCE ESTIMATORS

Anderson and Rubin [1] established the asymptotic normality of MLEs of factor loadings and unique variances. Their theorem holds under the following two assumptions: (i) \( \Phi \otimes \Phi \) is nonsingular, where \( \Phi \) is defined in (10); and (ii) \( \Theta \) has ordered, distinct, diagonal elements. We state a theorem regarding the asymptotic normality of covariance estimators of MLEs of factor loadings which holds under the same set of assumptions. Let \( \text{vech}(D) \) denote the column vector consisting of elements on and below the diagonal of the square matrix \( D \), starting with the first column (cf., e.g., Searle [8]).

**Theorem 3.1.** Let the above assumptions (i) and (ii) hold. Let \( e = n' \cdot \text{vech}(\text{Cov}(\hat{\theta}_n)) \) and let \( \hat{v}_n = n' \cdot \text{vech}(\tilde{\text{Cov}}(\hat{\theta}_n)) \) be the estimator of \( v \). Then, as \( n \to \infty, \hat{v} \to v \) a.s., and \( \sqrt{n}(\hat{v}_n - v) \) is asymptotically multinormal, though singular.

**Proof.** By virtue of assumptions (i) and (ii), \( v \) is free from \( n, \hat{v}_n \) is a continuous function of \( S_n \), and \( v \) is the same function of \( \Sigma \). \( S_n \) being a \( U \)-statistic (matrix), by virtue of the reverse martingale property, \( S_n \to \Sigma \) a.s. as \( n \to \infty \) (see, e.g., Sen and Singer [9]). The same \( U \)-statistic characterization of \( S_n \) leads to asymptotic multinormality of \( \sqrt{n}(\hat{v}_n - v) \). Furthermore, we can express \( \sqrt{n}(\hat{v}_n - v) \) as
\[ \sqrt{n}(\hat{v}_n - v) = \left( \frac{\partial v}{\partial \sigma} \right) \text{vech}(\sqrt{n}(S_n - \Sigma)) + o_p(1), \hfill (11) \]
where \( \frac{\partial v}{\partial \sigma} \) is the matrix of partial derivatives of \( \hat{v}_n \) with respect to \( s_n = \text{vech}(S_n) \) evaluated at \( S_n = \Sigma \), and is a nonstochastic matrix depending only on the population parameters. Thus the asymptotic normality of \( \sqrt{n}(\hat{v}_n - v) \) follows from differentiability of \( \hat{v}_n \) in (11) and the asymptotic normality of \( \sqrt{n}(s_n - \Sigma) \). Q.E.D
4. COVARIANCE MATRIX FOR COVARIANCE ESTIMATORS

In this section we give an explicit expression for the asymptotic covariance matrix for \( \hat{\varepsilon}_n \). The asymptotic covariance matrix for \( \hat{\varepsilon}_n \) is given by

\[
\text{Cov}(\hat{\varepsilon}_n) = \left( \frac{\partial \hat{v}}{\partial \hat{\sigma}} \right) (\text{Cov}(s_n)) \left( \frac{\partial \hat{v}}{\partial \hat{\sigma}} \right)',
\]

(12)

where \( \frac{\partial \hat{v}}{\partial \hat{\sigma}} \) is a \((1/2) \, pm(\, pm + 1) \times (1/2) \, p(p + 1)\) matrix of partial derivatives connecting the differential of \( \hat{v}_n \) and the differential of \( s_n \) such that \( \hat{v}_n = (\hat{v} / \hat{\sigma}) \, (d S_n) \). Note also that \( \hat{v} / \hat{\sigma} \) in (12) is evaluated at \( S_n = \Sigma \), and we will omit it thereafter for notational simplicity. In the case of normal sampling, the formula for \( \text{Cov}(s_n) \) is

\[
(1/2) \cdot H_p \cdot I_p \cdot G_p \cdot H_p'
\]

(13)

where \( K_{pm} \) is defined such that \( K_{pm} \, \text{vec}(S) = \text{vec}(S') \) for any \( p \times p \) matrix \( S \); \( H_p = (G_p' \, G_p)^{-1} \) \( G_p' \) and \( G_p \) is defined such that \( \text{vec}(S) = G_p \, \text{vech}(S) \) for any \( p \times p \) matrix \( S \).

