PSYCHOMETRIKA—VOL. 55, NO. 2, 277–291
JUNE 1990

NONITERATIVE ESTIMATION AND THE CHOICE OF THE NUMBER OF FACTORS IN EXPLORATORY FACTOR ANALYSIS

YUTAKA KANO

OSAKA UNIVERSITY

Based on the usual factor analysis model, this paper investigates the relationship between improper solutions and the number of factors, and discusses the properties of the noniterative estimation method of Ihara and Kano in exploratory factor analysis. The consistency of the Ihara and Kano estimator is shown to hold even for an overestimated number of factors, which provides a theoretical basis for the rare occurrence of improper solutions and for a new method of choosing the number of factors. The comparative study of their estimator and that based on maximum likelihood is carried out by a Monte Carlo experiment.

Key words: exploratory factor analysis, identifiability, improper solutions, maximum likelihood estimator, noniterative estimator, number of factors, quasi-loadings.

1. Introduction and Examples

Factor analysis can be characterized as a multivariate technique for analyzing the internal relationships among a set of variables. Formally, a $p \times 1$ random vector $\mathbf{x} = (X_1, \ldots, X_p)'$ of observations is represented in the form

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u},$$

where Λ is a $p \times k$ matrix of factor loadings, $\mathbf{f}$ and $\mathbf{u}$ are $k \times 1$ and $p \times 1$ random vectors of common and unique factors, respectively, and $k(<p)$ is the number of factors. It is usually assumed that $\mathbf{f}$ and $\mathbf{u}$ are uncorrelated and the covariance matrix Ψ of $\mathbf{u}$ is diagonal and positive definite. These assumptions lead to the fundamental equation of factor analysis:

$$\text{Cov}\,(\mathbf{x}) = \Lambda\Phi\Lambda' + \Psi(= \Sigma, \text{ say}),$$

where Φ is the covariance matrix of the common factors (see, e.g., Lawley & Maxwell, 1971, p. 13). In this paper we discuss the problem of estimation in orthogonal exploratory factor analysis: that is, Φ is assumed to be the identity matrix and no prior information exists about Λ, Ψ, and k. Given orthogonality, the fundamental equation becomes

$$\Sigma = \Lambda\Lambda' + \Psi, \tag{1}$$

where Λ and Ψ are free parameters to be estimated, and k is unknown.

When the number of factors is given, an estimator $(\hat{\Lambda}, \hat{\Psi})$ is usually determined by the solution minimizing a suitable discrepancy function, $F(\mathbf{S}, \Sigma)$, which gives an indication of the difference between the sample covariance matrix $\mathbf{S}$ and the model $\Sigma = \Lambda\Lambda' + \Psi$ (see, e.g., Browne, 1982). Typical examples are estimators based on maximum

The author would like to express his thanks to Masashi Okamoto and Masamori Ihara for helpful comments and to the editor and referees for critically reading the earlier versions and making many valuable suggestions. He also thanks Shigeo Aki for his comments on physical random numbers.

Requests for reprints should be sent to Yutaka Kano, Department of Applied Mathematics, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, JAPAN.

0033-3123/90/0600-9102 $00.75/0

likelihood (MLE, say) and generalized least squares. Unfortunately, solutions generated by such estimation procedures cannot be represented as explicit functions of S because the derivatives of $F(\mathbf{S}, \boldsymbol{\Sigma})$ with respect to $(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$ are nonlinear and complicated; consequently, estimates are usually calculated by an iterative method (e.g., Jennrich & Robinson, 1969; Jöreskog, 1967). In spite of the efforts that have been made to devise good estimation techniques, we often meet the following difficulties: (I) Some diagonal elements of $\hat{\boldsymbol{\Psi}}$ are not positive, which is called an improper solution or Heywood case, (II) the iteration does not terminate, and (III) solutions may depend on the choice of initial estimates. It is well-known that these difficulties frequently arise when we assume a large number of factors.

One of the main purposes of the present paper is to relate some of the difficulties commonly encountered in iterative estimation to the choice of the number of factors, another important and longstanding problem in factor analysis. Many methods of finding suitable values of k have been proposed (e.g., Hakstian & Muller, 1973; Hakstian, Rogers & Cattell, 1982); at least two have been justified from the statistical point of view, namely the likelihood ratio (LR) test (Lawley, 1940) and the AIC (Akaike, 1974). It has been reported, however, that all methods often fail to work well (see, e.g., Akaike, 1987; Jöreskog, 1978; Sato, 1987). To illustrate, we shall use Davis' data (1944; $p = 9$, $n = 421$), which have been reanalyzed by Rao (1955) and by Jennrich and Robinson (1969), who showed that Rao's solution does not correspond to the MLE. In Davis' data, the iterative process calculating the MLE for $k = 2$ does not converge if the restriction that $\psi_i \geqq 0$ for all i is dropped (see Tumura & Sato, 1981). If we impose this restriction to calculate the MLE, then for $k = 1$, the MLE is proper and the LR test is highly significant (a χ^2 value of 54.95 for 27 df; $p < .005$); for $k = 2$, the LR test suggests acceptance of the model but the solution is improper. The AIC also indicates $k = 2$, where the MLE's generated for $k > 1$ are all improper. To illustrate the difficulty labeled as (III), we calculated the MLE's ($k = 1, 2$) starting from ten different initial estimates $I_0, I_1, \ldots, I_9$ that were constructed as follows: I_0 is the initial estimate recommended by Jöreskog (1967) written as $\hat{\psi}_i = (1 - k/2p)/s^{ii}$, where s^{ij} stands for the (i, j) element of S^{-1}, and I_k was formed by replacing the k-th element of I_0 by zero ($k = 1, 2, \ldots, 9$). The iteration was terminated if either the root mean square of the gradient vector or of $\Delta\boldsymbol{\Psi} = (\hat{\boldsymbol{\Psi}}_{(t+1)} - \hat{\boldsymbol{\Psi}}_{(t)})$ is less than 0.00001 as in Lee and Jennrich (1979); here, $\hat{\boldsymbol{\Psi}}_{(t)}$ denotes the t-th solution in the iterative process.

