PSYCHOMETRIKA—VOL. 37, NO. 1
MARCH, 1972

SOME NEW RESULTS ON FACTOR INDETERMINACY

PETER H. SCHÖNEMANN AND MING-MEI WANG

PURDUE UNIVERSITY

Some relations between maximum likelihood factor analysis and factor indeterminacy are discussed. Bounds are derived for the minimum average correlation between equivalent sets of correlated factors which depend on the latent roots of the factor intercorrelation matrix Ψ. Empirical examples are presented to illustrate some of the theory and indicate the extent to which it can be expected to be relevant in practice.

1. *Introduction*

The issue of factor indeterminacy is an issue of long standing [Wilson, 1928] but it was relatively dormant for several decades, with the notable exception of the important contributions by L. Guttman in the fifties (especially Guttman, [1955]). Only recently have several authors drawn attention to it again, among them Heermann [1964, 1966] and Schönemann [1971]. The latter presented a considerably simplified statement of some of the mathematical aspects of this issue which to some extent parallels, but in others is quite distinct from, the development given by Guttman. Historically these two approaches go back to E. B. Wilson [1928] and Thomson [1935], on the one hand, and to Piaggio [1933], on the other.

One way of approaching this issue is to state the factor model for a given set of factor variables and to inquire what class of transformations exists which carry the given random variables into another, equivalent set of random variables. This "transformation approach" is especially simple if the common factors are uncorrelated. It was used by Wilson [1928] and later elaborated upon by Thomson [1935] for the single common factor case and by Ledermann [1938] for the multiple common factor case. The advantage of this approach is that the geometry is intuitively straightforward. For a more recent statement of this geometry see Heermann [1964] and Schönemann [1971].

Another way of approaching this issue is by way of construction of the factor variables as sums of determinate and indeterminate parts. This "construction approach" has been used by Piaggio [1933] and later by Kestelman [1952] for the uncorrelated case and by Guttman [1955] for the correlated case. A more recent statement of this approach can also be found in Heermann [1966].

Quite recently Schönemann [1971] used the transformation approach to establish the simple but somewhat surprising fact that the minimum average correlation between equivalent, uncorrelated factors is a constant which depends on p, the number of observed variables, and on m, the number of common factors, but not on the pattern (A, U). That is to say, this average can be computed, on the assumption of a fit of the factor model, "without looking at the data," once m and p are known. While this result should be of some theoretical interest, its practical impact is limited because (i) the result was established only for the case of uncorrelated common factors and (ii) the minimum average correlation includes both common and unique factors.

In the present paper we examine, among other things, the effect of removing either one of these restrictions. In addition, we shall employ the "construction approach" of Piaggio and Guttman to investigate the factor indeterminacy issue in the fallible case, *i.e.*, when the model does not fit exactly. In order to communicate effectively, we need a working definition of the term "factor scores." This poses some semantical problems which we will discuss in more detail in Sec. 5. As a preliminary definition for the purposes of this paper we shall mean by "factor scores" two sets of numbers which satisfy all the strictures of the factor model in the sample once C, the covariance matrix estimate, satisfies them. We shall say "the factor model fits in the sample exactly" when C, the observed covariance matrix can be written

$$C = \hat{A}\hat{A}' + \hat{U}^2 \tag{1.1}$$

for some $p \times m$ $(m < p - 1)$ matrix $\hat{A}$ and some positive definite (p.d.) diagonal matrix $\hat{U}^2$. We shall call an $m \times N$ matrix X of numbers "(uncorrelated) common factor scores" and a $p \times N$ matrix Z of numbers "(standardized) unique factor scores" if they jointly satisfy, for a given (deviation) score matrix Y $(p \times N)$ and two matrices $\hat{A}$ $(p \times m)$, $\hat{U}^2$ $(p \times p$, p.d.; diagonal):

$$Y = \hat{A}X + \hat{U}Z \tag{1.2}$$

$$\begin{pmatrix} X \\ Z \end{pmatrix}(X', Z')/N = I_{p+m}\,, \quad \begin{pmatrix} X \\ Z \end{pmatrix}J = \phi_{p+m} \tag{1.3}$$

