1.6 CONDITIONS FOR FACTOR (IN)DETERMINACY IN FACTOR ANALYSI

Wim P. Krijnen[1], Theo K. Dijkstra[2] and Richard D. Gill[3]

[1,2]University of Groningen
[3]University of Utrecht

[1]The first author is obliged to the Department of Economics for their post-doc grant. The current address of Wim Krijnen is Lisdodde 1, 9679 MC Scheemda, The Netherlands.
[2]Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands.
[3]Department of Mathematics, University of Utrecht, P.O. Box 80010, 3508 TA Utrecht, The Netherlands.
The authors are obliged to Willem Schaafsma and the reviewers for useful comments.

CONDITIONS FOR FACTOR (IN)DETERMINACY IN FACTOR ANALYSIS

Abstract

The subject of factor indeterminacy has a vast history in factor analysis (Wilson, 1928; Lederman, 1938, Guttman, 1955). It has lead to strong differences in opinion (Steiger, 1979). The current paper gives necessary and sufficient conditions for observability of factors in terms of the parameter matrices and a finite number of variables. Five conditions are given which rigorously define indeterminacy. It is shown that (un)observable factors are (in)determinate. Specifically, the indeterminacy proof by Guttman (1955) is extended to Heywood cases. The results are illustrated by two examples and implications for indeterminacy are discussed.
Keywords: Indeterminacy, Heywood cases, mean squared error, factor score prediction.

After Spearman (1904) proposed the model for factor analysis, it was shown under the assumption of positive definite error variance that a certain indeterminacy exists (Wilson, 1928; Lederman, 1938; Guttman, 1955). Guttman (1955) proposed a measure for factor indeterminacy which was criticized by McDonald (1974) and defended by Elffers, Bethlehem, and Gill (1978). This may illustrate that the subject of factor indeterminacy has lead to strong differences in opinion (Schönemann & Wang, 1972; McDonald, 1977; Steiger, 1979). The conjecture (Spearman, 1933) that factor indeterminacy vanishes when the number of loadings (bounded away from zero) per factor goes to infinity was proven by Guttman (1955). Similar sufficient conditions for the least squares predictor to converge in quadratic mean to the unique common factor were given by Williams (1978) and Kano (1986). An extension of such conditions to multiple-factor factor analysis is given by Schneeweiss and Mathes (1995). McDonald (1974) has pointed out that it is unclear which sampling process is implied by the indeterminate factor model. Thomson (1950, p. 372) conjectured that zero error variances are part of a sufficient condition for factors to be determinate.

(In)determinacy has not been shown to exist under the condition that an error variance equals zero. We briefly review some of the issues associated with zero estimates for error variances. It has been empirically demonstrated by Jöreskog (1967) that such "Heywood" (1931), "improper", or "boundary" cases occur frequently when constrained factor analysis is applied to prevent negative error variances. Other procedures for constrained factor analysis have been proposed on the basis of Gauss-Seidel iteration (Howe, 1955; Bargmann, 1957; Browne, 1968), modified Gauss-Newton (Lee, 1980), and alternating least squares (Ten Berge & Nevels, 1977; Krijnen, 1996).

If a solution exists (cf. Krijnen, 1997a), estimates of the parameters can generally be obtained as solutions which optimize certain functionals. Unconstrained solutions with negative error variances, however, are "inadmissible" in the sense of not being a member of the parameter set. The inadmissibility may be caused by sampling fluctuations (Browne, 1968; van Driel, 1978) or by a zero population error variance. Furthermore, problems of non-convergence have been reported for unconstrained factor analysis (Jöreskog, 1967; Boomsma, 1985) but not for constrained factor analysis. Obviously, statistical inference is impossible on the basis of non-convergent, suboptimal, or inadmissible solutions. Moreover, an estimate which optimizes a functional for constrained factor analysis violates standard regularity conditions for statistical inference if it is not an internal point of the parameter set (Ferguson, 1958; Browne, 1984). Hence, when in an application the constraints appear to be active in yielding a Heywood case, nonstandard estimation theory is required (cf. Shapiro, 1986; Dijkstra, 1992).

The purpose of the current paper is to give necessary and sufficient conditions for the factors to be observable, to give conditions for indeterminacy, and to show that (un)observable factors are (in)determinate. The latter extends Guttman's (1955) proof for indeterminacy to Heywood cases. Two sampling processes are given and the implications of indeterminacy are discussed.

Definitions

The model for factor analysis assumes that the observations are generated by

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

where X is the random vector of order p with observed scores on the variables, $\mathcal{E}[X] = \mu_o$ its expectation, F the random vector with factor scores of order m, E the unobservable random error vector of order p, and Λ_o the loadings matrix of order p by m. Without loss of generality it will be assumed that $\mu_o = o$ and that the factors are standardized such that $\mathrm{Var}[F] = \Phi_o$ is the factor correlations matrix. It will furthermore be assumed that $\mathcal{E}[F] = o$, $\mathcal{E}[E] = o$, $\mathrm{Cov}[\boldsymbol{F}, \boldsymbol{E}] = \boldsymbol{O}$, and $\mathcal{E}[EE'] = \Psi_o$ diagonal. It follows that

$$\boldsymbol{\Sigma}_o = \boldsymbol{\Lambda}_o \boldsymbol{\Phi}_o \boldsymbol{\Lambda}_o' + \boldsymbol{\Psi}_o, \tag{2}$$

where $\Sigma_o = \mathrm{Var}[X]$ (Lawley & Maxwell, 1971). Throughout it will be assumed that $\mathrm{rank}(\Lambda_o) = m$ and Φ_o positive definite, so that $\mathrm{rank}(\Sigma_o) \geq m$. For notational brevity, the population matrices Λ_o, Φ_o, Ψ_o and their estimates $\widehat{\Lambda}$, $\widehat{\Phi}$, $\widehat{\Psi}$ will be denoted by the mathematical variables Λ, Φ, Ψ when distinctions between these do not matter.

