Some Remarks to Using Single Factor Analysis as a Measurement Model

Wim P. Krijnen[1]

Abstract

Single factor analysis is frequently applied as a measurement model (Jöreskog & Goldberger, 1975). It is shown that three loadings bounded away from zero for the population vector as well as for its least squares estimator is necessary and sufficient for the regularity conditions of asymptotic normality to hold. In particular, the model is identified, and the condition number of the Jacobian matrix as well as the asymptotic variance matrix are bounded. Least squares estimation is a practically feasible method for bootstrap (Monte Carlo) estimation, even when the population parameter is close to the boundary of the parameter set. Furthermore, bootstrap least squares estimation provides the possibility to test the normality of the empirical distribution of the statistics, to obtain nonnormal confidence intervals within the parameter set, and to draw logically consistent statistical inferences. A Monte Carlo experiment is reported as well as three bootstrap applications to empirical data.

Key Words: Mean square error of factor prediction, Validity, Reliability, Nonparametric estimation, Asymptotic normality.

[1] Part of this paper was written while the author was at the University of Groningen. The address of Wim Krijnen is Lisdodde 1, 9679 MC Scheemda, The Netherlands (phone: +31 597593743).

1. Introduction

Factor analysis (Spearman, 1904) with a single factor is frequently applied as a measurement model (Jöreskog & Goldberger, 1975). Application of the model involves estimation of the parameters for making inferences to the population and prediction of the factor for making inferences to the cases.

From the theory of the generalized method of moments, sufficient conditions are known for the parameters to be asymptotically normal. One of these conditions involves the Jacobian matrix to have full column rank (Browne, 1984). However, even when this holds, the Jacobian matrix may be ill-conditioned. A necessary and sufficient condition for the Jacobian matrix to be well-conditioned is not known. Additionally, the regularity conditions for asymptotic normality have not been verified with respect to the existence of the estimator, the practical feasibility of the estimator, and the Jacobian matrix to have full rank.

Generally, the single factor is said to be "observable" or "determinate" if and only if it is a linear combination of the observable variables. This is almost surely the case if and only if the mean squared error (MSE) of factor prediction is zero (e.g. Krijnen, Dijkstra, & Gill, 1998). Moreover, a single factor is observable if and only if the validity attains unity. However, on the basis of asymptotic normality, empirical cases can occur for which it is impossible to draw logically consistent statistical inferences. In particular, this impossibility occurs when one is unable to reject the null hypothesis of unit validity or that of zero MSE.

The purpose of this paper is to give a necessary and sufficient condition for the Jacobian matrix to be well-conditioned. Furthermore, the purpose is to illustrate that (bootstrap) Monte Carlo approximations do not suffer from the above mentioned logical inconsistency, yield confidence intervals within the parameter set that correct for non-normality, and are useful for testing the normality of the empirical distribution of the least squares statistics.

2. The single-factor model

The single-factor model is

$$\boldsymbol{X} = \boldsymbol{\mu}_o + \boldsymbol{\lambda}_o F + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{X}$ is the random vector of order p with observable scores on the variables, $\boldsymbol{\mu}_o = \mathcal{E}[\boldsymbol{X}]$ its expectation, F the random factor, $\boldsymbol{E}$ the random error vector of order p, and $\boldsymbol{\lambda}_o$ the loadings vector of order p (e.g. Anderson & Rubin, 1956; Lawley & Maxwell, 1971). The subscript zero is added to distinguish population parameters from mathematical variables. It will be assumed that $\boldsymbol{\mu}_o = \boldsymbol{o}$, $\text{var}[F] = 1$, $\mathcal{E}[F] = 0$, $\mathcal{E}[\boldsymbol{E}] = \boldsymbol{o}$, $\mathcal{E}[\boldsymbol{E}F] = \boldsymbol{o}$, and $\mathcal{E}[\boldsymbol{E}\boldsymbol{E}'] = \boldsymbol{\Psi}_o$ diagonal positive semi definite (Lawley & Maxwell, 1971). It will be convenient to collect the parameters to be estimated into the population vector $\boldsymbol{\theta}_o = \begin{pmatrix} \boldsymbol{\lambda}_o \\ \boldsymbol{\Psi}_o \boldsymbol{\iota} \end{pmatrix}$, where $\boldsymbol{\iota}$ is the p-vector with unit elements. Let λ_{i0} be element i of $\boldsymbol{\lambda}_o$, ψ_{ii0} be element ii of $\boldsymbol{\Psi}_o$, and Θ the set of parameter vectors. Defining $\text{var}[\boldsymbol{X}] = \boldsymbol{\Sigma}_o$, it follows that

$$\boldsymbol{\Sigma}_o = \boldsymbol{\lambda}_o \boldsymbol{\lambda}_o' + \boldsymbol{\Psi}_o. \tag{2}$$

When no loading or error variance is fixed, invariance considerations reveal that the observable variables may be standardized without loss of generality (cf. Browne & Shapiro, 1991). Hence, it will be assumed that $\boldsymbol{\Sigma}_o$ is a correlations matrix. It follows that $\boldsymbol{\lambda}_o$ contains the correlations between the observable variables and the factor.

The factor F is a random variable which may be predicted by the linear combination of the observable variables which yields a minimum amount of mean squared error. The latter is accomplished by $\widehat{F} = \boldsymbol{\lambda}_o' \boldsymbol{\Sigma}_o^{-1} \boldsymbol{X}$, the projection of F on the space spanned by the observable variables (Luenberger, 1969, p.51). This predictor maximizes $\text{cor}[F, \widehat{F}]$, the product moment correlation between F and $\widehat{F}$ (Rao, 1973, p.264), which is commonly called the "validity" (Lord & Novick, 1968, p.261). The "reliability" of a measurement is defined as $\text{cor}^2[F, \widehat{F}]$, the squared correlation between $\widehat{F}$ and F (Lord & Novick, 1968, p.61; Jöreskog, 1971).

For notational brevity let $\gamma_o = \boldsymbol{\lambda}_o' \boldsymbol{\Psi}_o^{-1} \boldsymbol{\lambda}_o$, given that $\boldsymbol{\Psi}_o$ is diagonal positive definite. It is immediate that $\text{MSE}[\widehat{F}] (= \frac{1}{1+\gamma_o})$ is positive. Hence, there is no linear combination of the observable variables which yields the factor, so that the factor is "unobservable". Furthermore, it follows that $\text{cor}[F, \widehat{F}] = \left(\frac{\gamma_o}{1+\gamma_o}\right)^{1/2}$.