(where J is a vector of N ones and ϕ a vector of $p + m$ zeros),

$$YX'/N = \hat{A}, \qquad YZ'/N = \hat{U}. \tag{1.4}$$

An essential ingredient of such a definition is the requirement that the factor model "fit in the sample exactly," *i.e.*, that $YY'/N = C = \hat{A}\hat{A}' + \hat{U}^2$, exactly. Only if this is the case will "factor scores" X, Z, computed by the formulae in Kestelman [1952] and Guttman [1955] reproduce all the observable information (1.2), (1.4) and the stipulated covariance matrix (1.3) exactly. But in practice the observed covariance matrix $C = YY'/N$ will almost always be different from $\hat{\Sigma} = \hat{A}\hat{A}' + \hat{U}^2$, We are usually satisfied if the factor model "fits statistically" in the sample in the sense that we cannot reject the hypothesis of a fit in the population although there are some discrepancies in the sample.

In this paper we address ourselves to the problem of factor indeterminacy when the model fits only statistically. We are interested in measures of factor indeterminacy when $\hat{A}$, $\hat{U}^2$ have been obtained by the method of maximum likelihood. We shall find that this particular estimation method permits us to calculate measures of factor indeterminacy separately for each factor which are meaningful for the common factors even if the model does not fit in the sample exactly. We shall also find that these measures can be written down at once as simple functions of the latent roots of an eigenproblem which occurs in maximum likelihood factor analysis (MLFA). Moreover, it will be shown that a particular identifiability constraint which is often used in MLFA leads to a set of common factors which contains the most determinate and the least determinate of all factors obtainable upon orthogonal and oblique rotation. We shall also develop weak bounds for the minimum average correlation of equivalent factors when some of the common factors are correlated. Finally, a number of empirical analyses will be presented to illustrate the theory and to indicate the extent to which it may be expected to be relevant in practice.

2. *Factor Indeterminacy and Maximum Likelihood Factor Analysis*

Detailed accounts of the method of maximum likelihood factor analysis can be found in Anderson and Rubin [1956], Bargmann [1957], Browne [1968, 1969], Howe [1955], Jöreskog [1963, 1967], Lawley [1940, 1942], Lawley and Maxwell [1963], Rao [1955], and elsewhere. Here we simply state some of the results which are relevant for our discussion of the relations between factor indeterminacy and the method of maximum likelihood factor analysis (MLFA).

The conditional equations one obtains upon setting the derivative of the (log-) likelihood function (or, equivalently, of the determinant $|\hat{U}^{-1}(C - \hat{A}\hat{A}')\hat{U}^{-1}|$) equal to zero are

$$\hat{\Sigma}^{-1}(\hat{\Sigma} - C)\hat{\Sigma}^{-1}\hat{A} = \phi \tag{2.1}$$

$$\operatorname{diag} \hat{\Sigma}^{-1}(\hat{\Sigma} - C)\hat{\Sigma}^{-1} = \operatorname{diag} \phi \tag{2.2}$$

where

$$\hat{\Sigma} = \hat{A}\hat{A}' + \hat{U}^2 \tag{2.3}$$

is the maximum likelihood estimate (MLE) of Σ under the hypothesis that the factor model holds for exactly m common factors, and where

$$C = (c_{ij}). \qquad c_{ij} = \sum_{k}^{N} (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)/N \tag{2.4}$$

is the biased maximum likelihood estimate of Σ in the unconstrained parameter space. These equations define $\hat{A}$ up to a rotation (provided certain additional mild constraints are met which are discussed in Anderson and Rubin [1956]. For convenience we shall assume here they are met.) To remove the rotational indeterminacy one sometimes imposes the "identifiability constraint"

$$\hat{A}'\hat{U}^{-2}\hat{A} = D_c^2 = \text{diagonal} \tag{2.5}$$

on $\hat{A}$. For convenience we shall assume the diagonal elements c_j^2 in D_c^2 are ordered so that $c_1^2 \geq c_2^2 \geq, \cdots, \geq c_m^2$. This constraint can be imposed without loss of generality since a ("orthogonal") rotation does not affect

$$\hat{U}^2 = \text{diag}\,(C - \hat{A}\hat{A}'). \tag{2.6}$$

Thus, if A^* does not satisfy the constraint (2.5), so that $A^{*\prime}\hat{U}^{-2}A^* = M = LD_c^2L'$ is a symmetric but non-diagonal matrix M with orthogonal eigenvectors L, then $\hat{A} = A^*L$ will satisfy the diagonality constraint (2.5) for a suitable ordering of the columns in L.