The results are shown in Table 1. For $k = 2$, all solutions were improper and seven different solutions were obtained (the solutions starting from I_0, I_1, I_5 and I_6 are the same), whereas all solutions for $k = 1$ were identical and proper.

To explain why such difficulties may be encountered, assume that a matrix $\boldsymbol{\Sigma}$ is decomposed as in (1), and many different common factor decompositions with $k + 1$ factors exist because the following equalities hold:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$$

$$= \begin{bmatrix} & c \\ & 0 \\ \boldsymbol{\Lambda} & 0 \\ & \vdots \\ & \vdots \\ & 0 \end{bmatrix} \begin{bmatrix} & c \\ & 0 \\ \boldsymbol{\Lambda} & 0 \\ & \vdots \\ & \vdots \\ & 0 \end{bmatrix}' + \begin{bmatrix} \psi_1 - c^2 & & & & \\ & \psi_2 & & & \\ & & \psi_3 & & \\ & & & \ddots & \\ & & & & \psi_p \end{bmatrix}$$

$$= \begin{bmatrix} & 0 \\ & c \\ \Lambda & 0 \\ & \vdots \\ & \vdots \\ & 0 \end{bmatrix} \begin{bmatrix} & 0 \\ & c \\ \Lambda & 0 \\ & \vdots \\ & \vdots \\ & 0 \end{bmatrix}' + \begin{bmatrix} \psi_1 & & & & \\ & \psi_2 - c^2 & & & \\ & & \psi_3 & & \\ & & & \ddots & \\ & & & & \psi_p \end{bmatrix}$$

.
.
.

Here c is any number, and thus, the factor loadings and unique variances for $k + 1$ factors are not identified. This nonidentifiability can cause a number of serious problems, if we intend to estimate the parameters under the assumption of $k + 1$ factors. For instance, c may become so large that $\psi_i - c^2$ is negative; c may oscillate or tend to infinity so that iteration is not terminated; and the position of the ψ_i where c appears may change if different initial estimates are chosen. For convenience, we will call such a loading, c, a quasi-loading; also, when a column vector of $\hat{\Lambda}$ consists of just one element with a large value and the others have rather small value, the element will be also called a quasi-loading. The presence of quasi-loadings may be one of the possible causes of the three difficulties mentioned earlier; also this could be one of reasons for the failure of the AIC, which requires the MLE to be consistent.

Considering Davis' data and observing the results in Table 1, we see that the first factor loading vectors with $k = 2$ are almost the same and are close to the factor loading vector with $k = 1$. Each of the second vectors includes only one element with a large value; the others are considerably smaller ($\hat{\Lambda}$'s were obtained by Sato's, 1987, method of rotation). Thus, the elements with a large value can be considered quasi-loadings, and the one-factor solution would be appropriate in exploratory factor analysis.

Akaike (1987) suggested the existence of local maxima of the likelihood function with $k = 2$ for Davis' data. In fact, two different improper solutions for the 1st and 2nd variables were reported by Mattsson, Olssen, and Rosén (1966) and Akaike (1987), respectively. These improper solutions are very close to the corresponding estimates in Table 1, which suggests that these solutions are due to the appearance of quasi-loadings. Although Boomsma (1985) stated that different initial values have little effect in estimation, this would be true only when the number of factors is properly chosen. One might choose as the MLE the solution with the maximum likelihood of the seven shown in Table 1, but quasi-loadings are loadings not for common but for unique factors, and it is meaningless to analyze estimates that include them.

Other examples illustrating quasi-loadings could be given in a similar manner. For instance, the MLE with $k = 3$ for the data of Eight Physical Variables (Harman, 1976, p. 22) has the following form:

$$\hat{\Lambda}' = \begin{bmatrix} .59 & .53 & .50 & .55 & .94 & .79 & .75 & .69 \\ .72 & .77 & .73 & .73 & -.12 & -.12 & -.16 & -.01 \\ -.07 & .35 & .15 & -.09 & -.10 & -.01 & .00 & .16 \end{bmatrix}$$

and

$$\hat{\Psi} = \text{diag}\,(.13 \quad .00 \quad .21 \quad .15 \quad .09 \quad .36 \quad .41 \quad .51),$$

where again $\hat{\Lambda}$ was obtained by Sato's method of rotation. Sato (1987) also investigated the treatment of quasi-loadings in Maxwell's data (1961).

TABLE 1

MLE's Starting from Ten Different Initial Estimates in Davis' Data with k=1, 2

(The underlined values represent quasi-loadings as defined in the text.)

Number of Factors	Initial Estimate		Variable 1	2	3	4	5	6	7	8	9
k=1	All	$\hat{\Lambda}$:	.81	.81	.48	.41	.67	.89	.84	.66	.84
		$\hat{\Psi}$:	.34	.34	.77	.83	.55	.20	.29	.57	.30
k=2	I_0 I_1 I_5 I_6	$\hat{\Lambda}$:	.81	.80	.47	.42	.68	.90	.84	.66	.84
			.59	.13	.05	-.10	-.05	-.03	.00	-.04	.01
		$\hat{\Psi}$:	.00	.36	.77	.82	.53	.19	.30	.56	.30
	I_2	$\hat{\Lambda}$:	.80	.81	.48	.41	.68	.90	.84	.66	.84
			.13	.58	-.09	.05	-.04	-.04	-.00	-.03	.00
		$\hat{\Psi}$:	.36	.00	.76	.83	.54	.18	.30	.56	.30
	I_3	$\hat{\Lambda}$:	.81	.82	.47	.41	.67	.89	.84	.66	.84
			.03	-.05	.88	-.04	.02	.01	.03	-.04	.02
		$\hat{\Psi}$:	.34	.33	.00	.83	.55	.20	.30	.56	.30
	I_4	$\hat{\Lambda}$:	.81	.81	.48	.41	.67	.89	.84	.66	.84
			-.06	.03	-.04	.91	.03	-.01	.01	.02	.02
		$\hat{\Psi}$:	.34	.34	.77	.00	.55	.20	.30	.57	.30
	I_7	$\hat{\Lambda}$:	.81	.81	.47	.41	.68	.89	.84	.65	.84
			-.01	-.01	.04	.01	-.03	.02	.54	.08	-.06
		$\hat{\Psi}$:	.34	.34	.78	.83	.54	.21	.00	.58	.29
	I_8	$\hat{\Lambda}$:	.81	.81	.48	.41	.67	.90	.84	.66	.84
			-.04	-.02	-.05	.03	.01	-.03	.05	.75	.04
		$\hat{\Psi}$:	.34	.34	.77	.83	.55	.20	.30	.00	.30
	I_9	$\hat{\Lambda}$:	.81	.81	.48	.41	.68	.89	.85	.65	.84
			-.01	-.01	.02	.03	-.03	.02	-.06	.06	.54
		$\hat{\Psi}$:	.34	.34	.77	.84	.54	.21	.28	.58	.00