3. The Jacobian Matrix

The Jacobian matrix $\boldsymbol{\Delta}_o = \left.\frac{\partial\boldsymbol{\sigma}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}$ plays an important role in the regularity conditions for the asymptotic normality of estimators for the model of factor analysis (Browne, 1984). By the rules of taking first order differentials (e.g. Magnus & Neudecker, 1991) it follows that

$$\boldsymbol{\Delta}_o = \left[(\boldsymbol{\lambda}_o \otimes \boldsymbol{I}) + (\boldsymbol{I} \otimes \boldsymbol{\lambda}_o), (\boldsymbol{e}_1 \otimes \boldsymbol{e}_1, .., \boldsymbol{e}_p \otimes \boldsymbol{e}_p\right], \tag{3}$$

where $\boldsymbol{e}_j$ is column j of the identity matrix $\boldsymbol{I}$. Part of the conditions is that the Jacobian matrix $\boldsymbol{\Delta}_o$ has full column rank (Browne, 1984). Obviously, the matrix $\boldsymbol{\Delta}_o$ has full column rank if its condition number $\kappa(\boldsymbol{\Delta}_o) = \text{sv}_{max}(\boldsymbol{\Delta}_o)/\text{sv}_{min}(\boldsymbol{\Delta}_o)$ is bounded, where "sv" is shorthand for singular value. The condition number quantifies the sensitivity of the linear equation problem (cf. Golub & Van Loan, 1983, p.26), on the solution of which the asymptotic normality of an estimator is based (e.g. Ferguson, 1958; Browne, 1974, 1984). A necessary condition is given in

Result 1. In case $p = 3$, the condition number $\kappa(\boldsymbol{\Delta}_o) = \infty$ if one of the loadings is zero.

Proof. From standard properties of the Kronecker product it follows that

$$\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o = \begin{bmatrix} 2\boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o\boldsymbol{I} + 2\boldsymbol{\lambda}_o\boldsymbol{\lambda}_o' & 2\text{diag}(\boldsymbol{\lambda}_o) \\ 2\text{diag}(\boldsymbol{\lambda}_o) & \boldsymbol{I} \end{bmatrix}. \tag{4}$$

Hence, $\lambda_{io}^2 \leq 1$ for all i, implies $1 \leq \text{ev}_{max}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o) \leq \text{tr}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o) \leq 2p^2 + 2p + p$, where "ev" is shorthand for eigen value. Let a pair of single bars denote the determinant of a certain matrix. From a well-known result for the determinant of a partitioned matrix (e.g. Rao, 1973, p.32), it follows that

$$|\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o| = 2^p|\boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o\boldsymbol{I} + \boldsymbol{\lambda}_o\boldsymbol{\lambda}_o' - 2\text{diag}(\boldsymbol{\lambda}_o * \boldsymbol{\lambda}_o)| \tag{5}$$

$$= 2^p \left| \begin{bmatrix} \lambda_{2o}^2 + \lambda_{3o}^2 & \lambda_{1o}\lambda_{2o} & \lambda_{1o}\lambda_{3o} \\ \lambda_{2o}\lambda_{1o} & \lambda_{1o}^2 + \lambda_{3o}^2 & \lambda_{2o}\lambda_{3o} \\ \lambda_{3o}\lambda_{1o} & \lambda_{3o}\lambda_{2o} & \lambda_{1o}^2 + \lambda_{2o}^2 \end{bmatrix} \right|, \tag{6}$$

where $*$ is the element-wise Hadamard product. Since the order of the loadings is immaterial, suppose $\lambda_3 = 0$. Then the first 2 by 2 principal submatrix has rank one, so that $\text{ev}_{min}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$=0. Hence, $\kappa(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$ and its root $\kappa(\boldsymbol{\Delta}_o)$ are infinite. This completes the proof.

Since eigenvalues are continuous functions of the elements of their matrix, its follows for the $p = 3$ case that the condition number of the Jacobian matrix is unbounded when any of the loadings tends to zero. When $p > 3$ the necessary condition is that three loadings are bounded away from zero, since permutations would lead to a principal submatrix equal to that in Equation (6). That this condition is also sufficient is stated in

Result 2. $\kappa(\boldsymbol{\Delta}_o)$ is bounded if three loadings are bounded away from zero.
Proof. It is convenient to use permutations such that $\lambda_p^2 = \max\{\lambda_1^2, .., \lambda_p^2\}$. Since $\text{ev}_{max}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$ is positive and bounded, it suffices to prove that $\text{ev}_{min}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$ is bounded away from zero. This will be accomplished by showing that $|\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o|$ is bounded away from zero. Using the partitioning $\boldsymbol{\lambda}_o = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \lambda_p \end{pmatrix}$ in (5) and a result for the determinant of a partitioned matrix (e.g. Rao, 1973, p.32), it follows that

$$|\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o| = 2^p\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1 \left| \boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o\boldsymbol{I}_{p-1} - 2\text{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1) + \left(1 - \frac{\lambda_p^2}{\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1}\right)\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1' \right|. \tag{7}$$

Let $\alpha = \lambda_p^2/\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1$. The proof proceeds by separating the cases $\alpha \leq 1$ and $\alpha > 1$.

Let $\alpha \leq 1$ and three loadings bounded away from zero. It follows that $\boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o\boldsymbol{I}_{p-1} -$ $2\text{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)$ is positive definite, and $(1-\alpha)\,\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1'$ is positive semi definite. Hence, (7) implies that

$$|\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o| \geq 2^p\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1 \left| \boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o\boldsymbol{I}_{p-1} - 2\text{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1) \right| = 2^p\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1 \prod_{i=1}^{p-1} \left(\boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o - 2\lambda_i^2\right), \tag{8}$$

which is bounded away from zero since three loadings are. Hence, $\mathrm{ev}_{min}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$ is bounded away from zero.