This particular constraint is convenient because it can be used to express (2.1) in the form of an eigenproblem. If

$$\hat{U}^{-1}C\hat{U}^{-1} = (V_1\ ,\ V_2)\begin{bmatrix} D_{b1}^2 & \\ & D_{b2}^2 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = VD_b^2V' \tag{2.7}$$

is the eigendecomposition of the (unknown) matrix $\hat{U}^{-1}C\hat{U}^{-1}$ (so that $V = (V_1\ ,\ V_2)$ is orthonormal and $D_b^2 = \begin{bmatrix} D_{b1}^2 & \\ & D_{b2}^2 \end{bmatrix}$ is a diagonal matrix which contains the latent roots of $\hat{U}^{-1}C\hat{U}^{-1}$ in descending order), then the first m ("retained") roots in D_{b1}^2 and their associated eigenvectors in V_1 define the solution matrix $\hat{A}$ as

$$\hat{A} = \hat{U}V_1(D_{b1}^2 - I_m)^{1/2} \tag{2.8}$$

while the last $p - m$ ("rejected") roots in D_{b2}^2 can be used in a likelihood ratio test

$$\begin{aligned} T(X) &= -k \ln |E| \\ &= -k \sum_{i=m+1}^{p} \ln b_i^2 \end{aligned} \tag{2.9}$$

where

$$E = (e_{ij}) = \hat{U}^{-1}(C - \hat{A}\hat{A}')\hat{U}^{-1} = (V_1\ ,\ V_2)\begin{bmatrix} I_m & \\ & D_{b2}^2 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

and

$$k = N - (2p + 5)/6 - 2m/3 - 1$$

to assess the fit of the model statistically, if one is willing to assume an underlying multivariate normal distribution. (2.1), (2.2) can be arrived at without this assumption either within the context of a canonical correlation paradigm [Rao, 1955] or within the context of a partial correlation paradigm [Bargmann, 1957; Howe, 1955].

For our present purposes it is important to note that the conditional equations (2.1), (2.2) and the identifiability constraint (2.5) are independent of each other. Certain computer programs for MLFA, *e.g.*, M. Browne's [1968] which is based on a Gauss-Seidel iteration for a solution of (2.1), (2.2), do not employ the identifiability constraint (2.5). Conversely, the constraint (2.5) can be imposed on any factor pattern $\hat{A}$ for a set of uncorrelated common factors, whether $\hat{A}$ was obtained by the method of MLFA or not.

Now suppose we have drawn an N-fold sample on p tests and we have computed a $p \times N$ matrix of deviation scores $Y = (y_{ij})$ and the associated sample covariance matrix $C = YY'/N$. We have obtained a matrix $(p \times m)$ $\hat{A}$ and a p.d. diagonal matrix $\hat{U}^2$ which jointly satisfy (2.1), (2.2). We have found that the factor model fits statistically, but not exactly, *i.e.*, $\hat{\Sigma}$ in (2.3) is not identical to C in (2.4). We now proceed to calculate an $m \times N$ matrix of deviation scores $X = (x_{ij})$ and a $p \times N$ matrix of deviation scores $Z = (z_{kj})$ in accordance with the formulae in Kestelman [1952] or Guttman [1955]:

Given $\hat{A}$, $\hat{\Sigma} = \hat{A}\hat{A}' + \hat{U}^2$ we compute an arbitrary Gram factor P $(m \times m)$ of

$$I - \hat{A}'\hat{\Sigma}^{-1}\hat{A} = PP' \tag{2.10}$$

and a matrix of deviation scores S $(m \times N)$ which satisfies

$$SJ = \phi, \qquad SS'/N = I_m, \qquad SY'/N = \phi. \tag{2.11}$$

The matrix S can be constructed in various ways, for example by orthogonalizing the rows of a row centered matrix of random numbers S^* relative to the rows of Y and relative to each other. These two matrices P, S we use, in conjunction with Y, $\hat{A}$, $\hat{U}^2$ (and $\hat{\Sigma} = \hat{A}\hat{A}' + \hat{U}^2$) to compute

$$X = \hat{A}'\hat{\Sigma}^{-1}Y + PS \tag{2.12}$$

and

$$Z = \hat{U}\hat{\Sigma}^{-1}Y - \hat{U}^{-1}\hat{A}PS. \tag{2.13}$$

We now wish to check the properties of these two matrices under three different conditions:

(i) when the model does not fit exactly in the sample, *i.e.*, when $C \neq \hat{\Sigma}$ and $\hat{A}$, $\hat{U}^2$ do not satisfy (2.1), (2.2),

(ii) when the model does not fit exactly in the sample, *i.e.*, when $C \neq \hat{\Sigma}$, but $\hat{A}$, $\hat{U}^2$ have been obtained by the method of MLFA, so that (2.1), (2.2) are satisfied, and

(iii) when the model fits in the sample exactly, *i.e.*, when $C = \hat{\Sigma}$, regardless of the method of estimation used.

The last condition is assumed in the articles by Kestelman [1952] and Guttman [1955]. They show that in this case all observable score and covariance information (*i.e.*, Y, C, $\hat{A}$, $\hat{U}^2$) and all strictures of the factor model are reproduced precisely by X, Z in the sample, although evidently neither X nor Z are unique, since P, for example, is determined only up to an $m \times m$ rotation G $(GG' = I)$ by (2.10).

The other two conditions are less obvious, although one may anticipate that under (i) none of the reproduced matrices will be exact. To appreciate the derivations under (ii) it will be helpful to note in advance that (2.1) implies

$$C^{-1}\hat{A} = \hat{\Sigma}^{-1}\hat{A}, \tag{2.14}$$

which leads to considerable simplifications.

We first check the reproduced score matrix $\hat{Y}$, given $\hat{A}$, $\hat{U}^2$, X and Z:

$$\begin{aligned} \hat{Y} &= \hat{A}X + \hat{U}Z \\ &= \hat{A}\hat{A}'\hat{\Sigma}^{-1}Y + \hat{A}PS + \hat{U}^2\hat{\Sigma}^{-1}Y - \hat{A}PS \\ &= Y \end{aligned} \tag{2.15}$$

under (i), (ii), and (iii). *I.e.*, $\hat{A}$, $\hat{U}$, and X, Z will always reproduce the observed scores exactly. We now check the properties of X:

$$\begin{aligned} XX'/N &= I + \hat{A}'\hat{\Sigma}^{-1}(C - \hat{\Sigma})\hat{\Sigma}^{-1}\hat{A} \neq I_m && \text{under (i)} \\ &= I_m && \text{under (ii) and (iii).} \end{aligned} \tag{2.16}$$

$$\begin{aligned} XY'/N &= \hat{A}'\hat{\Sigma}^{-1}C \neq \hat{A}' && \text{under (i)} \\ &= \hat{A}' && \text{under (ii) and (iii)} \end{aligned} \tag{2.17}$$

$$\begin{aligned} XZ'/N &= \hat{A}'\hat{\Sigma}^{-1}(C - \hat{\Sigma})\hat{\Sigma}^{-1}\hat{U} \neq \phi && \text{under (i)} \\ &= \phi && \text{under (ii) and (iii).} \end{aligned} \tag{2.18}$$

Thus, the score matrix X has all the properties of a matrix of "(uncorrelated) common factor scores," once $\hat{A}$, $\hat{U}^2$ are MLE's, whether C fits the model exactly or not. On the other hand, we find for

$$\begin{aligned} ZZ'/N &= I_p + \hat{U}\hat{\Sigma}^{-1}(C - \hat{\Sigma})\hat{\Sigma}^{-1}\hat{U} \neq I_p && \text{under (i) and (ii)} \\ &= I_p && \text{under (iii),} \end{aligned} \tag{2.19}$$

and

$$\begin{aligned} (2.20) \qquad ZY'/N &= \hat{U}\hat{\Sigma}^{-1}C \neq \hat{U} && \text{under (i) and (ii)} \\ &= \hat{U} && \text{under (iii)}, \end{aligned}$$