4.1. Matrix of Partial Derivatives of \( \hat{\varepsilon}_n \)

Our main task remaining is to report the expression for \( \hat{v} / \hat{\sigma} \) in (12). The actual derivation of the expression for \( \hat{v} / \hat{\sigma} \) is given in Hayashi and Sen [2]. We simply show the final results here. First let \( a = \text{vech}(A) \), \( b = \text{vec}(B) \), and \( e = \text{vech}(E) \). Then by the chain rule, the expression for \( \hat{v} / \hat{\sigma} \) is

\[
\hat{v} / \hat{\sigma} = \left( \frac{\partial v}{\partial a} \right) + \left( \frac{\partial v}{\partial b} \right) \left( \frac{\partial b}{\partial \sigma} \right) + \left( \frac{\partial v}{\partial e} \right) \left( \frac{\partial e}{\partial \sigma} \right),
\]

(13)

where

\[
\frac{\partial v}{\partial a} = I_{pm \times (pm + 1)/2},
\]

\[
\frac{\partial v}{\partial b} = 2H_{pm}(I_p \cdot \pm + K_{pm, \cdot \cdot}) \left( I_p \otimes B' \right) E',
\]

(14)

\[
\frac{\partial v}{\partial e} = 2H_{pm}(B' \otimes B') \, G_p,
\]

and \( K_{pm} \) is defined such that \( \text{vec}(B') = K_{pm, \cdot \cdot} \, \text{vec}(B) \); and \( H_{pm} \) is defined such that \( H_{pm} = (G_p \cdot G_p)^{-1} \) \( G_p \). Hence, \( G_p \) is defined such that \( \text{vec}(R) = G_p \, \text{vech}(R) \) for any \( pm \times pm \) matrix \( R \).
4.2. Matrices of Partial Derivatives of $\dot{a}, \dot{b}, \text{ and } \dot{e}$

In Section 2, we gave the matrix expressions for $A, B, \text{ and } E$. Now, we give the expressions for the matrices of partial derivatives of these matrices. Let $\dot{a}_1 = \text{vec}(A_1), \dot{a}_2 = \text{vec}(A_2), \dot{b}_1 = \text{vec}(B_1), \dot{b}_2 = \text{vec}(B_2), \dot{\phi} = \text{vec}(\Phi), \mu = \text{vec}(M), \theta = \text{vec}(\Theta), \lambda = \text{vec}(A), \text{ and } \Psi = \text{vec}(\Psi)$. First, the matrix of partial derivatives of $\dot{a}$ is obtained immediately from Eq. (2) and is of the form

$$\frac{\partial \dot{a}}{\partial \sigma} = \frac{\partial \dot{a}_1}{\partial \sigma} - \frac{\partial \dot{a}_2}{\partial \sigma},$$

where the matrix of partial derivatives of $\dot{a}_i$ is

$$\frac{\partial \dot{a}_i}{\partial \sigma} = A_{1(i)} \left( \frac{\partial \mu}{\partial \sigma} \right) + A_{1(2)} \left( \frac{\partial \phi}{\partial \sigma} \right) + A_{1(3)} \left( \frac{\partial \lambda}{\partial \sigma} \right) + A_{1(4)},$$

with

$$A_{1(1)} = H_{pm}(I_m \otimes K_{pm} \otimes I_p)(I_m \otimes \text{vec}(\Sigma)) + ((I_m \otimes A)(\text{diag}(K_{mn} \gamma^*)) \otimes I_m)(I_m \otimes K_{mm} \otimes I_p) \times (I_m \otimes \text{vec}(AM) + (\mu \otimes I_m)(I_m \otimes A))/1,$n

$$A_{1(2)} = H_{pm}(I_m \otimes A \otimes M \otimes AM) K_{mm} K_{mn},$$

$$A_{1(3)} = H_{pm}((I_m \otimes A)(\text{diag}(K_{mn} \gamma^*)) \otimes I_m)(I_m \otimes K_{mm} \otimes I_p) \times (\mu \otimes I_m)(M \otimes I_p) + (I_m \otimes (M \otimes AM)(\text{diag}(K_{nm} \gamma^*))) \times (I_m \otimes K_{pm} \otimes I_m)(I_m \otimes I_p) K_{pm}/1,$n

$$A_{1(4)} = H_{pm}(I_m \otimes K_{pm} \otimes I_p)(\mu \otimes I_p) G_p,$n

and $K_{mm} = \sum_{i=1}^{m^2} (J_m^2, J_m^3, i \otimes J_m^2, i)$ with a $m^2$-dimensional unit vector $J_m^2, i$ whose $i$th element is 1 and the rest are 0’s. Next, the matrix of partial derivatives of $\dot{b}_2$ is