It may be noted that (2) implies that Λ and Ψ are in the column space of Σ so that we have $\Sigma\Sigma^+(\Lambda, \Psi) = (\Lambda, \Psi)$ (e.g. Magnus & Neudecker, 1991, p. 58), where $^+$ denotes the unique Moore-Penrose inverse (Penrose, 1955).

Prediction by Projection

The existence of the various (co)variances allows us to define the inner product as the covariance between two random variables. Consequently, for the purpose of predicting F by $\widehat{F}$ from X we have

$$||\boldsymbol{F} - \widehat{\boldsymbol{F}}||^2 = \mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}] = \mathrm{MSE}[\widehat{\boldsymbol{F}}] = \boldsymbol{\mathcal{E}}\left[(\boldsymbol{F} - \widehat{\boldsymbol{F}})(\boldsymbol{F} - \widehat{\boldsymbol{F}})'\right]. \tag{3}$$

From the classical projection theorem (Luenberger, 1969, p. 51) it follows that for $\widehat{F}$ to satisfy $\|F - \widehat{F}\| \leq \|F - A'X\|$ for all $A'X$, it is necessary and sufficient that

$F-\widehat{F}$ is orthogonal to the space spanned by X. The latter condition is equivalent to $\boldsymbol{O}=\operatorname{Cov}[\boldsymbol{X},\, \boldsymbol{F}-\widehat{\boldsymbol{F}}]=\boldsymbol{\Lambda\Phi}-\boldsymbol{\Sigma A}$, which is, due to $\Sigma\Sigma^{+}\Lambda=\Lambda$, equivalent to $A=\Sigma^{+}\Lambda\Phi+N$, where N is orthogonal to Σ. For notational brevity we set $N=O$. Thus by taking $\widehat{F}=\Phi\Lambda'\Sigma^{+}X$ we have obtained the orthogonal projection $F-\widehat{F}$ of F onto the space spanned by X. When the dimension, rank$(\boldsymbol{\Sigma})$, of the space spanned by X equals p, then $N=O$ and the representation of $\widehat{F}$ in terms of X is unique (Luenberger, 1969, p. 51). A predictor $\widehat{F}$ may be called best linear, in the sense of Löwner's (1934) partial matrix order, when $\text{MSE}[\widehat{F}]\leq\text{MSE}[A'X]$ for all linear predictors $A'X$, which means that $\text{MSE}[A'X]-\text{MSE}[\widehat{F}]$ is positive semi definite (cf. Krijnen, Wansbeek, & Ten Berge, 1996). For the error of prediction $F-\widehat{F}$ we have

$$\operatorname{Var}[\boldsymbol{F}-\widehat{\boldsymbol{F}}]=\boldsymbol{\Phi}-\boldsymbol{\Phi\Lambda'\Sigma^{+}\Lambda\Phi}. \tag{4}$$

Obviously, the right hand side is non-negative definite since it is a variance matrix. It will be said that the jth factor F_j is *observable* if $\boldsymbol{F_j}=\boldsymbol{a_j'X}$ almost surely (a.s.), where a_j is column j of A. Obviously, the condition in this definition is equivalent to the condition that the jth diagonal element of $\operatorname{Var}[\boldsymbol{F}-\widehat{\boldsymbol{F}}]$ is equal to zero. However, what the definition means in terms of the parameter matrices is far from transparent. Below we will characterize observability via conditions on the parameter matrices being of finite order.

Conditions for Observable Factors

By multiplications with permutations matrices it follows without loss of generality that any order in the elements of F and in those of E can be arranged for. Hence, when Ψ contains p_1 zero diagonal elements, it will be understood that its first p_1 diagonal elements are zero and that its remaining diagonal elements are positive. Consider the partitions $\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X_1}\\ \boldsymbol{X_2}\end{bmatrix}$, $\boldsymbol{\Lambda}=\begin{bmatrix}\boldsymbol{\Lambda_1}\\ \boldsymbol{\Lambda_2}\end{bmatrix}$, $\boldsymbol{E}=\begin{bmatrix}\boldsymbol{E_1}\\ \boldsymbol{E_2}\end{bmatrix}$, such that $\operatorname{Var}[E_1]\equiv\Psi_{11}=O$ and $\operatorname{Var}[E_2]\equiv\Psi_{22}$ positive definite. Let λ_j be column j of Λ_1 and let Λ_{-j} have its jth column equal to zero and its other columns equal to those of Λ_1. Thus λ_j not in span(Λ_{-j}) is equivalent to rank(Λ_{-j})+1=rank(Λ_1).