Let $\alpha > 1$. It will be shown that the smallest eigenvalue of the matrix of the determinant to the right hand side of (7) is bounded away from zero. After taking its negative, the eigenvector which corresponds to this smallest eigenvalue is the solution of

$$\max_{\boldsymbol{x}'\boldsymbol{x}=1} \boldsymbol{x}'\left[(\alpha-1)\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1' + 2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)\right]\boldsymbol{x}. \tag{9}$$

Since, this eigenvector will have a zero element if the corresponding element in $\boldsymbol{\lambda}_1$ is zero, the problem can be redefined by partitioning in such a manner that the diagonal matrix in (9) is positive definite. For the sake of simple notation it will be supposed, without loss of generality, that $2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)$ is positive definite. By letting $\boldsymbol{y} = [2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{1/2}\boldsymbol{x}$, the problem can be written as

$$\max_{\boldsymbol{y}'[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1}\boldsymbol{y}=1} \boldsymbol{y}'\left[(\alpha\text{-}1)[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1/2}\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1'[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1/2} + \boldsymbol{I}\right]\boldsymbol{y}. \tag{10}$$

The solution of this problem is well-known to be equal to the suitably normalized vector $[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1/2}\boldsymbol{\lambda}_1$. Hence

$$\boldsymbol{x} = \left(\boldsymbol{\lambda}_1'[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-2}\boldsymbol{\lambda}_1\right)^{-1/2}[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1}\boldsymbol{\lambda}_1 \tag{11}$$

is the solution to (9). Using that $\boldsymbol{\lambda}_1'[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-1}\boldsymbol{\lambda}_1 = (p-1)/2$ and

$$\boldsymbol{\lambda}_1'[2\mathrm{diag}(\boldsymbol{\lambda}_1 * \boldsymbol{\lambda}_1)]^{-2}\boldsymbol{\lambda}_1 = 1/(4\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1),$$

it follows that the value of the maximum of (9) may be written as

$$\boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1\left[(\alpha-1)(p-1)^2 + 2(p-1)\right]. \tag{12}$$

Hence, the smallest eigenvalue of the matrix of the determinant to the right hand side of (7) is

$$\boldsymbol{\lambda}_o'\boldsymbol{\lambda}_o - \boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1\left[(\alpha-1)(p-1)^2 + 2(p-1)\right] = \boldsymbol{\lambda}_1'\boldsymbol{\lambda}_1\left[1 + (p-1)^2 - 2(p-1) + \alpha p(2-p)\right]. \tag{13}$$

This eigenvalue is positive if and only if

$$\alpha > \frac{1 + (p-1)^2 - 2(p-1)}{p(2-p)} = 1 - \frac{2}{p}. \tag{14}$$

The latter condition holds under our supposition $\alpha > 1$. This completes the proof.

From Result 1 and 2 it follows that three loadings bounded away from zero is necessary and sufficient for the condition number of the Jacobian matrix to be bounded.

4. Estimation by least squares

To verify the regularity conditions for least squares factor analysis, the main properties of alternating least squares will briefly be reiterated. Let $\boldsymbol{R}$ be the matrix with sample correlation coefficients, vec$\boldsymbol{R} = \boldsymbol{r}$, and vec$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\sigma}(\boldsymbol{\theta})$, where "vec" transforms a matrix into a vector by stacking its columns one underneath the other. The least squares function is defined as

$$q(\boldsymbol{\theta},\boldsymbol{r}) = ||\boldsymbol{r} - \boldsymbol{\sigma}(\boldsymbol{\theta})||^2 = ||\boldsymbol{R} - \boldsymbol{\lambda}\boldsymbol{\lambda}' - \boldsymbol{\Psi}||^2, \tag{15}$$

where $||.||$ denotes the Euclidean norm. A vector $\widehat{\boldsymbol{\theta}}_n$ which minimizes q over Θ is called the estimator of the factor model. This vector may be obtained by an algorithm based on the idea of alternating least squares. In particular, substitution of the optimal choice Diag$(\boldsymbol{R} - \boldsymbol{\lambda}\boldsymbol{\lambda}')$ for $\boldsymbol{\Psi}$ allows the function to be written as a constant plus $||\boldsymbol{r}_j - \underline{\boldsymbol{\lambda}}_j\lambda_j||^2$, where $\boldsymbol{r}_j$ is column j of $\boldsymbol{R}-\boldsymbol{I}$, $\underline{\boldsymbol{\lambda}}_j$ is equal to $\boldsymbol{\lambda}$ except for its jth element which is zero. Now take $\lambda_j = (\underline{\boldsymbol{\lambda}}_j'\underline{\boldsymbol{\lambda}}_j)^{-1}\underline{\boldsymbol{\lambda}}_j'\boldsymbol{r}_j$ if $\left[(\underline{\boldsymbol{\lambda}}_j'\underline{\boldsymbol{\lambda}}_j)^{-1}\underline{\boldsymbol{\lambda}}_j'\boldsymbol{r}_j\right]^2 < 1$ and $\lambda_j = 1(-1)$ if $(\underline{\boldsymbol{\lambda}}_j'\underline{\boldsymbol{\lambda}}_j)^{-1}\underline{\boldsymbol{\lambda}}_j'\boldsymbol{r}_j > 1(< -1)$ (Zegers & Ten Berge, 1983). In order to prevent that statistical inferences are drawn from locally instead of the globally optimal vector, the algorithm was run five times in this study for each sample correlation matrix. After randomly chosing $\boldsymbol{\lambda}$, a run of the algorithm generates a sequence of function values $\{q(\boldsymbol{\theta}_k,\boldsymbol{r})\}$ and a sequence of parameter vectors $\{\boldsymbol{\theta}_k\}$.