i.e., the score matrix Z has the properties of a matrix of "(standardized) unique scores" only if the model fits exactly, whether $\hat{A}$, $\hat{U}^2$ are MLE's or not. These results further strengthen the already strong case [Browne, 1969] for MLFA as a method for obtaining $\hat{A}$, $\hat{U}^2$. They also provide a strong case for calling the elements x_{ij} in X "(common) factor scores": these numbers behave precisely the way "(common) factor scores" should behave, according to the model, and they do so regardless whether the model fits in the sample exactly or not, as long as MLFA was employed. The only thing "wrong" is the fact that these scores are not unique, which, perhaps, should not be faulted to the x_{ij} but rather to the model. If one dislikes score indeterminacies, one should probably look for other models which do not contain any, *e.g.*, models which define the latent variables as linear combinations of the (possibly rescaled) observed variables ("Component analysis," which contains "Principal component analysis" as a special case). Once one embarks on the factor model, one has to live with this indeterminacy. To call another set of numbers which satisfies none of the strictures of the model (*e.g.*, the so-called "regression estimates") "factor scores" in preference to X, which satisfies them all, does not strike us as very rational.

Our present interest is in the factor model. We now inquire to which extent X is indeterminate, referring to earlier work by Kestelman [1952], Guttman [1955], and others. These results can, of course, also be interpreted along more traditional lines by considering the factor indeterminacy measures in the sample estimates (in fact, MLE's, since we now assume $\hat{A}$, $\hat{U}^2$ are MLE's, *cf.* Anderson [1958, p. 47]) of the population parameters which indicate the extent of the indeterminacy of the latent random variables in the population.

In the sample we consider the correlations between corresponding rows in X and another score matrix X^* which also has been constructed in accordance with (2.12) for the same matrix S but a different Gram factor $P^* = PG$, where G is an $m \times m$ rotation. We choose G so as to minimize the sum of all diagonal elements in $XX^{*\prime}/N$. As Guttman [1955] has shown, this happens when $G = -I$. We therefore obtain

$$(2.21) \qquad R_{xx^*} = XX^{*\prime}/N = 2\hat{A}'\hat{\Sigma}^{-1}\hat{A} - I_m$$

for the correlations between minimally correlated equivalent "factor scores" X, X^*. The diagonal elements of this matrix

$$(2.22) \qquad r_{x_jx_j^*} = 2\hat{a}_j'\hat{\Sigma}^{-1}\hat{a}_j - 1 = 2\hat{a}_j'C^{-1}\hat{a}_j - 1 = \hat{\rho}_{x_jx_j^*} \qquad j = 1, \cdots, m$$

(where $\hat{a}_j'$ denotes the j'th row of $\hat{A}'$) give the sample correlations between two equivalent columns of factor scores for the same factor x_j.

On the other hand, the elements in (2.21), (2.22) could also be interpreted as MLE's of the corresponding population parameters, *i.e.*, the population correlations between minimally correlated equivalent common factors, under the hypothesis that the factor model fits in the population. The latter interpretation also applies to

$$\hat{\rho}_{z_i z_i^*} = 2\hat{u}_i'\hat{\Sigma}^{-1}\hat{u}_i - 1 \qquad i = 1 \cdots, p \tag{2.23}$$

(where $\hat{u}_i'$ is the i'th row of $\hat{U}$). The former, score interpretation, does not apply to the unique variables because Z, as we saw, does not qualify as a matrix of "unique scores" except when the model fits exactly.

In practice, however, our main interest revolves around the common factors. We will now show that the elements in (2.22) can be written down at once as simple functions of the "retained roots" in D_{b1}^2 once $\hat{A}$, $\hat{U}^2$ have been estimated by the method of MLFA if $\hat{A}$ has been identified as in (2.5). We find

$$\hat{\rho}_{x_j x_j^*} = (c_j^2 - 1)/(c_j^2 + 1) = 1 - 2/b_j^2 \tag{2.24}$$

where c_j^2 are the diagonal elements in (2.5) and b_j^2 are the elements in (2.7): From

$$\hat{A}'C^{-1}\hat{A} = \hat{A}'\hat{\Sigma}^{-1}\hat{A} = \hat{A}'(\hat{A}\hat{A}' + \hat{U}^2)^{-1}\hat{A}$$
$$= \hat{A}'[\hat{U}^{-2} - \hat{U}^{-2}\hat{A}(\hat{A}'\hat{U}^{-2}\hat{A} + I)^{-1}\hat{A}'\hat{U}^{-2}]\hat{A} = D_c^2 - D_c^4(D_c^2 + I)^{-1}$$

it is seen that (2.1), (2.2), and (2.5) imply that $\hat{A}'C^{-1}\hat{A}$ is diagonal. The minimum correlation between equivalent common factors is given by (2.22) which thus reduces to $(c_j^2 - 1)/(c_j^2 + 1)$. From the definition of $\hat{A}$ in (2.8) one further concludes that $\hat{A}'\hat{U}^{-2}\hat{A} = D_c^2 = D_{b1}^2 - I_m$, which gives (2.24).