Result 1. The factors $F_1,..,F_{m_1}$ are observable if and only if the first $p_1\geq m_1$ diagonal elements in Ψ are zero and λ_j is not in span(Λ_{-j}), for $j=1,..,m_1$.

Proof. (Necessity) Consider the partitions $\boldsymbol{F}=\begin{bmatrix}\boldsymbol{F_1}\\ \boldsymbol{F_2}\end{bmatrix}$ and $A=[A_1\ A_2]$, where F_1' and A_1 have m_1 columns. Assume $A_1'X=F_1$. Let Φ_{11} be the m_1 by m_1 left upper submatrix of Φ. Because Φ positive definite, rank(Φ_{11}) $=m_1$. Then $\Phi_{11}=\operatorname{Var}[F_1]=\operatorname{Var}[A_1'X]=A_1'\Sigma A_1$, and rank$(\Sigma)$ $\geq m$ implies that rank(A_1) $=m_1$. From $A_1'X=F_1$, $\mathcal{E}[F]=o$, $\mathcal{E}[E]=o$, and $\operatorname{Cov}[\boldsymbol{F},\,\boldsymbol{E}]=\boldsymbol{O}$, it follows upon Equation (1) that $\boldsymbol{O}=\operatorname{Cov}[\boldsymbol{E},\,\boldsymbol{F_1}]=\operatorname{Cov}[\boldsymbol{E},\,\boldsymbol{A_1'X}]=\boldsymbol{\Psi A_1}$ (a.s.). Hence, A_1 is in the nullspace of Ψ. Thus rank(Ψ) $\leq p-m_1$. From this and Ψ diagonal, it follows that Ψ has at least m_1 zero diagonal elements is necessary for the factors $F_1,..,F_{m_1}$ to be observable.

Let column j of A be partitioned by $\boldsymbol{a_j}=\begin{bmatrix}\boldsymbol{a_{1j}}\\ \boldsymbol{a_{2j}}\end{bmatrix}$, where a_{1j} is of order p_1. It will be useful to prove that $a_{2j}=o$ when F_j is observable. From the partition of E, $\mathcal{E}[E_1]=o$, $\operatorname{Var}[E_1]=O$, it follows that $E_1=o$ (a.s.). Hence, $\boldsymbol{F_j}=\boldsymbol{a_j'X}$ (a.s.), F

uncorrelated with E, and Equation (1), implies that

$$\mathrm{Cov}[\boldsymbol{a}_j'\boldsymbol{X},\,\boldsymbol{E}] = \boldsymbol{a}_j'\mathrm{Cov}[\boldsymbol{\Lambda F}+\boldsymbol{E},\,\boldsymbol{E}] = \boldsymbol{a}_j'\begin{bmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Psi_{22}} \end{bmatrix} = \boldsymbol{o}'.$$

Hence, $a_{2j} = o$ follows from $\mathrm{Var}[E_2] = \Psi_{22}$ positive definite.

Assume $a_{2j} = o$, the first $p_1 \geq m_1$ diagonal elements of Ψ zero, and $\lambda_j \in \mathrm{span}(\Lambda_{-j})$, for a j $(1 \leq j \leq m_1)$. From the partition of E, $\mathcal{E}[E_1] = o$, $\mathrm{Var}[E_1] = O$, it follows that $E_1 = o$ (a.s.). Hence, (1) implies that $X_1 = \Lambda_1 F = \lambda_j F_j + \Lambda_{-j}F$. Obviously, $a_{2j} = o$ implies $a_j'X = a_{1j}'X_1$. From $\lambda_j \in \mathrm{span}(\Lambda_{-j})$, it follows that $\lambda_j = \Lambda_{-j}\Lambda_{-j}^{+}\lambda_j$. Hence, there is no vector a_{1j} such that $a_{1j}'\Lambda_{-j} = o$ and $a_{1j}'\lambda_j = 1$. This completes the proof for the necessity of the condition for the factors $F_1, .., F_{m_1}$ to be observable.

(Sufficiency) Assume that the first $p_1 \geq m_1$ diagonal elements of Ψ are zero and that λ_j is not an element of $\mathrm{span}(\Lambda_{-j})$, for a j $(1 \leq j \leq m_1)$. If $M_j = I - \Lambda_{-j}\Lambda_{-j}^{+}$, then it is the orthogonal projection matrix that projects vectors onto the ortho-complement column subspace of Λ_{-j}. It follows immediately that $M_j = M_j'$ and $M_j\Lambda_{-j} = O$. Because λ_j is not an element of $\mathrm{span}(\Lambda_{-j})$, there is no vector b_j such that $\lambda_j = \Lambda_{-j}b_j$. Hence, $M_j\lambda_j = \lambda_j - \Lambda_{-j}\Lambda_{-j}^{+}\lambda_j \neq o$. Then by taking $\boldsymbol{a_{1j}} = \left(\boldsymbol{\lambda_j' M_j \lambda_j}\right)^{\boldsymbol{-1}}\boldsymbol{M_j\lambda_j}$, using the properties for M_j, we obtain $a_j'X = a_{1j}'\lambda_j F_j = F_j$. Because the reasoning holds for $j = 1, .., m_1$, the sufficiency of the condition follows. This completes the proof.