It is well-known that the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$ is based on

$$\frac{\partial q(\boldsymbol{\theta},\boldsymbol{r})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n} = \boldsymbol{o}, \tag{16}$$

where from the rules of taking the differential (e.g. Magnus, & Neudecker, 1991), it can be found that

$$\frac{\partial q(\boldsymbol{\theta},\boldsymbol{r})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} 4\left(\|\boldsymbol{\lambda}\|^2 \boldsymbol{I} + \boldsymbol{\Psi} - \boldsymbol{R}\right)\boldsymbol{\lambda} \\ 2\left(\boldsymbol{\Psi}\boldsymbol{\iota} - \boldsymbol{\iota} + \boldsymbol{\lambda} * \boldsymbol{\lambda}\right) \end{bmatrix}. \tag{17}$$

After a finite number of iterations (16) does not hold exactly for $\boldsymbol{\theta} = \boldsymbol{\theta}_k$. To ascertain that it holds with a prescribed degree a accuracy, we need a convergence criterion ϵ (which was fixed to 10^{-8} in this study). For a vector interior to the parameter set, the algorithm was stopped when the maximum of the absolute values of the elements in (17) is smaller than ϵ. However, for Monte Carlo applications to small samples a solution on the boundary of the parameter set may occur, so that (16) does not hold almost surely. In such a case the algorithm was be stopped when the absolute values of the difference between the loadings before and after updating are smaller than ϵ.

5. Regularity Conditions and Asymptotic Normality

Browne (1984) has given regularity conditions under which the estimator of the factor model is asymptotically normal. It can be shown that these conditions actually hold if three loadings are bounded away from zero for the least squares estimator as well as for the population vector, both being interior points of the parameter set. This can be seen as follows.

Since $\lambda_i \in [-1, 1]$ and $\psi_{ii} \in [0, 1]$, and Θ is the cartesian product of these closed and bounded real intervals (cf. Copson, 1968, p.80), it follows that Θ is closed and bounded, and hence compact (Rudin, 1976, p.40). From Θ compact and q continuous on Θ, it follows that q attains its infimum on Θ (Rudin, 1976, p.89). That is, there exists a vector $\widehat{\boldsymbol{\theta}}_n \in \Theta$ such that $q(\widehat{\boldsymbol{\theta}}_n,\boldsymbol{r}) = \inf_{\boldsymbol{\theta}\in\Theta} q(\boldsymbol{\theta},\boldsymbol{r})$.

Obviously, three loadings bounded away from zero implies that there are three non-zero loadings. The latter is necessary and sufficient for the identification of the parameters in a neighborhood of e.g. $\boldsymbol{\theta}_o$ (Anderson & Rubin, 1956; Shapiro, 1984). That is, $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$ implies $\boldsymbol{\theta} = \boldsymbol{\theta}_o$. Furthermore, by the law of contraposition, the latter is quivalent to $\boldsymbol{\theta} \neq \boldsymbol{\theta}_o$ implies $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$. That is, different parameters yield different correlations matrices.

Since each update is based on linear projection, $\{q(\boldsymbol{\theta}_k, \boldsymbol{r})\}$ is monotonically decreasing, as $k \to \infty$. Hence, apart from local optima, the sequence of function values converges to the infimum of q, because q is bounded from the below (Rudin, 1976, p.55). For convergence of the sequence of parameter vectors a slightly stronger condition must hold. In particular, the condition that q attains its infimum at a unique point is neccesary and sufficient for $\{\boldsymbol{\theta}_k\} \to \widehat{\boldsymbol{\theta}}_n$, as $k \to \infty$ (Krijnen, Kroonenberg, & Dijkstra, 1998).

Moreover, all derivatives involved exist (and are continuous) since all partials are taken of polynomials of elements of the parameter vector (cf. Rudin, 1976, p.105). The first order derivative of q is zero, since the estimator is an internal point of the parameter set. From Result 3 and the continuity of $\boldsymbol{\Delta}_o$, it follows that the Jacobian has full rank and that it is bounded in a neighborhood of $\boldsymbol{\theta}_o$. Additionally, the computation of $(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1}$ is practically feasible since $\mathrm{ev}_{min}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)$ is bounded away from zero. Finally, it is straightforward to show that the second order *partial* derivatives of $\boldsymbol{\sigma}(\boldsymbol{\theta})$ are continuous and bounded in a neighborhood of $\boldsymbol{\theta}_o$.

It is well-known (cf. Ferguson, 1958; Browne, 1984; Bentler & Dijkstra, 1985) that these regularity conditions imply, under e.g. finite fourth order moments of independently and identically distributed observable variables,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_o) \xrightarrow{d} N\left(\boldsymbol{o}, (\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1}\boldsymbol{\Delta}_o'\boldsymbol{\Omega}_o\boldsymbol{\Delta}_o(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1}\right), \tag{18}$$

where $\boldsymbol{\Omega}_o$ is the asymptotic variance matrix of $\sqrt{n}\boldsymbol{r}$ (e.g. Hsu, 1949; Browne & Shapiro, 1986). In particular, Result 2, $(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1/2}\boldsymbol{\Delta}_o'\boldsymbol{\Omega}_o\boldsymbol{\Delta}_o(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1/2} \leq \mathrm{ev}_{max}(\boldsymbol{\Omega}_o)\boldsymbol{I}$, and $(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1} \leq (\mathrm{ev}_{min}(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o))^{-1}\boldsymbol{I}$, implies that the asymptotic variance matrix in (18) is bounded.

The asymptotic normality of estimators of the MSE of prediction, the validity, and the reliability can be obtained as follows. Since these statistics are estimators of functions of $\gamma_o = \boldsymbol{\lambda}_o' \boldsymbol{\Psi}_o^{-1} \boldsymbol{\lambda}_o = \sum_{j=1}^{p} \frac{\lambda_{jo}^2}{\psi_{jo}}$, it is convenient to go first into the asymptotic distribution of $\widehat{\gamma}_n = \widehat{\boldsymbol{\lambda}}_n' \widehat{\boldsymbol{\Psi}}_n^{-1} \widehat{\boldsymbol{\lambda}}_n$. It is easy to see that $\frac{\partial \gamma}{\partial \boldsymbol{\theta}} \Big| \boldsymbol{\theta} = \boldsymbol{\theta}_o$ $= (\frac{2\lambda_{1o}}{\psi_{1o}}, .., \frac{2\lambda_{po}}{\psi_{po}}, -\frac{\lambda_{1o}^2}{\psi_{1o}^2}, .., -\frac{\lambda_{po}^2}{\psi_{po}^2})'$. Lets denote this vector by $\boldsymbol{\delta}_o$. Then (18) and the multivariate version of the delta theorem (or "method"), implies

$$\sqrt{n}(\widehat{\gamma}_n - \gamma_o) \overset{d}{\to} N\left(o, \sigma_{\gamma_o}^2\right), \tag{19}$$

where $\sigma_{\gamma_o}^2 = \boldsymbol{\delta}_o'(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1}\boldsymbol{\Delta}_o'\boldsymbol{\Omega}_o\boldsymbol{\Delta}_o(\boldsymbol{\Delta}_o'\boldsymbol{\Delta}_o)^{-1}\boldsymbol{\delta}_o$ (Serfling, 1980, p.122).