$$\frac{\partial \dot{b}_2}{\partial \sigma} = A_{2(1)} A_{21(1)} \left( \frac{\partial \lambda}{\partial \sigma} \right) + A_{2(2)} A_{22(1)} \left( \frac{\partial \theta}{\partial \sigma} \right),$$

where

$$A_{2(1)} = (\text{diag}(H_{pm} K_{pm, pm} a_{21} \# a_{22})) H_{pm} + (\text{diag}(H_{pm} a_{21} \# a_{22})) H_{pm} K_{pm, pm},$$

$$A_{2(2)} = \text{diag}(H_{pm} a_{21} \# H_{pm} K_{pm, pm} a_{21}),$$
Next, from (8), the matrix of partial derivatives of $b$ is derived as
\[ \frac{\partial b}{\partial \sigma^2} = -(\text{diag}(b_3)) \left( \frac{\partial b_1}{\partial \sigma^2} \right) - (\text{diag}(b_2)) \left( \frac{\partial b_2}{\partial \sigma^2} \right), \] (18)

where
\[ \frac{\partial b_1}{\partial \sigma^2} = B_{1(1)} \left( \frac{\partial \theta}{\partial \sigma^2} \right) + B_{1(2)} \left( \frac{\partial \lambda}{\partial \sigma^2} \right) + B_{1(3)} \left( \frac{\partial \psi}{\partial \sigma^2} \right), \] (19)
\[ \frac{\partial b_2}{\partial \sigma^2} = B_{2(1)} \left( \frac{\partial \theta}{\partial \sigma^2} \right) + B_{2(2)} \left( \frac{\partial \lambda}{\partial \sigma^2} \right) + B_{2(3)} \left( \frac{\partial \psi}{\partial \sigma^2} \right), \] (20)

with
\[ B_{1(1)} = -(I_m \otimes K_{pp})(I_{pq} \otimes 1_p)(((\Theta - I_m)^{-1} \otimes \Psi^{-2})A(I \otimes -I_m)^{-1})], \]
\[ B_{1(2)} = (I_m \otimes K_{pp})(I_{pq} \otimes 1_p)((\Theta - I_m)^{-1} \otimes \Psi^{-2}), \]
\[ B_{1(3)} = -(I_m \otimes K_{pp})(I_{pq} \otimes 1_p)((\Theta - I_m)^{-1} \otimes A^\top I_p) \times (\Psi^{-1} \otimes \Psi^{-2} + \Psi^{-2} \otimes \Psi^{-1}), \]
\[ B_{2(1)} = \left\{ (I_m \otimes A)(\text{vec}(K_{mn}^\top \theta^\top)) \otimes I_p \right\} (I_m \otimes \lambda I_m) + (I_m \otimes A \otimes I_m \otimes A) T_{mn}, \]
\[ B_{2(2)} = \left\{ (I_m \otimes A)(\text{diag}(K_{mn} \theta^\top)) \otimes I_p \right\} (\Theta I_m \otimes I_p) + \left\{ I_m \otimes ((I_m \otimes A)(\text{vec}(K_{mn} \theta^\top))) \right\} \times (I_m \otimes K_{mn} \otimes I_m)(\text{vec}(I_m) \otimes I_{pq} I_{pq}), \]
\[ B_{2(3)} = I_m \otimes I_p, \]
\[ T_{mn} = K_{mn}^\top I_{mn}(1_m \otimes \Theta \otimes \Theta^*) \times \left\{ 2K_{mn}^\top I_m \otimes (2\Theta - I_m) \right\} + (I_m \otimes K_{mn} \otimes I_m) \times (I_m \otimes \text{vec}(I_m) - \text{vec}(I_m) \otimes I_{pq}). \]
Finally, the matrix of partial derivatives of $\hat{e}$ is obtained from (10) and is

$$\frac{\partial e}{\partial \sigma} = -2H_p(E \otimes E)G_p(\text{diag}(\phi)) \left( \frac{\partial \phi}{\partial \sigma} \right),$$

(21)

where

$$\frac{\partial \phi}{\partial \sigma} = P_1 \left( \frac{\partial \theta}{\partial \sigma} \right) + P_2 \left( \frac{\partial \lambda}{\partial \sigma} \right) + P_3 \left( \frac{\partial \psi}{\partial \sigma} \right),$$

(22)

with

$$P_1 = H_p(\Psi^{-1}A(\Theta - I_m)^{-1} \otimes \Psi^{-1}A(\Theta - I_m)^{-1}),$$

$$P_2 = -H_p(I_p + K_p)(\Psi^{-1}A(\Theta - I_m)^{-1} \otimes \Psi^{-1}),$$

$$P_3 = H_p(\Psi^{-1}A(\Theta - I_m)^{-1} A'\Psi^{-1} \otimes \Psi^{-1}$$

$$+ \Psi^{-1} \otimes \Psi^{-1}A(\Theta - I_m)^{-1} A'\Psi^{-1} - \Psi^{-1} \otimes \Psi^{-1}).$$