Some remarks seem in order. The condition in Result 1 is general in the sense that it holds for Heywood cases and for singular Σ matrices. The necessary condition is new. The condition in Result 1 relates observable factors to the parameter matrices for a finite number of variables. Provided that the first $p_1 \geq m_1$ diagonal elements in Ψ are zero, a simpler but stronger condition is $\mathrm{rank}(\boldsymbol{\Lambda_1}) = \boldsymbol{m}$. When this stronger condition holds, it also holds for all rotations of Λ. Finally, λ_j is not a member of $\mathrm{span}(\Lambda_{-j})$ when $\lambda_j'\Lambda_{-j} = o$.

Conditions for Indeterminacy

To give conditions under which indeterminacy exists, let $\widetilde{F}$, $\widetilde{E}$ be a factor, error vector, respectively. We have

Condition 1: $\mathcal{E}[\widetilde{F}', \widetilde{E}']' = o$.
Condition 2: $\mathrm{Var}[\widetilde{\boldsymbol{F}}] = \boldsymbol{\Phi}$ and $\mathrm{Var}[\widetilde{\boldsymbol{E}}] = \boldsymbol{\Psi}$.
Condition 3: $\mathrm{Cov}[\widetilde{\boldsymbol{E}},\,\widetilde{\boldsymbol{F}}] = \boldsymbol{O}$.
Condition 4: $\boldsymbol{X} = [\,\boldsymbol{\Lambda} \;\; \boldsymbol{I}\,]\begin{bmatrix} \widetilde{\boldsymbol{F}} \\ \widetilde{\boldsymbol{E}} \end{bmatrix}$.
Condition 5: $F \neq \widetilde{F}$ and $E \neq \widetilde{E}$.

Conditions 1 through 3 hold for $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$ as defined previously. Condition 4 ensures that the basic model equation holds for the same observable variables as those in Equation (1). Condition 4 implies that the loadings are fixed, so that $\widetilde{F}$ cannot be a rotation of F. This distinguishes indeterminacy from rotational indeterminacy. Condition 5 ensures that $\begin{bmatrix} \widetilde{\boldsymbol{F}} \\ \widetilde{\boldsymbol{E}} \end{bmatrix}$ differs from $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$.

Random Variables for which the Conditions Hold

It will now be shown that there are random variables which satisfy the five conditions. Let

$$\widetilde{\boldsymbol{F}} = \widehat{\boldsymbol{F}} + \boldsymbol{Y} \tag{5}$$

and

$$\widetilde{\boldsymbol{E}} = \boldsymbol{\Psi\Sigma}^{+}\boldsymbol{X} - \boldsymbol{\Lambda Y}, \tag{6}$$

where the random variable Y satisfies $\mathcal{E}[Y] = o$, $\mathrm{Var}[\boldsymbol{Y}] = \boldsymbol{\Phi} - \boldsymbol{\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi}$, and $\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{O}$. We will start by showing that Condition 1 through 4 hold without further specifying Y for the moment.

It is clear that Condition 1 holds. Using that $\mathrm{Var}[\widehat{\boldsymbol{F}}] = \boldsymbol{\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi}$, and $\mathrm{Var}[\widetilde{\boldsymbol{F}}] = \mathrm{Var}[\widehat{\boldsymbol{F}}] + \mathrm{Var}[\boldsymbol{Y}]$, it follows that $\mathrm{Var}[\widetilde{\boldsymbol{F}}] = \boldsymbol{\Phi}$, so that the first part of Condition 2 holds. From $\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{O}$ it follows that

$$\begin{aligned}\mathrm{Var}[\widetilde{\boldsymbol{E}}] &= \mathrm{Var}[\boldsymbol{\Psi\Sigma}^{+}\boldsymbol{X}] + \mathrm{Var}[\boldsymbol{\Lambda Y}] \\ &= \boldsymbol{\Psi\Sigma}^{+}\boldsymbol{\Psi} + \boldsymbol{\Lambda\Phi\Lambda}' - \boldsymbol{\Lambda\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi\Lambda}'. \end{aligned} \tag{7}$$

From $\Sigma\Sigma^{+}(\Lambda, \Psi) = (\Lambda, \Psi)$, $\Sigma\Sigma^{+}\Sigma = \Sigma$ (Penrose, 1955), and $\Psi = \Sigma - \Lambda\Phi\Lambda'$, it follows that

$$\boldsymbol{\Psi\Sigma}^{+}\boldsymbol{\Psi} = \boldsymbol{\Sigma} - 2\boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Lambda\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi\Lambda}'.$$

Using this in (7) shows that $\mathrm{Var}[\widetilde{\boldsymbol{E}}] = \boldsymbol{\Psi}$. Hence, Condition 2 holds.