Let g be a function of γ_o having a non-zero first order derivative $g'(\gamma_o)$. Then, (19) and the univariate version of the delta theorem (Serfling, 1980, p.118) imply

$$\sqrt{n}\left(g(\widehat{\gamma}_n) - g(\gamma_o)\right) \overset{d}{\to} N\left[o, [g'(\gamma_o)]^2\sigma_{\gamma_o}^2\right]. \tag{20}$$

Let $g(\gamma_o) = \frac{1}{1+\gamma_o}$, the MSE of $\widehat{F}$. Then its asymptotic variance follows from (20) and $[g'(\gamma_o)]^2 = \left(\frac{1}{1+\gamma_o}\right)^4$. Let $g(\gamma_o) = \frac{\gamma_o}{1+\gamma_o}$, the reliability of $\widehat{F}$. Then its asymptotic variance follows from (20) and $[g'(\gamma_o)]^2 = \left(\frac{1}{1+\gamma_o}\right)^4$. Therefore the asymptotic variance of the estimated reliability is equal to that of the MSE of prediction. Let $g(\gamma_o) = \left(\frac{\gamma_o}{1+\gamma_o}\right)^{1/2}$, the validity of $\widehat{F}$. Then its asymptotic variance follows from (20) and $[g'(\gamma_o)]^2 = \frac{\gamma_o}{4}\left(\frac{1}{1+\gamma_o}\right)^3$. In many cases, the loadings are large or the number of variables is large in the sense that $\gamma_o > 2$. This implies that $\left(\frac{1}{1+\gamma_o}\right)^4 < \frac{\gamma_o}{4}\left(\frac{1}{1+\gamma_o}\right)^3$, so that the asymptotic variance of the estimated reliability is smaller than that of the estimated validity.

Since all asymptotic variances are estimated with probability one by their sample analogs, Slutsky's theorem (Serfling, 1980, p.19) implies that these can be used for empirical purposes.

7. Monte Carlo Experiment

An estimator which uniquely optimizes a continuous function of the parameters and the sample correlation coefficients, such as q, is a continuous function of the correlation coefficients (Jennrich, 1969). Furthermore; it is well-known that extreme correlations require a large sample size for asymptotic normality (cf. Cramér, 1946, p.378). Hence, the possibility of drawing valid statistical inferences on the basis of asymptotic normality may be in danger when the population parameter vector is close to the boundary of the parameter set and the sample size is not large enough. This will be investigated by the following Monte Carlo experiment.

The model was constructed by taking $\boldsymbol{\lambda}_o = (.40, .50, .60, .70)'$, and the diagonal matrix $\boldsymbol{\Psi}_o$ such that $\boldsymbol{\lambda}_o\boldsymbol{\lambda}_o' + \boldsymbol{\Psi}_o$ is a correlations matrix according to Equation (2). Samples of size 20, 50, 100, and 200 were drawn from the $N(\boldsymbol{o}, \boldsymbol{\lambda}_o\boldsymbol{\lambda}_o' + \boldsymbol{\Psi}_o)$ distribution. Per size 2000 samples were drawn and analyzed. Let $\bar{\bar{\theta}}_i$ be the mean of parameter i and $\widehat{\sigma}$ the standard deviation over the 2000 points. From the 2000 points per sample size, the mean, the probability of the Kolmogorov-Smirnov (KS) statistic, the normal 95% confidence interval $[\bar{\bar{\theta}}_i - \widehat{\sigma}1.95996, \bar{\bar{\theta}}_i + \widehat{\sigma}1.95996]$, and the 2.5 and 97.5 percentile points were computed. These are given in Table 1.

From Table 1 it can be observed that the probability of the KS statistic increases with the sample size, but that it is low for the MSE of prediction, the validity, and the reliability. Hence, the empirical distribution of the parameter estimates approaches the normal distribution with increasing sample size, but it is far from being normal for the MSE of prediction, the validity, and the reliability. For all sample sizes, the mean of the estimates for the loadings is fairly close to the corresponding population value. However, for sample sizes up to 100, the error variances are underestimated by the means.

It can be observed that a large probability of the KS statistic occurs together with small differences between the two types of confidence intervals and vice versa. For sample sizes smaller than 200, the left point of the normal interval is larger than that of the percentile method. This implies that the parameter estimates are not normally distributed around their mean. Furthermore, the percentile intervals are subsets of the parameter set, whereas the normal confidence intervals are not. The latter is disturbing because it implies a positive mass of the distribution of a statistic out side the parameter set whereas from the algorithm we are completely sure that there is no.

For the percentile method the occurrence of 2.5% times a unit loading coincides

Table 1: The mean, the probability of the Kolmogorov-Smirnov statistic, lower (N2.5) and upper (N97.5) points of confidence intervals from normality, lower (MC2.5) and upper (MC97.5) points of the Monte Carlo percentile confidence intervals, for the parameters (pmrt), MSE of prediction, the validity (val), the reliability (rel), based on 2000 estimates from samples of size 20, 50, 100, and 200 drawn from the $N(\boldsymbol{o}, \boldsymbol{\lambda}_o\boldsymbol{\lambda}_o' + \boldsymbol{\Psi}_o)$ distribution.