In these examples, the estimates $\hat{\Lambda}$ were suitably rotated so that we can easily find the quasi-loadings. In practical situations, however, it is difficult to recognize the presence of quasi-loadings because, as a rule, only the estimate $\hat{\Lambda}$ considered is one that satisfies some restrictions required by the estimation method employed or has been generated by some method of rotation such as varimax. For example, a quasi-loading is not easily detected from the varimax solution ($k = 2$) for Davis' data given in Table 5 (a) in Martin and McDonald (1975).

To find an appropriate number of factors, solutions for several values of k must be calculated and compared; but quasi-loadings become an issue when k is overestimated. New estimation methods preventing the presence of quasi-loadings are needed that are different from the usual methods based on minimizing a discrepancy function, because, as shown above, factor loadings and unique variances are not identified in cases with overestimated values of k. Akaike (1987) made use of Bayesian methods to eliminate the indefiniteness and concluded that $k = 1$ and $k = 2$ are appropriate for the data sets of Davis and the Eight Physical Variables, respectively. Another approach is to use some prior information about (Λ, Ψ) and k in a type of confirmatory factor analysis, where an acceptable two-factor solution is obtained for Davis' data (this was pointed out by one of referees).

In this paper, the properties of the noniterative estimator of Ihara and Kano (1986) are investigated, both analytically and experimentally. Their estimation method is shown not only to produce a simple consistent estimator but also to prevent quasi-loadings. In section 2, the analytical properties of their estimator are developed, and it is shown that the consistency of the estimator holds even if the number of extracted factors is greater than the true number. A new criterion for choosing the number of factors is then proposed based on this consistency property. Section 3 presents an algorithm for the best choice of the Ihara and Kano estimators, and the new estimation procedure is applied to seven real data sets in section 4. The comparative study of the Ihara and Kano estimator and the MLE is carried out using a Monte Carlo experiment in section 5.

2. Ihara and Kano's Estimator and Its Analytical Properties

Let $\Sigma = \Lambda\Lambda' + \Psi$, with Λ of order $p \times k$, and suppose that the parameter (Λ, Ψ) satisfies Anderson and Rubin's sufficient condition for identifiability (see Anderson & Rubin, 1956, Theorem 5.1). This condition will be referred to as the A-R condition hereafter. Note that the A-R condition requires $p \geqq 2k + 1$.

Let m be a positive integer with $k \leqq m \leqq (p - 1)/2$, representing the number of extracted factors, and partition Λ, Ψ and $\underline{\Sigma}$ as follows:

$$\Lambda = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \Lambda_3 \\ \Lambda_4 \end{bmatrix} \begin{matrix} \} \\ \} \\ \} \\ \} \end{matrix} \begin{matrix} 1 \\ m \\ m \\ p - 2m - 1 \end{matrix}, \qquad \Psi = \begin{bmatrix} \psi_1 & & & \mathbf{0} \\ & \Psi_2 & & \\ & & \Psi_3 & \\ \mathbf{0} & & & \Psi_4 \end{bmatrix}, \tag{2}$$

and

$$\Sigma = \begin{bmatrix} \sigma_{11} & & & \\ \sigma_{21} & \Sigma_{22} & \text{symmetric} & \\ \sigma_{31} & \Sigma_{32} & \Sigma_{33} & \\ \sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{bmatrix}.$$

The sample covariance matrix $\mathbf{S}$ is also partitioned in the same fashion. Define

$$\hat{\psi}_1^{(m)} = s_{11} - s_{12}\mathbf{S}_{32}^{-1}s_{31}. \tag{3}$$

(This estimator is not invariant under permutation of the variables $X_2, \ldots, X_p$; a problem to be discussed in section 3.) Estimators of $\psi_2, \ldots, \psi_p$ are constructed in a similar manner to (3), and an overall estimator $\hat{\mathbf{\Psi}}^{(m)}$ is defined by diag $(\hat{\psi}_1^{(m)}, \ldots, \hat{\psi}_p^{(m)})$. Once $\hat{\mathbf{\Psi}}^{(m)}$ is given, an estimator of $\mathbf{\Lambda}$ is constructed as follows: Using spectral decomposition, we have

$$\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)} = \mathbf{PDP}',$$

where $\mathbf{D}$ is a $p \times p$ diagonal matrix of eigenvalues arranged in descending order, and $\mathbf{P}$ is a $p \times p$ orthogonal matrix of eigenvectors. Partitioning

$$\underset{\quad m}{\mathbf{P} = [\mathbf{P}_1}\quad \underset{\quad p-m}{\mathbf{P}_2]} \quad \text{and} \quad \mathbf{D} = \begin{matrix}\begin{bmatrix}\mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2\end{bmatrix} \\ \begin{matrix}\quad m & \quad p-m\end{matrix}\end{matrix},$$

an estimator $\hat{\mathbf{\Lambda}}^{(m)}$ is defined by

$$\hat{\mathbf{\Lambda}}^{(m)} = \mathbf{P}_1\mathbf{D}_1^{1/2}. \tag{4}$$

Note that the squared length of the i-th column vector of $\hat{\mathbf{\Lambda}}^{(m)}$ is equal to the i-th eigenvalue of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)}$. The estimator $(\hat{\mathbf{\Lambda}}^{(m)}, \hat{\mathbf{\Psi}}^{(m)})$ constructed in this way will be called the I-K estimator. Ihara and Kano (1986) showed that the consistency and asymptotic normality of $\hat{\mathbf{\Psi}}^{(m)}$ hold when $m = k$; that is, when the number of factors is known. In this case it can also be proved that the same properties hold for $\hat{\mathbf{\Lambda}}^{(m)}$.