From (5), (6), $\mathrm{Cov}[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{O}$, $\mathrm{Var}[\boldsymbol{Y}] = \boldsymbol{\Phi} - \boldsymbol{\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi}$, $\widehat{F} = \Phi\Lambda'\Sigma^{+}X$, and $\Sigma\Sigma^{+}\Sigma = \Sigma$, it follows that

$$\mathrm{Cov}[\widetilde{\boldsymbol{E}}, \widetilde{\boldsymbol{F}}] = \boldsymbol{\Psi\Sigma}^{+}\boldsymbol{\Lambda\Phi} - \boldsymbol{\Lambda\Phi} + \boldsymbol{\Lambda\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda\Phi}. \tag{8}$$

Using that $\Psi = \Sigma - \Lambda\Phi\Lambda'$ and $\Sigma\Sigma^{+}\Lambda = \Lambda$, it follows that the right hand side of (8) is zero. Hence, Condition 3 holds.

To show that Condition 4 holds we shall use $\Psi\Psi^{+}E = E$. To see this let ψ_{jj}^{+} be element jj of Ψ^{+}. Then Ψ^{+} is uniquely defined by $\psi_{jj}^{+} = 0$ if $\psi_{jj} = 0$ and $\psi_{jj}^{+} = \frac{1}{\psi_{jj}}$ if $\psi_{jj} > 0$, $j = 1, .., p$. From $\mathcal{E}[E] = o$ and $\psi_{jj} = 0$, it follows that $E_j = 0$ (a.s.), so that $E = \Psi\Psi^{+}E$. Furthermore, from (5) and (6), it is immediate that

$$[\ \boldsymbol{\Lambda} \quad \boldsymbol{I}\ ]\begin{bmatrix} \widetilde{\boldsymbol{F}} \\ \widetilde{\boldsymbol{E}} \end{bmatrix} = \boldsymbol{\Sigma\Sigma}^{+}\boldsymbol{X}. \tag{9}$$

From this, substitution of $\Lambda F + \Psi\Psi^{+}E$ for X, using that $\Sigma\Sigma^{+}(\Lambda, \Psi) = (\Lambda, \Psi)$, and $\Psi\Psi^{+}E = E$, it follows that Condition 4 holds.

The orthogonal projection $F - \widehat{F}$ of F onto the space spanned by X suggests two choices for Y which are in the space spanned by $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$. Suppose that $Y = F - \widehat{F}$ (cf. Guttman, 1955; Elffers, Bethlehem, & Gill, 1978). Then Equation (5) implies $\widetilde{F} = F$. Equation (6) implies

$$\widetilde{\boldsymbol{E}} = \boldsymbol{\Psi\Sigma}^{+}\boldsymbol{X} - \boldsymbol{\Lambda F} + \boldsymbol{\Lambda\Phi\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{X} = \boldsymbol{\Sigma\Sigma}^{+}\boldsymbol{X} - \boldsymbol{\Lambda F}. \tag{10}$$

But, $\boldsymbol{X} = [\ \boldsymbol{\Lambda} \quad \boldsymbol{\Psi\Psi}^{+}\ ]\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$, and $\Psi\Psi^{+}E = E$, implies that $\Sigma\Sigma^{+}X = \Lambda F + E$. This and Equation (10), implies that $\widetilde{E} = E$. Hence, Condition 5 does not hold when $Y = F - \widehat{F}$. Nevertheless, this shows that the model, as it is formulated in (1), can be formulated in terms of $\widetilde{F}$ and $\widetilde{E}$. This will be useful in deriving the key property for Guttman's (1955) measure for factor indeterminacy. At this place it may also be noted that for an observable factor F_j it holds that $Y_j = F_j - \widehat{F}_j = 0$, so that (5) implies $F_j = \widetilde{F}_j = \widehat{F}_j$. Hence, Condition 5 does not hold for observable factors. Therefore, observable factors are not indeterminate and may thus be called determinate.

Suppose F unobservable and $Y = \widehat{F} - F$. The supposition $\widetilde{F} = F$ leads to a contradiction, as follows. Using $\widetilde{F} = F$, substitution of $\widehat{F} - F$ for Y in (5) implies $F = \widehat{F}$. That is, F observable, which is contradictory.

Similarly, $\widetilde{E} = E$ implies that $\boldsymbol{O} = \mathrm{Cov}[\widetilde{\boldsymbol{E}}, \boldsymbol{F}]$. This, $\Psi = \Sigma - \Lambda\Phi\Lambda'$, and (6), implies that $\boldsymbol{O} = \boldsymbol{\Phi} - \boldsymbol{\Phi\Lambda'\Sigma^{+}\Lambda'\Phi} = \mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}]$. This contradicts the supposition F unobservable. We conclude that Condition 1 through 5 hold when $Y = \widehat{F} - F$. This generalizes Guttman's (1955) sufficient condition for indeterminacy to Heywood cases.

Sampling

It will now be shown how the model equations can be used to sample observable variables from a distribution, in particular, from the normal distribution. The sampling process indicates how "Nature" may proceed when observable variables are constructed according to the model for factor analysis.