n	statistic	λ_1	λ_2	λ_3	λ_4	ψ_{11}	ψ_{22}	ψ_{33}	ψ_{44}	MSE	val	rel
20	pmrt	.40	.50	.60	.70	.84	.75	.64	.51	.33	.82	.67
	mean	.40	.50	.60	.69	.76	.68	.58	.47	.16	.91	.84
	N2.5	-.13	-.02	.12	.22	.33	.18	.03	-.12	-.10	.77	.59
	N97.5	.94	1.02	1.09	1.15	1.20	1.18	1.12	1.06	.41	1.06	1.10
	MC2.5	-.19	-.10	.05	.17	.12	.00	.00	.00	.00	.75	.57
	MC97.5	.93	1.00	1.00	1.00	1.00	1.00	.99	.97	.43	1.00	1.00
	P_{KS}	.01	.03	.00	.00	.00	.00	.00	.00	.00	.00	.00
50	pmrt	.40	.50	.60	.70	.84	.75	.64	.51	.33	.82	.67
	mean	.40	.50	.60	.70	.82	.73	.61	.48	.26	.86	.74
	N2.5	.08	.20	.30	.40	.56	.42	.25	.06	.05	.74	.53
	N97.5	.72	.80	.90	1.01	1.07	1.03	.98	.90	.47	.98	.95
	MC2.5	.07	.19	.28	.39	.52	.37	.19	.00	.00	.74	.55
	MC97.5	.69	.79	.90	1.00	.99	.96	.92	.85	.45	1.00	1.00
	P_{KS}	.51	.80	.31	.09	.00	.00	.00	.00	.00	.00	.00
100	pmrt	.40	.50	.60	.70	.84	.75	.64	.51	.33	.82	.67
	mean	.40	.50	.60	.71	.83	.74	.63	.49	.29	.84	.71
	N2.5	.17	.29	.39	.49	.65	.53	.37	.19	.15	.76	.56
	N97.5	.62	.71	.81	.92	1.01	.95	.88	.79	.44	.93	.85
	MC2.5	.17	.28	.38	.49	.62	.50	.35	.13	.11	.76	.58
	MC97.5	.62	.70	.81	.93	.97	.92	.86	.76	.42	.94	.88
	P_{KS}	.87	.76	.46	.39	.00	.00	.00	.00	.00	.00	.00
200	pmrt	.40	.50	.60	.70	.84	.75	.64	.51	.33	.82	.67
	mean	.40	.50	.60	.70	.83	.74	.64	.50	.31	.83	.69
	N2.5	.24	.35	.45	.55	.71	.60	.46	.29	.22	.77	.59
	N97.5	.56	.65	.75	.85	.96	.89	.81	.71	.41	.89	.78
	MC2.5	.23	.35	.44	.55	.69	.58	.44	.27	.21	.77	.59
	MC97.5	.56	.65	.74	.85	.94	.88	.80	.70	.41	.89	.78
	P_{KS}	.65	.90	.71	.41	.02	.06	.16	.06	.01	.04	.01

with the occurrence of 2.5% times a corresponding zero error variance, 2.5% times a zero MSE of prediction, and 2.5% times of unit validity. In this sense the percentile point method yields inferences which are logically consistent. The results demonstrate that this does not hold for the normal intervals. In particular, for the sample size equal to 50 one would reject with 95% certainty from the normal confidence interval that the MSE of prediction is positive, but one cannot reject the false hypothesis that λ_{4_o} is smaller than unity. Consequently, Equation (2) can only hold if $\psi_{44_o} = 0$, which implies the observability of the factor. Furthermore, from the percentile points, one can not reject the false hypotheses that λ_{4_o} equals unity, ψ_{44_o} equals zero, the MSE of prediction equals zero, and the validity equals unity. At least these inferences are logically consistent. From the results where the sample size equals 200 it can be seen that the danger of not rejecting false hypotheses is less severe. However, when the sample size is 100, from the normal confidence intervals one cannot reject the false hypothesis that $\psi_{44_o} = 1$. Additionally, from the KS statistic the hypothesis that the statistics are normally distributed must be rejected. From the percentile points the false hypothesis $\psi_{44_o} = 1$ is rejected.

7. Bootstrapping

The Bootstrap can be seen as a Monte Carlo estimation method of the empirical distribution of the statistics (Efron, 1979). It is based on the empirical data at hand. In particular, for the purpose of confidence interval estimation, 2000 independent samples of size n were drawn from the original with replacement (Hall, 1986b; Efron, 1987). Running the algorithm with each of these independent correlations matrices as input gives an estimate of the empirical distribution of the statistics. Non-symmetric confidence intervals of size 95% around the bootstrap mean can be constructed from the empirical distribution by the percentile method as was done in the Monte Carlo experiment.

Experience with empirical sets of data reveals that it is extremely difficult to forecast, given a fixed sample size, the closeness of the distribution of the statistics to the normal distribution. In addition, to the best knowledge of the author, the accuracy (or rate) of the convergence to the normal distribution is analytically unknown (except for certain specialized univariate cases (e.g. Hall, 1988b)). We may,

Table 2: The least squares parameter estimate, lower (AN2.5) and upper (AN97.5) points of confidence intervals from asymptotic normality, Bootstrap mean (BSmean), lower (BS2.5) and upper (BS97.5) points of Bootstrap %95 percentile confidence intervals, the probability of the Kolmogorov-Smirnov statistic (KSprob), from the 88 x 5 test score data (Efron & Tibshirani, 1993, p.62).

parameter	estimate	AN2.5	AN97.5	BSmean	BS2.5	BS97.5	KSprob
λ_1	.61	.44	.79	.61	.41	.77	.00
λ_2	.69	.56	.82	.68	.54	.80	.01
λ_3	.91	.86	.97	.91	.85	.97	.12
λ_4	.76	.66	.86	.76	.66	.85	.02
λ_5	.71	.59	.83	.71	.57	.82	.02
ψ_1	.62	.41	.84	.62	.40	.83	.85
ψ_2	.53	.35	.71	.53	.35	.71	.71
ψ_3	.16	.06	.27	.16	.05	.28	.41
ψ_4	.42	.27	.57	.42	.28	.57	.27
ψ_5	.50	.32	.67	.49	.32	.68	.44
MSE	.10	.06	.14	.10	.05	.15	.02
val	.95	.74	1.15	.95	.92	.98	.03
rel	.90	.85	.95	.90	.85	.95	.02

however, use the bootstrap estimate of the empirical distribution of the statistics to test for normality, to construct nonnormal confidence intervals, and to obtain logically consistent statistical inferences. This will be illustrated by the analysis of three empirical applications.