We shall now investigate how the I-K estimator works when the number m is misspecified. We are concerned especially with the case where m is overestimated, that is, $m > k$, because in the usual iterative methods, the quasi-loadings would appear and the difficulties (I), (II) and (III) would arise. Partition $\mathbf{\Lambda}$ as follows:

$$\underset{\quad 1}{\mathbf{\Lambda} = [\boldsymbol{\lambda}_1'}\quad \underset{\quad k}{\boldsymbol{\lambda}_2'}\quad \underset{\quad m}{\mathbf{\Lambda}_3'}\quad \underset{\quad k}{\mathbf{\Lambda}_4'}\quad \underset{\quad m}{\mathbf{\Lambda}_5'}\quad \underset{\quad p-2m-1}{\mathbf{\Lambda}_6']'},$$

with $\mathbf{\Psi}$, $\mathbf{\Sigma}$ and $\mathbf{S}$ partitioned correspondingly. We may assume that $\mathbf{\Lambda}_2$ and $\mathbf{\Lambda}_4$ are nonsingular, and thus, it can be proved in the same way as in Ihara and Kano (1986) that

$$s_{11} - s_{12}\mathbf{S}_{42}^{-1}s_{41} \xrightarrow{\text{P}} \sigma_{11} - \sigma_{12}\mathbf{\Sigma}_{42}^{-1}\sigma_{41} = \psi_1, \tag{5}$$

where $\xrightarrow{\text{P}}$ denotes convergence in probability. The estimator (3) is now represented as

$$\hat{\psi}_1^{(m)} = s_{11} - [s_{12}s_{13}]\begin{bmatrix}\mathbf{S}_{42} & \mathbf{S}_{43} \\ \mathbf{S}_{52} & \mathbf{S}_{53}\end{bmatrix}^{-1}\begin{bmatrix}s_{41} \\ s_{51}\end{bmatrix}.$$

By the following equality

$$\begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^{-1} & 0 \\ 0 & 0\end{bmatrix} + \begin{bmatrix}\mathbf{A}^{-1}\mathbf{B} \\ -I\end{bmatrix}W[\mathbf{CA}^{-1} - I],$$

with $W = (D - CA^{-1}B)^{-1}$ (see, e.g., Anderson, 1984, Theorem A.3.3.), we have

$$\begin{aligned}\hat{\psi}_1^{(m)} &= (s_{11} - s_{12}\mathbf{S}_{42}^{-1}s_{41}) \\ &\quad - (s_{13} - s_{12}\mathbf{S}_{42}^{-1}\mathbf{S}_{43})(\mathbf{S}_{53} - \mathbf{S}_{52}\mathbf{S}_{42}^{-1}\mathbf{S}_{43})^{-1}(s_{51} - \mathbf{S}_{52}\mathbf{S}_{42}^{-1}s_{41}).\end{aligned} \tag{6}$$

Each element of $n^{1/2}(\mathbf{S}_{ij} - \mathbf{S}_{i2}\mathbf{S}_{42}^{-1}\mathbf{S}_{4j})(i \neq j,\ i, j \neq 2, 4)$ has an asymptotic normal distribution with zero mean and positive and finite variance. There exist nonzero $(m - k)$-th derivatives of the determinant of $\mathbf{S}_{53} - \mathbf{S}_{52}\mathbf{S}_{42}^{-1}\mathbf{S}_{43}$ with respect to the elements of $\mathbf{S}$ (in fact, the derivative with respect to the diagonal elements of $\mathbf{S}_{53}$ is equal to 1), and thus, the determinant of $n^{1/2}(\mathbf{S}_{53} - \mathbf{S}_{52}\mathbf{S}_{42}^{-1}\mathbf{S}_{43})$ does not vanish as n tends to infinity. This implies that $n^{1/2}$ times the second term in (6) is bounded in probability, and from (5) that

$$\hat{\psi}_1^{(m)} \xrightarrow{\mathrm{P}} \psi_1, \tag{7}$$

that is, the consistency of $\hat{\mathbf{\Psi}}^{(m)}$ holds even if the number of factors is overestimated. In this case, however, asymptotic normality of $\hat{\mathbf{\Psi}}^{(m)}$ would not hold, because $\hat{\mathbf{\Psi}}^{(m)}$ is not a continuous function of $\mathbf{S}$ at $\mathbf{S} = \mathbf{\Sigma}$. For the asymptotic distribution in a case when $m > k$, see Kano (in press).

We shall investigate the estimator $\hat{\mathbf{\Lambda}}^{(m)}$ defined by (4). Since $\mathbf{S}$ converges to $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ in probability, it follows from (7) that

$$\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)} \xrightarrow{\mathrm{P}} \mathbf{\Lambda}\mathbf{\Lambda}', \tag{8}$$

and hence, from (4):

$$\underset{p \times m}{\hat{\mathbf{\Lambda}}^{(m)}} \xrightarrow{\mathrm{P}} [\underset{p \times k}{\mathbf{\Lambda}\mathbf{Q}} : \underset{p \times (m-k)}{\mathbf{0}}]$$

for some orthogonal matrix Q of order k.

Summarizing these results, we have the following theorem.