Before going into these processes it will be convenient to note that since, $\widehat{F}$, $Y = \widehat{F} - F$, and X are in the columnspace of $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$, it follows that $\begin{bmatrix} \widetilde{\boldsymbol{F}} \\ \widetilde{\boldsymbol{E}} \end{bmatrix}$ is in the columnspace of $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$. More specifically, it can be verified that

$$\begin{bmatrix} \widetilde{\boldsymbol{F}} \\ \widetilde{\boldsymbol{E}} \end{bmatrix} = \begin{bmatrix} (2\boldsymbol{\Phi\Lambda'\Sigma^{+}\Lambda} - \boldsymbol{I}) & 2\boldsymbol{\Phi\Lambda'\Sigma^{+}} \\ (\boldsymbol{\Psi\Sigma^{+}} - \boldsymbol{\Lambda\Phi\Lambda'\Sigma^{+}} + \boldsymbol{I})\boldsymbol{\Lambda} & (\boldsymbol{\Psi} - \boldsymbol{\Lambda\Phi\Lambda'})\boldsymbol{\Sigma^{+}} \end{bmatrix} \begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}. \tag{11}$$

The first process is according to the factor model as it is given by Equation (1). In particular, let n independent vectors $\left[\begin{pmatrix} \boldsymbol{f}_1 \\ \boldsymbol{e}_1 \end{pmatrix}, .., \begin{pmatrix} \boldsymbol{f}_n \\ \boldsymbol{e}_n \end{pmatrix}\right]$ be drawn from the normal distribution $N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{O} \\ \boldsymbol{O}' & \boldsymbol{\Psi} \end{bmatrix}\right)$. Then take $x_i = \Lambda f_i + e_i$, for $i = 1, .., n$.

The second process can be based on the sample $\left[\begin{pmatrix} \boldsymbol{f}_1 \\ \boldsymbol{e}_1 \end{pmatrix}, .., \begin{pmatrix} \boldsymbol{f}_n \\ \boldsymbol{e}_n \end{pmatrix}\right]$, just obtained. Premultiplication of $\begin{pmatrix} \boldsymbol{f}_i \\ \boldsymbol{e}_i \end{pmatrix}$, for $i = 1, .., n$, with the matrix in Equation (11) yields the sample $\left[\begin{pmatrix} \widetilde{\boldsymbol{f}}_1 \\ \widetilde{\boldsymbol{e}}_1 \end{pmatrix}, .., \begin{pmatrix} \widetilde{\boldsymbol{f}}_n \\ \widetilde{\boldsymbol{e}}_n \end{pmatrix}\right]$. Now take, according to Condition 4, $x_i = \Lambda\widetilde{f}_i + \widetilde{e}_i$, for $i = 1, .., n$.

It follows from Kolmogorov's theorem that $\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$ converges to Σ with probability 1 as $n \to \infty$ (Serfling, 1980, p. 27, Th. B).

Issues of Prediction

It is well-known that if $\boldsymbol{\mathcal{L}}\begin{pmatrix} \boldsymbol{F} \\ \boldsymbol{X} \end{pmatrix} = \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\Phi\Lambda'} \\ \boldsymbol{\Lambda\Phi} & \boldsymbol{\Sigma} \end{bmatrix}\right)$, then

$$\boldsymbol{\mathcal{L}}(\boldsymbol{F}|\boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{N}(\boldsymbol{\Phi\Lambda'\Sigma^{+}x}, \boldsymbol{\Phi} - \boldsymbol{\Phi\Lambda'\Sigma^{+}\Lambda\Phi}) \tag{12}$$

e.g. Anderson(1984, p. 37). Thus $\Phi - \Phi\Lambda'\Sigma^{+}\Lambda\Phi$ is the dispersion of the prediction error $F - \widehat{F}$. Obviously, $\boldsymbol{\mathcal{L}}\begin{pmatrix} \boldsymbol{F} \\ \boldsymbol{X} \end{pmatrix} = \boldsymbol{\mathcal{L}}\begin{pmatrix} \widetilde{\boldsymbol{F}} \\ \boldsymbol{X} \end{pmatrix}$, implies that exactly the same result holds for indeterminate factors.

In case two researchers have a different opinion on which of the sampling procesess is the correct one, their degree of disagreement can be measured by the correlation between the factors $\widehat{F}_j + Y_{j1}$ and $\widehat{F}_j + Y_{j2}$ (Guttman, 1955). A lower bound for the correlation between these can be obtained as follows. Let u_j be column j from the identity matrix. The Cauchy-Schwarz inequality implies

$$\mathrm{Cov}[\widehat{\boldsymbol{F}}_j + \boldsymbol{Y}_{j1}, \widehat{\boldsymbol{F}}_j + \boldsymbol{Y}_{j2}] = \boldsymbol{u}_j'\boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{u}_j + \mathrm{Cov}[\boldsymbol{Y}_{j1}, \boldsymbol{Y}_{j2}] \geq$$

$$\boldsymbol{u}_j'\boldsymbol{\Phi}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{+}\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{u}_j - (\mathrm{Var}[\boldsymbol{Y}_{j1}]\mathrm{Var}[\boldsymbol{Y}_{j2}])^{\frac{1}{2}}.$$

Thus the minimum correlation occurs when $Y_1 = -Y_2$. Taking $Y_1 = Y$, using that $\widetilde{F} = F$ when $Y = F - \widehat{F}$, leads to, F and $\widetilde{\boldsymbol{F}} = 2\widehat{\boldsymbol{F}} - \boldsymbol{F}$, so that the minimum value equals

$$\mathrm{Cov}[\boldsymbol{F}, \widetilde{\boldsymbol{F}}] = \mathrm{Var}[\widehat{\boldsymbol{F}}] - \mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}]. \tag{13}$$

Obviously, F observable, implies $F = \widetilde{F}$ and $\mathrm{Cov}[\boldsymbol{F}, \widetilde{\boldsymbol{F}}] = \boldsymbol{\Phi}$.