It thus turns out that both sets of roots in D_b^2 carry relevant information: the rejected roots reflect the fit of the model through the (likelihood ratio) test (2.9) and the retained roots reflect the degree of determinateness of the common factors, both in the sample and in the population. The identifiability constraint (2.5) has an even stronger consequence:

Theorem: If $\hat{A}$, $\hat{U}^2$ are obtained by the method of MLFA and the identifiability constraint (2.5) is employed, then the (uncorrelated) common factors associated with $\hat{A}$ are ordered from most to least determinate among all common factors with unit variance obtainable by orthogonal or oblique rotation. In particular, the least determinate factor has minimum correlation $\hat{\rho}_{x_m x_m^*} = 1 - 2/b_m^2$ and the most determinate common factor has minimum correlation $\hat{\rho}_{x_1 x_1^*} = 1 - 2/b_1^2$.

This follows from the fact that the minimum correlations between equivalent common factors relate monotonically to the eigenvalues c_j^2 of $M = A^{*\prime}\hat{U}^{-2}A^*$, where A^* is any factor pattern (structure) obtained by rotation through L, $A^* = \hat{A}L$. It is well known that the eigenvalues of M are the extreme values of all quadratic forms $q = \lambda_i' M \lambda_i$ under choice of unit length vectors λ_i, and under the side condition that $\lambda_i'\lambda_j = 0$ for $i \neq j$ (Bellman, 1960, p. 111). In particular, $c_1^2 = \max_{\lambda'\lambda=1} \lambda' M \lambda$, and $c_m^2 = \min_{\lambda'\lambda=1} \lambda' M \lambda$. If the roots c_j^2 are all distinct, all λ_j will be mutually orthogonal, so that $L = (\lambda_1 \cdots \lambda_m)$ is a rotation. If not all c_j^2 are distinct, then L can be chosen orthogonal, as is well known.

Thus, inspection of the latent roots c_j^2 of the matrix $M = A^{*\prime}\hat{U}^{-2}A^*$ if A^* does not satisfy the constraint (2.5), or of $M = D_c^2 = \hat{A}'\hat{U}^{-2}\hat{A}$ if it does shows at once the range of all minimum correlations between equivalent common factors obtainable by orthogonal or "oblique" rotation.

Harris [1962, p. 259] suggests one retain only those factors which correspond to $b_j^2 > 1$. If one were to insist on factors which are better determined than a set of standardized random numbers are (in the sense that the scores on factor x_j can be predicted better from scores on its equivalent twin x_j^* than from a set of random numbers), then one would have to raise that standard to retaining only factors for which $b_j^2 > 2$.*

One might say that the identifiability constraint (2.5) rotates the common factors into a "canonical position" relative to the measures of factor indeterminacy $\hat{\rho}_{x_j x_j^*}$ in much the same sense (and, basically for the same reasons [Anderson, 1958, p. 274]) as a set of principal components can be understood as a new set of variables obtained from a set of given variables upon rotation into a "canonical position" relative to the variances. Once the indeterminacy measures $\hat{\rho}_{x_j x_j^*}$ are known for the m factors x_j implied by (2.5), we can compute the indeterminacy measures $\hat{\rho}_{v_j v_j^*}$ for any new set of common factors $v_1, \cdots, v_m$ obtained by orthogonal or "oblique" rotation as weighted averages of the indeterminacy measures $\hat{\rho}_{x_j x_j^*}$:

Corollary: If a set of factors $v_1, \cdots, v_m$ is obtained by orthogonal or "oblique" rotation from a set of uncorrelated factors $x_1, \cdots, x_m$ whose pattern $\hat{A}$ satisfies (2.5), then the minimum correlation between equivalent common factors v_j, v_j^* is given by the weighted average

$$\hat{\rho}_{v_j v_j^*} = \sum_i^m t_{ij}^2 \hat{\rho}_{x_i x_i^*} \tag{2.25}$$

where the t_{ij} in $\tau_j' = (t_{1j}, \cdots, t_{mj})$ are the elements in the j'th column τ_j of the transformation matrix T which carries the factor structure of the x_i into the factor structure of the v_j.