Theorem. Assume that $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$ (with $\mathbf{\Lambda}$ of order $p \times k$), and suppose that Anderson and Rubin's sufficient condition for identifiability is satisfied. Let m $(k \leqq m \leqq (p-1)/2)$ be a number of extracted (or assumed) factors and let the I-K estimator $(\hat{\mathbf{\Lambda}}^{(m)}, \hat{\mathbf{\Psi}}^{(m)})$ be defined by (3) and (4). When $m > k$, $(\hat{\mathbf{\Lambda}}^{(m)}, \hat{\mathbf{\Psi}}^{(m)})$ converges to $([\mathbf{\Lambda}\mathbf{Q}:\mathbf{0}], \mathbf{\Psi})$ in probability for some $k \times k$ orthogonal matrix $\mathbf{Q}$ (but the asymptotic normality is not guaranteed). When $m = k$, the I-K estimator is consistent and has an asymptotic normal distribution.

For a given $\mathbf{\Sigma}$, there exist many common factor decompositions of $\mathbf{\Sigma}$ if we take several large numbers of factors, but the decomposition satisfying the A-R condition is unique (up to orthogonal rotation) when it exists. If the I-K estimator converges to $([\mathbf{\Lambda}:\mathbf{0}], \mathbf{\Psi})$, the $(\mathbf{\Lambda}, \mathbf{\Psi})$ is the parameter satisfying the A-R condition, and the factor loading matrix $\mathbf{\Lambda}$ includes no quasi-loadings (if quasi-loadings are included, $(\mathbf{\Lambda}, \mathbf{\Psi})$ does not satisfy the A-R condition). The I-K estimator therefore produces no quasi-loadings. When the number of factors is known, $(\hat{\mathbf{\Lambda}}^{(k)}, \hat{\mathbf{\Psi}}^{(k)})$ would be the best of all $(\hat{\mathbf{\Lambda}}^{(m)}, \hat{\mathbf{\Psi}}^{(m)})$ with $m \geqq k$ because the second terms in (6) may be unstable due to sample fluctuations.

We shall consider application of the I-K estimator to the choice of the number of factors. The relation (8) means that the eigenvalues of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)}$ converge to those of $\mathbf{\Lambda}\mathbf{\Lambda}'$, and the $p - k$ smallest eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}'$ are all zero. Thus, the smallest $p - k$ eigenvalues of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)}$ would be close to zero, unless the sample size is too small. It is natural to determine the number of factors by the number of eigenvalues of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)}$ greater than some constant d, and this means disregarding the column vectors of $\hat{\mathbf{\Lambda}}^{(m)}$ that are close to the zero vector. Applications of this method to real and simulated data sets will be given in sections 4 and 5, respectively.

3. Algorithm for the Choice of Λ_2 and Λ_3

There are $\binom{p-1}{m}\binom{p-m-1}{m}/2$ alternative choices of Λ_2 and Λ_3 in (2), and the same number of estimators (3) whose values are usually distinct from each other. We need to find estimators that satisfy the A-R condition and possess good properties. If the absolute value (Δ, say) of the determinant of $\mathbf{S}_{32}$ is small, it is doubtful whether the submatrix $\mathbf{\Sigma}_{32}$ corresponding to $\mathbf{S}_{32}$ is nonsingular, and hence it is natural to choose estimates with a large value of Δ. This criterion was reported to work well for some data sets in Ihara and Kano (1986). Calculating Δ for all cases often requires heavy computation, and the procedure proposed by Kano (1989) is inadequate for large p.

To develop a more efficient algorithm, we first state a simple procedure for calculating the estimate (3). Partition $\mathbf{S}$ as $\mathbf{\Lambda}$ in (2), and sweep out the elements of $\mathbf{S}$ except the first row and column vectors by choosing diagonal elements of $\mathbf{S}_{32}$ as pivots. If X and Y are matrices that sweep out the column and row vectors, respectively, we have

$$XSY = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & I & X_{23} & 0 \\ 0 & 0 & X_{33} & 0 \\ 0 & 0 & X_{43} & I \end{bmatrix} \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & \mathbf{S}_{22} & \mathbf{S}_{23} & \mathbf{S}_{24} \\ s_{31} & \mathbf{S}_{32} & \mathbf{S}_{33} & \mathbf{S}_{34} \\ s_{41} & \mathbf{S}_{42} & \mathbf{S}_{43} & \mathbf{S}_{44} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Y_{22} & Y_{23} & Y_{24} \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} * & * & * & * \\ * & 0 & * & * \\ * & I & 0 & 0 \\ * & 0 & * & * \end{bmatrix},$$

which implies that

$$XSY = \begin{bmatrix} s_{11} & s_{12}Y_{22} & s_{13} - s_{12}\mathbf{S}_{32}^{-1}\mathbf{S}_{33} & s_{14} - s_{12}\mathbf{S}_{32}^{-1}\mathbf{S}_{34} \\ s_{21} - \mathbf{S}_{22}\mathbf{S}_{32}^{-1}s_{31} & 0 & \mathbf{S}_{23} - \mathbf{S}_{22}\mathbf{S}_{32}^{-1}\mathbf{S}_{33} & \mathbf{S}_{24} - \mathbf{S}_{22}\mathbf{S}_{32}^{-1}\mathbf{S}_{34} \\ X_{33}s_{31} & I & 0 & 0 \\ s_{41} - \mathbf{S}_{42}\mathbf{S}_{32}^{-1}s_{31} & 0 & \mathbf{S}_{43} - \mathbf{S}_{42}\mathbf{S}_{32}^{-1}\mathbf{S}_{33} & \mathbf{S}_{44} - \mathbf{S}_{42}\mathbf{S}_{32}^{-1}\mathbf{S}_{34} \end{bmatrix}$$

$$= [t_{ij}], \quad \text{(say)}$$

and that

$$Y_{22}X_{33} = \mathbf{S}_{32}^{-1}.$$

We note

$$t_{11} - t_{12}t_{31} \tag{9}$$

is identical with the I-K estimator (3), showing that the estimate can be calculated by sweeping $\mathbf{S}$ out only m times and calculating (9). One merit of the procedure is that the estimates with fewer than m factors can be simultaneously obtained by calculating (9) at each step of the sweeping-out process.