Two Examples

To illustrate, at first glance counterintuitive facts, two examples will be given. The first shows that the factors may be indeterminate (unobservable) for singular Σ, and the second shows that the factors may be determinate (observable) for non-singular Σ. The matrices Σ in the examples are correlation matrices.

Example 1. Let ι be the vector with unit elements having suitable order, $\Phi = I_2$, $\iota'\Psi = (0, 0, 0, \frac{1}{2}, \frac{1}{2})$, so that $p = 5$ and $p_1 = 3$. Furthermore let

$$\boldsymbol{\Lambda} = \sqrt{\tfrac{1}{2}}\left[\ \text{[illegible]}\ \right].$$

Then, the condition in Result 1 does not hold since the two columns of Λ_1 are dependent, hence both factors are not observable. The matrix Σ has rank 3, so that it is singular. For completeness we mention that by (4) it is found that $\mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}] = \frac{1}{4}\left[\ -1 \quad -1\ \right]$.

Example 2. Let $\Phi = I_2$ and

$$\boldsymbol{\Lambda} = \begin{bmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \\ \sqrt{1/2} & 0 \\ \sqrt{1/2} & 0 \\ 0 & \sqrt{1/2} \end{bmatrix}, \quad \boldsymbol{\Psi\iota} = \begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1.00 & 0.00 & 0.57 & 0.57 & -0.42 \\ 0.00 & 1.00 & 0.42 & 0.42 & 0.57 \\ 0.57 & 0.42 & 1.00 & 0.50 & 0.00 \\ 0.57 & 0.42 & 0.50 & 1.00 & 0.00 \\ -0.42 & 0.57 & 0.00 & 0.00 & 1.00 \end{bmatrix},$$

so that $p = 5$ and $p_1 = 2$. Then, the condition in Result 1 holds since the first two diagonal elements of Ψ are zero and the columns of Λ_1 are independent. Hence, both factors are observable. All eigenvalues of Σ are larger than zero, so that it is non-singular. It may be noted that the correlations in Σ seem realistic with respect to empirical applications of the factor model. For completeness we mention that by (4) it is found that $\mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}] = \boldsymbol{O}$.

Conclusions and Discussion

Result 1 gives necessary and sufficient conditions in terms of the parameter matrices for the factors to be observable. The five conditions define indeterminacy rigorously and distinguish issues of indeterminacy from rotational indeterminacy. By extending Guttman's(1955) proof to Heywood cases, it follows that (un)observable factors are (in)determinate. The examples illustrate that Result 1 contains a construction device for population matrices useful for Monte Carlo Research. More specifically, Result 1 shows how to construct factors arbitrarily close to being (in)determinate. The latter can be accomplished for parameter points on or arbitrarily close to the boundary of the parameter set. In particular, this can be arranged for by choosing population error variances arbitrarily close to zero.

We have stressed that from well-known dimensionality type of conditions with respect to projection, the uniqueness of the best linear predictor is implied. Hence, the criterion in Equation (3) allows the predictor to be unique. There are various criteria in the literature on factor prediction which do not allow uniqueness. Examples are "reliability" (Jöreskog, 1971) or "validity" or multiple correlation (McDonald & Burr, 1967; Lord & Novick, 1968, p. 261; Muirhead, 1982, p. 165). Furthermore, there are factor score predictors in the literature which satisfy a certain constraint (Thurstone, 1935; Bartlett, 1937; Anderson & Rubin, 1956; Ten Berge, Krijnen, Wansbeek & Shapiro, 1997). These are, however, not best linear (Krijnen, Wansbeek, & Ten Berge, 1996).

Under certain regularity conditions, estimation procedures based on maximum likelihood or general method of moments yield estimates $\widehat{\Lambda}$, $\widehat{\Phi}$, $\widehat{\Psi}$ that converge with probability 1 to Λ_o, Φ_o, Ψ_o (Cramér, 1946, p. 500; Ferguson, 1958; Browne, 1984; Sen & Singer, 1993, p. 205). This implies that continuous functions of these, such as $\mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}]$, can be estimated with probability 1 (Serfling, 1980, p. 24). Furthermore, because functions such as $\mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}]$, are continuously differentiable with respect to the parameters (e.g. Magnus & Neudecker, 1991, p. 154), their asymptotic normality is obtainable (Serfling, 1980, p. 122). It may happen, in practice, that a diagonal element of $\mathrm{Var}[\boldsymbol{F} - \widehat{\boldsymbol{F}}]$ does not differ significantly from zero and that the estimated point does not differ significantly from a point for which the conditions of Result 1 hold. Such empirical cases exist for single-factor factor analysis (Krijnen, 1997b).