Efron and Tibshirani (1993, p.62) reported scores for 88 college students who took tests in: "Mechanics", "Vectors", "Algebra", "Analysis", and "Statistics". These data were analyzed by several authors (Beran & Srivastava, 1985; Efron & Tibshirani, 1993). The sample correlations between the observable variables are in the $[.39, .71]$ interval. The results from least squares single factor analysis are given in Table 2. From the probability of the KS statistic it can be observed that the empirical distribution of six out of 13 statistics do not differ significantly from normality at the 95% level. From the interval based on the asymptotic normality one cannot reject the hypothesis that the error variances and the MSE of prediction are positive. However, from the same analysis, the hypothesis that the validity is smaller than unity cannot be rejected. This logical inconsistency does not arise from statistical inferences based on percentile confidence intervals. In particular, from the latter it cannot be rejected that the error variances are positive, the MSE

Table 3: The least squares parameter estimate, lower (AN2.5) and upper (AN97.5) points of confidence intervals from asymptotic normality, Bootstrap mean (BSmean), lower (BS2.5) and upper (BS97.5) points of Bootstrap %95 percentile confidence intervals, the probability of the Kolmogorov-Smirnov statistic (KSprob), from the 62 x 3 Earthquakes data (Fuller, 1987, p.57).

parameter	estimate	AN2.5	AN97.5	BSmean	BS2.5	BS97.5	KSprob
λ_1	.88	.80	.96	.88	.77	.95	.00
λ_2	.84	.73	.95	.84	.71	.94	.00
λ_3	.92	.85	.99	.92	.84	.99	.10
ψ_1	.23	.08	.37	.23	.09	.40	.01
ψ_2	.30	.12	.48	.29	.12	.49	.15
ψ_3	.15	.03	.28	.15	.02	.29	.29
MSE	.08	.04	.12	.07	.02	.13	.01
val	.96	.70	1.22	.96	.93	.99	.00
rel	.92	.88	.96	.93	.87	.98	.01

of prediction is small but positive, and the validity is large but smaller than unity. This suggests the statistical inference that the factor is unobservable with a small amount of prediction error.

Fuller (1987, p.57) reported scores from the measures "Surface Wave", "Body Wave", "Trace" of 62 Alaskan earthquakes. The sample correlations between the observable variables are in the $[.74, .81]$ interval. The results of the analyses are given in Table 3. From the probability of the KS statistic it can be observed that the empirical distribution of seven out of nine statistics differs significantly from normality. From the confidence intervals based on normality one cannot reject the hypotheses that the error variances as well as the MSE of prediction are positive. However, from the same analysis the hypothesis that the validity equals unity cannot be rejected. This logical inconsistency does not arise from statistical inferences based on the percentile confidence intervals. In particular, from the latter one would statistically infer that the error variances are positive, the MSE of prediction is small but positive, and that the validity is large but smaller than unity.

Fuller (1987, p.65) reported 37 measures from "Aerial Photography", "Satellite Imagery", and "Personal Interview" of the area corn. The sample correlations between the observable variables are in the $[.83, .99]$ interval. The results of the analyses are given in Table 4. From the probability of the KS statistic the hypothesis that the statistics are normally distributed must be rejected. The estimated param-

Table 4: The least squares parameter estimate, lower (AN2.5) and upper (AN97.5) points of confidence intervals from asymptotic normality, Bootstrap mean (BSmean), lower (BS2.5) and upper (BS97.5) points of Bootstrap %95 percentile confidence intervals, the probability of the Kolmogorov-Smirnov statistic (KSprob), from the Corn area determination data (Fuller, 1987, p.65).

parameter	estimate	AN2.5	AN97.5	BSmean	BS2.5	BS97.5	KSprob
λ_1	.99	.98	1.01	.99	.98	1.00	.00
λ_2	.83	.72	.95	.83	.68	.92	.00
λ_3	.99	.98	1.00	.99	.98	1.00	.00
ψ_1	.01	-.02	.03	.01	.00	.04	.00
ψ_2	.31	.11	.50	.31	.15	.54	.00
ψ_3	.02	-.01	.05	.02	.00	.04	.00
MSE	.01	-.00	.02	.00	.00	.01	.00
val	1.00	.28	1.71	1.00	1.00	1.00	.00
rel	.99	.98	1.00	1.00	.99	1.00	.00

eter vector is extremely close to the boundary of the parameter set and the sample size is "small". Because several confidence intervals based on asymptotic normality are partially outside the parameter set and the non-normality of the statistics, it seems better to base the statistical inferences on the bootstrap estimates. From these it can be statistically inferred that the error variances of "Aerial Photography" and "Personal Interview" are zero, whereas that of "Satellite Imagery" is positive. Consequently, the hypotheses of zero MSE of prediction and unit validity cannot be rejected. Hence, we would statistically infer that the factor is observable.

8. Conclusions and discussion

The condition that three loadings are bounded away from zero for the population vector and its estimate both being internal points of the parameter set, is necessary and sufficient for the regularity conditions of asymptotic normality to hold. In particular, the model is identified and the Jacobian matrix as well as the asymptotic variance matrix are well-conditioned. Least squares factor analysis is a practically feasible method for bootstrap (Monte Carlo) estimation, even when the population parameter is close to the boundary of the parameter set. Furthermore, bootstrap estimation makes it possible to test the normality of the empirical distribution of the statistics, to obtain non-normal confidence intervals within the parameter set, and to draw logically consistent statistical inferences.

A result from the Monte Carlo experiment is that the mean tends to underestimate the MSE of prediction, and tends to overestimate the validity and the reliability. Obviously a quick glance on the definitions reveals that the overestimation is a consequence of the underestimation of the error variances. Underestimation is a well-known phenomenon for estimates which depend on a first order Taylor expansion such as those from the delta theorem (Hall, 1986a; Efron, 1992).