Using this procedure, we can easily find $\mathbf{S}_{32}$ or pivots with a large value of Δ. Interchanging appropriately the row and column vectors of $\mathbf{S}$, we choose as a pivot the element with the largest value of all possible elements of $\mathbf{S}$ at each step. The value of Δ is usually large because it is the product of all pivot values.

Kano (1990) reported that the arithmetic mean of several I-K estimators for ψ_i with a large value of Δ is a good estimator. In this paper we average p (the number of variables) estimates of ψ_i with a large value of Δ and calculate $\hat{\mathbf{A}}^{(m)}$ by (4). If there exist

TABLE 2

Root Mean Squared Differences (RMSD) between I-K Estimates and MLE in Empirical Data Sets

	Data Sets						
	Six School Subjects	Eight Physical Variables	Davis	Emmett	Nine Psychological Tests	Thirteen Psychological Tests	Twenty-Four Psychological Tests
p, k	6, 2	8, 2	9, 1	9, 2	9, 3	13, 3	24, 4
RMSD	.01	.03	.01	.02	.01	.07	.09

fewer than p estimators, we may average them all. To find such estimates, we first denote by $\mathbf{x}_{rt}$ the value of the t-th pivot in calculating the r-th estimate for $\psi_i (r = 1, 2, \ldots, p$ and $t = 1, 2, \ldots, m)$, and choose as $\mathbf{x}_{rt}$ the element with the largest value of all possible elements of S. For $r \geqq 2$ the following restrictions are imposed:

$$|\mathbf{x}_{r1} \ldots \mathbf{x}_{rt}| \neq |\mathbf{x}_{s1} \ldots \mathbf{x}_{st}| \qquad (s = 1, 2, \ldots r - 1),$$

which guarantees that all $\mathbf{\Delta}$'s are mutually distinct, and hence the p estimates of ψ_i can be chosen differently.

4. Applications

To investigate the closeness of the I-K estimate to the MLE, we begin by using seven sets of empirical data with the number of factors assumed known from the literature. The data sets are as follows: (a) Six School Subjects in Lawley and Maxwell (1971, p. 66), (b) Eight Physical Variables in Harman (1976, p. 22), (c) Davis (1944), (d) Emmett (1949), (e) Nine Psychological Tests (exploratory sample) in Lawley and Maxwell (1971, p. 96). (f) Thirteen Psychological Tests and (g) Twenty-Four Psychological Tests in Harman (1976, p. 124). The last three sets of data were originally given by Holzinger and Swineford (1939), and the data (f), analyzed by Harman (1976, p. 172), are the first thirteen of (g). Table 2 shows the numbers of factors and the root mean squared differences (RMSD) between the I-K estimates and the MLE for $\mathbf{\Psi}$. The differences are seen to be considerably small.

To illustrate the lack of quasi-loadings in the I-K estimates, we give $\hat{\mathbf{\Psi}}^{(1)}, \ldots, \hat{\mathbf{\Psi}}^{(4)}$ and $\hat{\mathbf{A}}^{(4)}$ for Davis' data in Tables 3 and 4, respectively. The four I-K estimates and the MLE (with $m = 1$) of $\mathbf{\Psi}$ are all mutually rather close (the MLE with $m = 1$ appears in Table 1). The first column of $\hat{\mathbf{A}}^{(4)}$ and the MLE (with $m = 1$) of $\mathbf{\Lambda}$ are also very close, and every element of the 2nd, 3rd and 4th columns of $\hat{\mathbf{A}}^{(4)}$ is small. These observations would also suggest the consistency of the I-K estimator in cases with $m > k$. Note that the RMSD between $\hat{\mathbf{\Psi}}^{(m)}$ and $\hat{\psi}^{(1)}_{\mathrm{MLE}}$, the MLE with $m = 1$, becomes large as m moves away from 1.

Consider the criterion for the choice of the number of factors based on the (squared) lengths of columns of $\hat{\mathbf{A}}^{(m)}$, which are equal to the eigenvalues of $\mathbf{S} - \hat{\mathbf{\Psi}}^{(m)}$. Theorem 5.6 in Anderson and Rubin (1956) gives a necessary condition for identifiabil-

TABLE 3

I-K Estimates $\hat{\Psi}^{(m)}$ for All Possible Values of m in Davis' Data

(The last row represents RMSD between $\hat{\Psi}^{(m)}$ and $\hat{\Psi}^{(1)}_{MLE}$.)

Variable	$m = 1$	$m = 2$	$m = 3$	$m = 4$
1	.36	.36	.37	.29
2	.35	.35	.37	.33
3	.78	.78	.78	.77
4	.83	.83	.84	.81
5	.55	.56	.55	.54
6	.21	.21	.19	.14
7	.31	.31	.32	.32
8	.58	.58	.57	.46
9	.31	.31	.32	.32
RMSD	.01	.01	.02	.05

ity, and states that each column of $\Lambda\mathbf{A}$ has at least three nonzero elements for every nonsingular matrix $\mathbf{A}$ of order k. This suggests that estimates would be rather unstable and difficulties could arise if there are only two elements with a large absolute value in a column of Λ. In fact, the analysis of Bechtoldt's (1961) data has encountered difficulties because only X_1 and X_2 are variables for the Memory factor that is identified as one of the common factors (see van Driel, 1978). Thus, each column of Λ should include at least three elements with a fairly large absolute value. The average of the absolute values of elements with a large absolute value will be denoted by x; the average of the others will be represented by y. The squared length of the i-th column vector of Λ is then greater than $3x^2 + (p - 3)y^2$. Thus, we propose to set the number of factors using the number of eigenvalues of $\mathbf{S} - \hat{\Psi}^{(m)}$ greater than

$$3x^2 + (p - 3)y^2 (= d, \text{ say}). \tag{10}$$

In practical situations the value x is expected to be greater than 0.3 or 0.4 and y would be about 0.1 or 0.2, when the sample correlation matrix is used. This method is similar to the Kaiser and Guttman rule (Kaiser, 1960), which determines k by the number of eigenvalues of $\mathbf{S}$ greater than 1. The difference is whether or not the effects of Ψ and p are taken into account.