Condition 4 says that the observable variables are a weighted sum of the loadings and the error vector. The random variable Y, however, is orthogonal to the space spanned by the observable variables X, although it does correlate with its constituting variables $\begin{bmatrix} \boldsymbol{F} \\ \boldsymbol{E} \end{bmatrix}$. In addition, it can be seen from (5) and (6) that observable variables are used to define observable variables. These properties complicate the understanding of the model in which indeterminate factors are involved. Most scientists are willing to consider a more complicated model when there is some evidence in favor for it. However, the orthogonality of Y to the observable variables X implies that its linear prediction is useless. It is thus impossible to empirically investigate Y in the sense of relating it to the observable variables. For these reasons the possibility of providing evidence in favor of the indeterminate factor model is at least questionable.

Finally, it may be noted that Guttman's (1955) measure for factor indeterminacy is closely related to other measures (cf. Elffers, Bethlehem, & Gill, 1978). That is, (12) shows that Var[$\boldsymbol{F} - \widehat{\boldsymbol{F}}$] is the dispersion matrix which reveals the degree of uncertainty with respect to making valid inferences to cases. Hence, for the latter purpose it is desirable that the entries of the dispersion matrix are small. When this is the case, however, the entries of Var[$\widehat{\boldsymbol{F}}$] are large, so that the entries of Guttman's (1955) measure are large, see (13). Possible means to obtain this in practice are decreasing the number of factors or increasing the number of variables with large loadings (Schneeweiss & Mathes, 1995).

References

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

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

Bargmann, R.E. (1957). A study of independence and dependence in multivariate normal analysis. Chapel Hill: University of North Carolina, Institute of Statistics. *Mimeo Series*, 186.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.

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

Browne, M.W. (1968). A comparison of factor analytic techniques. *Psychometrika*, 33, 267-334.

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

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

Dijkstra, T.K. (1992). On statistical inference with parameter estimates on the boundary of the parameter space. *British Journal of Statistical and Mathematical Psychology*, 45, 289-309.

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

Elffers, H., Bethlehem, J. & Gill, R.D. (1978). Indeterminacy problems and the interpretation of factor analysis results. *Statistica Neerlandica*, 32, 181-199.

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.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *The British Journal of Statistical Psychology*, 8, 65-81.

Heywood, H.B. (1931). On finite sequences of real numbers. *Proc. Roy. Soc. London*, 134, 486-501.

Howe, W.G. (1955). *Some contributions to factor analysis.* Oak Ridge: Oak Ridge National Laboratory (Report nr. ORNL-1919).

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

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

Kano, Y. (1986). A condition for the regression predictor to be consistent in a single common factor model. *British Journal of Mathematical and Statistical Psychology*, 39, 221-227.

Krijnen, W.P. (1996). Algorithms for unweighted least squares factor analysis. *Computational Statistics and Data Analysis*, 21, 2, 133-147.

Krijnen, W.P., Wansbeek, T.J., & Ten Berge, J.M.F. (1996). Best Linear Estimators for Factor Scores. *Communications in Statistics: Theory and Methods*, 25, 3013-3025.

Krijnen, W.P. (1997a). A note on the parameter set for factor analysis models. *Linear Algebra and its Applications.*

Krijnen, W.P. (1997b). Using single factor-factor analysis as a measurement model. *Submitted for publication.*

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

Lederman, W. (1938). The orthogonal transformations of a factorial matrix into itself. *Psychometrika*, 3, 181-187.

Lee, S.Y. (1980). Estimation of covariance structure models with parameters subject to functional restraints. *Psychometrika*, 45, 309-324.

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

Löwner, K. (1934). Über monotone Matrixfunctionen. *Mathematisches Zeitschrift*, 38, 177-216.

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

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

McDonald, R.P., & Burr, E.J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika*, 32, 381-401.

McDonald, R.P. (1974). The measurement of factor indeterminacy. *Psychometrika*, 39, 203-222.

Muirhead R.J. (1982). *Aspects of multivariate statistical theory.* New York: John Wiley & Sons.

Penrose, R (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51, 406-413.

Schneeweiss, H & Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, 55, 105-124.

Schönemann, P.H. & Wang, M-M (1972). Some new results on factor indeterminacy. *Psychometrika*, 37, 61-91.

Sen P.K. & Singer, J.M. (1993). *Large sample methods in statistics.* New York: Chapman & Hall.

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

Shapiro, A. (1986). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72, 133-144.

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

Spearman, C. (1933). The uniqueness and exactness of g. *British Journal Psychology*, 24, 106-108.

Steiger, J.H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika*, 44, 157-167.

Ten Berge, J.M.F. & Nevels, K. (1977). A general solution to Mosier's oblique Procrustes problem. *Psychometrika*, 42, 593-600.

Ten Berge, J.M.F., Krijnen, W.P. Wansbeek, T.J. & Shapiro, A. (1997). Some new results on correlation preserving factor scores prediction methods. *Linear Algebra and its Applications.*

Thurstone, L.L. (1935). *The vectors of mind.* Chicago: University of Chicago Press.

Thomson, G.H. (1950). *The Factorial Analysis of Human Ability.* London: University Press.

Williams, J.S. (1978). A definition for the common-factor analysis model and the elimination of problems of factor score indeterminacy. *Psychometrika*, 43, 293-306.

Wilson, E.B. (1928). On hierarchical correlation systems. *Proceedings, National Academy of Science*, 14, 283-291.