4.3. Matrices of Partial Derivatives of $\hat{\gamma}^*, \hat{\mu}, \hat{\theta}, \hat{\lambda}_m$, and $\hat{\psi}_n$

The matrices of partial derivatives in Section 4.2 are expressed in terms of the matrices of partial derivatives of $\hat{\gamma}^*, \hat{\mu}, \hat{\theta}, \hat{\lambda}_m$, and $\hat{\psi}_n$. Thus we need to obtain the expressions for these. First, the matrices of partial derivatives of $\hat{\gamma}^*, \hat{\mu}, \hat{\theta}$ are obtained from (5), (6), and $\Theta = A'\Psi^{-1}A + I_m$, respectively, and are as follows:

$$\frac{\partial \gamma^*}{\partial \sigma} = (2(\Theta^* - I_m) \otimes (\Theta - I_m) + (I_m \otimes (\Theta - I_m)^2) T_{\mu m}^*) \left( \frac{\partial \theta}{\partial \sigma} \right),$$

(23)

$$\frac{\partial \mu}{\partial \sigma} = ((\Theta - I_m)^{-1} \otimes I_m - (\Theta - I_m)^{-1} \otimes \Theta(\Theta - I_m)^{-1}) \left( \frac{\partial \theta}{\partial \sigma} \right),$$

(24)

$$\frac{\partial \theta}{\partial \sigma} = (I_m + K_mm)(I_m \otimes A'\Psi^{-1}) \left( \frac{\partial \lambda}{\partial \sigma} \right) - (A'\Psi^{-1} \otimes A'\Psi^{-1}) \left( \frac{\partial \psi}{\partial \sigma} \right).$$

(25)

Jennrich and Clarkson's [4] Eqs. (23), (24), (26), and (27) give the formulas that connect the differential of $\hat{\lambda}_m$ with the differentials of $\hat{\gamma}^*, \hat{\mu}, \hat{\theta}$, and $\hat{\psi}_n$, and the differential of $\text{vdg}(\Psi_p)$ with the differential of $S_n$, where $\text{vdg}(\Psi_p)$ denotes the diagonal elements of $\Psi_p$ arranged as a vector. From these equations, we derive the matrices of partial derivatives of $\hat{\lambda}_m$ and $\hat{\psi}_n$, which are

$$\frac{\partial \lambda}{\partial \sigma} = (W^{-1} \otimes I_m) \left( G_p - \left( \frac{\partial \psi}{\partial \sigma} \right) \right) - (W^{-1} \otimes I_A)(Y_1 + Y_2),$$

(26)

$$\frac{\partial \psi}{\partial \sigma} = K_p(Q \# Q) \otimes I_p)^{-1} K_p(Q \otimes Q) G_p,$$

(27)
where

\[ W = A' \Sigma^{-1} A, \quad Z = A' \Sigma^{-1}, \quad Q = I_p - AW^{-1} Z, \quad K_p \equiv \sum_{i=1}^{p} (J_{p,i} J_{p,i}^T \otimes J_{p,i}), \]

and

\[ Y_1 = \left( \frac{1}{2} \right) \left( \text{diag}(\text{vec}(W^{-1}))(Z \otimes Z) \left( G_p - \frac{\partial \psi}{\partial \sigma'} \right) \right), \quad (28) \]

\[ Y_2 = (I_m \otimes W - W \otimes I_m)^+ \left\{ (I_m - W)(Z \otimes Z) G_p - (Z \otimes Z) \left( \frac{\partial \psi}{\partial \sigma'} \right) \right\}, \quad (29) \]

with the Moore-Penrose inverse. Essentially the identical expression to (27), using \( \Phi \) instead of \( Q \), is obtained by Ihara and Kano [3]. [Note that an alternative matrix formula for \( \text{Cov}(\psi_{n}) \) is given by \( (\Phi_{n})\text{Cov}(\psi_{n})\Phi_{n}' \).]

ACKNOWLEDGMENTS

The authors are thankful to the Editor and the anonymous referee for their valuable suggestions. This paper is based on part of the first author’s dissertation research at the University of North Carolina, and he thanks members of his dissertation committee, particularly Dr. Yiu-Fai Yung, for introducing some properties of vec operator and Kronecker Product.

REFERENCES