TABLE 4

I-K Estimates $\hat{\Lambda}^{(4)}$ in Davis' Data

Variable	Factor Loading			
	1	2	3	4
1	.82	-.23	.12	-.04
2	.81	-.09	.14	.13
3	.47	-.09	-.10	-.12
4	.41	.14	.01	.15
5	.67	.05	-.13	.03
6	.90	-.01	-.21	.03
7	.84	.03	.01	-.06
8	.67	.26	.13	-.11
9	.83	.03	.01	-.00
Squared Length	4.83	.16	.12	.07

The new method was applied to the seven data sets. The estimates $\hat{\Psi}^{(m)}$ with $m = [(p - 1)/2]$ were calculated, where $[c]$ represents the minimum integer not greater than c, and the numbers of factors were then chosen by the proposed procedure. Table 5 shows the estimated numbers of factors as well as the numbers of variables, values of d with $x = 0.3$ and $y = 0.15$ in (10), and the eigenvalues of $\mathbf{S} - \hat{\Psi}^{(m)}$. The criterion appears to work rather well. Note that improper solutions do not appear for the I-K and ML methods with the estimated number of factors. In Emmett's data, the critical value $d(= 0.41)$ is rather close to the closest smaller eigenvalue $(= 0.34)$, and so some risk may be incurred if the number is decided solely on the basis of our criterion. The value of Δ sometimes becomes rather small especially for large m. In such cases $\hat{\psi}_i^{(m)}$ was calculated by the arithmetic mean of estimates with Δ greater than the half of the maximum value of Δ's.

5. A Monte Carlo Experiment

A comparative study of the I-K estimator and MLE was carried out using a small Monte Carlo experiment, employing a model with $(p, k) = (9, 2)$ and a true value of Λ using the 1st and 2nd factor loading vectors of the solution of Emmett's data rounded to one decimal place (Lawley & Maxwell, 1971, p. 43); Ψ was defined by I_p − Diag

TABLE 5

Choice of the Number of Factors

Data Sets	Estimated Number of Factors	p	d	Eigenvalues of $S - \hat{\Psi}^{(m)}$					
Six School Subjects	2	6	0.34	2.22	0.59				
Eight Physical Variables	2	8	0.38	4.45	1.51	0.11			
Davis	1	9	0.41	4.83	0.16	0.12	0.07		
Emmett	2	9	0.41	4.34	0.95	0.34	0.09		
Nine Psychological Tests	3	9	0.41	3.31	1.24	0.90	0.08		
Thirteen Psychological Tests	3	13	0.50	4.64	1.36	0.96	0.24	0.18	0.10
Twenty-Four Psychological Tests	4	24	0.74	7.67 0.32	1.71 0.30	1.23 0.24	0.96 0.14	0.48 0.10	0.39

($\Lambda\Lambda'$). For each sample size (n = 50, 100 and 200), 200 sample covariance matrices drawn from a Wishart distribution $W_p(\Sigma, n - 1)$ were generated by Smith and Hocking's program (1972), in which we used physical random numbers generated by the amplified thermal noise from Zener diode (for details see Niki, 1980). The I-K estimates with $m = 1, \ldots, [(p - 1)/2] = 4$ and the MLE for $m = k$ were calculated for each sample correlation matrix along with the root mean squared error (RMSE) of $\hat{\Psi}$ to the true value. The partial Gauss-Newton algorithm given by Okamoto and Ihara (1984) was employed for calculating the MLE.

The results are shown in Table 6 as well as the square root of the asymptotic variance obtained from the standard asymptotic theory; the number of improper solutions is given in parentheses. Table 6 also includes distributions of the estimated numbers of factors from the 200 simulated correlation matrices by using the rule proposed.

The Monte Carlo experiments suggest the following: (a) When $m = k$, the I-K estimator is not inferior to the MLE for moderate sample sizes both in RMSE and in the number of improper solutions; the I-K estimator is recommended particularly when the sample size is small. (b) $\hat{\Psi}^{(m)}$ with $m \geqq k$ is consistent of order $n^{-1/2}$, (c) $\hat{\Psi}^{(k)}$ is the best of all $\hat{\Psi}^{(m)}$ with $k \leqq m \leqq [(p - 1)/2]$, and the RMSE becomes large little by little as m moves away from k, and (d) the new method for choosing the number of factors works satisfactorily.

6. Discussion

Based on the previous discussion, the following procedure may be recommended for exploratory factor analysis:

1. Let $M = [(p - 1)/2]$ and average the p estimates of $\hat{\psi}_i^{(m)}$ for all m with $1 \leqq m \leqq M$, where p is the number of variables.

TABLE 6

Comparative Study of RMSE of I-K Estimates and MLE Using the Monte Carlo Experiment

(The number of improper solutions is given in parentheses.)

(multiplied by 1000)

		I-K Estimates	MLE with $m = 2$	Square Root of Asymptotic Variance of MLE	Distribution of Estimated Numbers of Factors
n=50	m=1	236 (0)	110 (29)	106	0
	m=2	107 (0)			178
	m=3	109 (3)			22
	m=4	137 (2)			0
n=100	m=1	225 (0)	76 (2)	74	0
	m=2	75 (0)			199
	m=3	78 (0)			1
	m=4	99 (2)			0
n=200	m=1	218 (0)	53 (0)	52	0
	m=2	53 (0)			200
	m=3	55 (0)			0
	m=4	71 (0)			0

2. Choose the number (k, say) of extracted factors by the number of eigenvalues of $S - \hat{\Psi}^{(M)}$ greater than some constant d.

3. $\hat{\Psi}^{(k)}$ is employed as an estimator of Ψ, and $\hat{\Lambda}$ is calculated by (4).