Several bootstrap methods have been proposed in the literature. In particular, the bias corrected bootstrap yields second order correct confidence intervals (Efron, 1987), but requires a least favorable distribution. The latter is absent in the current context. Moreover, its optimality properties depend on the estimation of the acceleration constant which is mathematically as well as practically complicated. The iterated bootstrap (Beran & Ducharme, 1991), where resamples are resampled, seems currently computationally too expensive for iterative estimation. Furthermore, the percentile-t bootstrap method is second order correct, but requires the statistic to be studentized to have a symmetric distribution and to have a stable estimated variance (Hall, 1988a). Ichikawa and Konishi (1995), however, conclude for maximum likelihood estimation that the symmetry requirement is often not met in practice. Furthermore, the stability of the estimated variances depends on the accuracy of estimates for $\boldsymbol{\Omega}_o$, $\boldsymbol{\Delta}_o$, and $\boldsymbol{\delta}_o$. In particular, it should be noted that $\boldsymbol{\Omega}_o$ is a rather intricate function of the first, second, and fourth order multivariate moments (see e.g. Hsu, 1949; Browne & Shapiro, 1986), and that stable estimates of higher order moments require the sample size to be "large" (cf. Kendall & Stuart, 1977, p.249).

Efron (1981) concluded from an extensive comparison between genuinely nonparametric methods of correlation estimation that the bootstrap performs best. Furthermore, under the assumption of finite fourth order moments, the bootstrap estimates the distribution of a statistic with probability one so that the percentile method yields asymptotically correct confidence intervals with probability one (e.g. Bickel & Freedman, 1981; Beran & Srivastava, 1985). Hence, the confidence intervals from asymptotic normality and percentile bootstrapping are asymptotically equal. Furthermore, the percentile bootstrap has the built in property to correct for non-normality of the empirical distribution. To base such corrections on Edgeworth or Cornish-Fisher expansions would be analytically tedious (Hall, 1983;1986a;1988b), and may lead to overcorrection when the sample size is not sufficiently large (Hall, 1986a).

Some remarks to the large sample concept of asymptotic efficiency seem in order, since it is well-known that least squares estimation is not asymptotically efficient. The asymptotic efficiency completely depends on the asymptotic normality. Obviously, in the current context, where the statistics are continuous functions of the sample correlations, accurate convergence to the normal distribution may require the sample size to be "large" (e.g. Browne, 1982), in particular when the population vector is close to the boundary of the parameter set. Obviously, testing the normality of the empirical distribution of the statistics is helpful in reducing the uncertainty whether the sample size is sufficiently "large".

References

Anderson, T.W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium*, 5, 111-150.

Bentler, P.M. & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P.R. Krishnaiah (Ed.) *Multivariate Analysis - VI Proceedings of the sixth international symposium on multivariate analysis* pp. 9-42. Amsterdam : Elsevier.

Beran, R. & Ducharme, G.R. (1991). *Asymptotic theory for bootstrap methods in statistics.* Montreal: Centre de recherche mathématique.

Beran, B. & Srivastava, M.S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 13, 95-115.

Bickel, P.J. & Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9, 1196-1217.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in Lisrel Maximum Likelihood estimation. *Psychometrika*, 50, 229-242.

Browne, M.W. (1974). Generalized least squares estimates in the analysis of covariance structures. *South African Statistical Journal*, 8, 1-24.

Browne, M.W. (1982). Covariance structures. In *Topics in Multivariate Analysis*

(D.M. Hawkins, ed.) pp.72-141. Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Browne, M.W. & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra and its Applications*, 52, 169-176.

Browne, M.W. & Shapiro, A. (1991). Invariance of covariance structures under groups of transformations. *Metrika*, 38, 345-355.

Copson, E.T. (1968). *Metric Spaces.* Cambridge: Cambridge University Press.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton: University Press.

Efron, B (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7, 1-26.

Efron, B. (1981). Non-parametric estimates of standard error: The Jackknife, the bootstrap and other methods. *Biometrika*, 68, 589-599.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-185.

Efron, B. (1992). Six questions raised by the bootstrap. In Lepage, R and Billard (Eds.): *Exploring the limits of Bootstrap.* New-York: Wiley.

Efron, B. & Tibshirani, R.F. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall

Ferguson, T.S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics, 29*, 1046-1062.

Fuller, W.A. (1987). *Measurement error models.* New York: Wiley.

Golub, G.H. & Van Loan, C.F. (1983). *Matrix Computations.* Baltimore: The John Hopkins University Press.

Hall, P. (1983). Inverting an Edgeworth Expansion. *The Annals of Statistics*, 11, 569-576.

Hall, P. (1986a). On the bootstrap and confidence intervals. *The Annals of Statis-*

tics, 14, 1431-1452.

Hall, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, 14, 1453-1462.

Hall, P. (1988a). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, 16, 927-953.

Hall, P. (1988b). Rate of convergence in bootstrap approximations. *The Annals of Statistics*, 16, 1665-1684.

Hsu, P.L. (1949). The limiting distribution of functions od sample means and application to testing hypotheses. J. Neyman (Eds.) *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (pp. 359-402). Berkeley: University of California Press.

Ichikawa, M. & Konishi, S. (1995). Application of the bootstrap methods to factor analysis. *Psychometrika*, 60, 77-94.

Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40, 633-643.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Jöreskog, K.G. & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70, 631-639.

Kendall, M. & Stuart, A. (1977). *The advanced Theory of Statistics* (Vol 1., 4th ed.). London: Griffin.

Krijnen, W.P., Dijkstra, T.K., & Gill, R.D. (1998). Conditions for factor (in)determinacy in factor analysis. *Psychometrika*, 63, 359-367.

Krijnen, W.P., Kroonenberg, P. M., & Dijkstra, T.K. (1998). Detecting diverging parameter estimates of the PARAFAC model. *Submitted for Publication.*

Lawley, D.N., & Maxwell, A.E. (1971). *Factor analysis as a statistical method.* Durban: Lawrence Erlbaum.

Lord, M. & Novick, M.R. (1968). *Statistical theories of mental test scores.* Massachusetts: Addison-Wesley.

Luenberger, D.G. (1969). *Optimization by vector space methods.* New York: John

Wiley.

Magnus, J.R. and Neudecker, H. (1991). *Matrix differential calculus with applications in statistics and economics.* Chichester:John Wiley.

Rao, C.R. (1973). *Linear statistical inference and its applications.* New York: John Wiley.

Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). New York: McGraw-Hill.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* New York: John Wiley.

Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 84-88.

Spearman, C. (1904). 'General Intelligence', objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Zegers, F.E. & Ten Berge, J.M.F. (1983). A fast and simple computational method of minimal residual factor analysis. *Multivariate Behavioral Research*, 18, 331-340.

Ontvangen: 10-07-1997
Geaccepteerd: 20-04-1999