Simplicity is one of the advantages of the I-K estimator. Data sets with a fairly large number of variables could be analyzed even by a personal computer with a rather small memory. For example, in the Twenty-Four Psychological Tests data, the computing time for the MLE with $m = 4$ is about sixteen times as much as that for the I-K estimate $(\hat{\Lambda}^{(4)}, \hat{\Psi}^{(4)})$; here the MLE based on the partial Gauss-Newton algorithm given by Okamoto and Ihara (1984) was employed that required five iterations for convergence (when the I-K estimates are calculated for all m, the proportion of computing times of the I-K estimate to the MLE is nearly one to six). The I-K estimates required much less computing time than the MLE for the other data sets as well. Several inexpensive estimation methods have been proposed recently (see, e.g., Bentler & Dijkstra, 1985; Hägglund, 1982; Jennrich, 1986; Okamoto, 1988), but the I-K estimator is the simplest and possesses the useful properties stated previously, though it is not asymptotically efficient.

The I-K estimates are all proper for all the data sets treated here and for every m with $1 \leqq m \leqq [(p - 1)/2]$, whereas the MLE are all improper for a large $m(\leqq [(p - 1)/2])$ (except the data of Six School Subjects). The same holds for the following data sets:

Maxwell's data for normal and neurotic children with $p = 10$ (Maxwell, 1961, p. 53 and p. 55), and Hemmerle's data with $p = 15$ (Hemmerle, 1965, p. 298). When the MLE is improper and the I-K estimate is proper in cases where m is rather large, the improper solution is likely to be due to quasi-loadings. Van Driel (1978) discussed three causes of improper solutions, and the appearance of quasi-loadings corresponds to the indefiniteness of the model.

As for the difficulties (I), (II) and (III), the I-K estimator is free from (II) and (III) because it is calculated by a noniterative procedure. In cases with $m > k$, the I-K estimate does not produce any improper solutions due to the quasi-loadings but the MLE does; in cases with $m = k$, the ML method more frequently leads to improper solutions than the I-K method according to our Monte Carlo experiment, although improper solutions could arise because of sample fluctuations (see, e.g., Anderson & Gerbing, 1984; Boomsma, 1985).

The present study leaves the following problems open. The first is to evaluate how many estimators should be averaged in calculating $\hat{\psi}_i^{(m)}$. The present I-K estimator could probably not be improved by choosing a more reasonable number because the Monte Carlo experiment shows that the present I-K estimator (the average of p estimators) is pretty good. The second is to study further how to choose d, especially the values of x and y in (10); here, applications to a great number of real data sets might provide a better assessment of d. Anderson and Rubin's necessary condition may be satisfied even when the value of x is quite small and y is equal to zero. However, factor loading vectors with such a small length that we cannot distinguish between minor factors and sampling errors should be ignored because of poor reproducibility.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

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

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155–173.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5*, (pp. 111–150). Berkeley: University of California Press.

Bechtoldt, H. P. (1961). An empirical study of the factor analysis stability hypothesis. *Psychometrika, 26*, 405–432.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42), Amsterdam: North-Holland.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50*, 229–242.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.

Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrika, 9*, 185–197.

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

Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika, 47*, 209–222.

Hakstian, A. R., & Muller, V. J. (1973). Some notes on the number of factors problem. *Multivariate Behavioral Research, 8*, 461–475.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research, 17*, 193–219.

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

Hemmerle, W. J. (1965). Obtaining maximum-likelihood estimates of factor loadings and communalities using an easily implemented iterative computer procedure. *Psychometrika, 30*, 291–302.

Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs, 48*, Chicago: University of Chicago.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika, 51*, 563–566.

Jennrich, R. I. (1986). A Gauss-Newton algorithm for exploratory factor analysis. *Psychometrika, 51*, 277–284.

Jennrich, R. I., & Robinson, S. M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis. *Psychometrika, 34*, 111–123.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443–477.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151.

Kano, Y. (1989). A new estimation procedure using g-inverse matrix in factor analysis. *Mathematica Japonica, 34*, 43–52.

Kano, Y. (1990). Comparative studies of non-iterative estimators based on Ihara and Kano's method in exploratory factor analysis. *Communications in Statistics, Part A, 19*.

Kano, Y. (in press). The asymptotic distribution of a non-iterative estimator in exploratory factor analysis. *Annals of Statistics*.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, Series A, 60*, 64–82.

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

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.

Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika, 40*, 505–517.

Mattsson, A., Olsson, U., & Rosén, M. (1966). *The maximum likelihood method in factor analysis with special consideration to the problem of improper solutions* (Research Report). Sweden: University of Uppsala, Institute of Statistics.

Maxwell, A. E. (1961). Recent trends in factor analysis. *Journal of the Royal Statistical Society, Series A, 124*, 49–59.

Niki, N. (1980). Machine generation of random numbers. *Proceedings of the Institute of Statistical Mathematics, 27*, 115–131. (In Japanese)

Okamoto, M. (1988). Early-step estimators in least-squares factor analysis. *Mathematica Japonica, 33*, 565–576.

Okamoto, M., & Ihara, M. (1984). Partial Gauss-Newton algorithm for least-squares and maximum likelihood methods in factor analysis. *Journal of the Japan Statistical Society, 14*, 137–144.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika, 20*, 93–111.

Sato, M. (1987). Pragmatic treatment of improper solution in factor analysis. *Annals of the Institute of Statistical Mathematics, Part B, 39*, 443–455.

Smith, W. B., & Hocking, R. R. (1972). Wishart variate generator. *Applied Statistics, 21*, 341–345.

Tumura, Y., & Sato, M. (1981). On the convergence of iterative procedures in factor analysis. *TRU Mathematics, 17*, 159–168.

van Driel, O. P. (1978). On various casues of improper solutions in maximum likelihood factor analysis. *Psychometrika, 43*, 225–243.

Manuscript received 5/17/87
Final version received 2/23/89