This follows from the definition of the minimum correlation between equivalent oblique common factors v_j, v_j^*

* Note that a negative correlation does not, of course, mean that the factors are better determined than if it were zero but rather that the range of correlations among all equivalent factors extends below and, thus, includes zero.

$$(2.26) \qquad \hat{\rho}_{v_j v_j *} = 2\hat{\rho}^2_{v_j \cdot y_1 \cdots y_p} - 1 = 2\hat{\sigma}_j' \hat{\Sigma}^{-1} \hat{\sigma}_j - 1$$

where $\hat{\sigma}_j$ is the j'th column of the (oblique) factor structure of the v's [Guttman, 1955] and the fact that $\hat{A}'\hat{\Sigma}^{-1}\hat{A}$ is diagonal if $\hat{A}$ satisfies (2.5) as was shown earlier:

$$(2.27) \qquad \hat{\rho}_{v_j v_j *} = 2\hat{\sigma}_j' \hat{\Sigma}^{-1} \hat{\sigma}_j - 1 = 2\tau_j'(\hat{A}'\hat{\Sigma}^{-1}\hat{A})\tau_j - 1 = \sum_i^m t_{ij}^2 \hat{\rho}_{x_i x_i *} \, .$$

The practical significance of this result is that knowledge of the c_i^2 in (2.5) (or, equivalently, knowledge of D_{b1}^2 in (2.7)) together with knowledge of the transformation T relating the two factor structures suffices to compute the minimum correlations $\hat{\rho}_{v_j v_j *}$ for any other set of factors $v_1, \cdots, v_m$ obtained by orthogonal or "oblique" rotation without need to use (2.22) again. Inspection of T will also explain why oblique rotation sometimes improves the average indeterminacy of the common factors: new factors v_j which have relatively large weights for well determined x_i will be better determined than v_k, which are largely composed of poorly determined x_i. In the case of orthogonal rotation the sum of the minimum correlations of the v_j (and hence their average) is the same as the sum of the minimum correlations for the x_i (since $\sum_j \sum_i t_{ij}^2 \hat{\rho}_{x_j x_j *} = \sum_i \rho_{x_i x_i *} \sum_j t_{ij}^2 = \sum_i \hat{\rho}_{x_i x_i *}$ if $\sum_j t_{ij}^2 = 1$, as it is if T is orthogonal). In the case of "oblique rotation," on the other hand, $\sum_i t_{ij}^2 = 1$ does not imply $\sum_j t_{ij}^2 = 1$, so that now a new set of factors v_j can be constructed whose average minimum correlation exceeds that of the x_i by repeatedly weighing in the more determinate factors among the x_i.

If a method for "oblique rotation" to simple structure is employed, a rise in the minimum average correlation among the common factors can be expected to occur if all factors are at acute angles to each other (disregarding orientation), *i.e.*, if they are all highly correlated. From (2.8) it is apparent that in this case the most determinate common factor x_1 of the uncorrelated set whose pattern $\hat{A}$ satisfies (2.5) is in the vicinity of the centroid of the tests. Upon "oblique rotation" to simple structure all other common factors v_k would be inclined towards x_1 so that the weights t_{1k} (the cosines of the angle of x_1 with v_k) are likely to be larger in magnitude than the remaining weights t_{jk} for $j \neq 1$. Thus, all new factors v_k contain a relatively large share of the most determinate factor x_1 and the sum of their minimum correlations $\hat{\rho}_{v_j v_j *}$ will exceed that of the x_i or any other set of orthogonal factors.

3. *Factor Indeterminacy and "Oblique Rotation"*

We have seen that in theory, at least, "oblique rotation" can increase or decrease the minimum average correlation of the common factors. In Schönemann [1971] it was shown that the average minimum correlation between equivalent sets of uncorrelated factors is a constant independent of the data. Since "oblique rotation" only affects the minimum correlations of the common factors, it follows that this simple result cannot be generalized directly to the oblique case. But it is possible to state somewhat weaker results for the minimum average correlation of oblique factors in the form of bounds which, as will be shown, depend on the latent roots of the correlation matrix for the common factors. For notational convenience we shall treat the population case. The derivations for the exact sample case are analogous.

In the oblique case the matrix of correlations between two sets of minimally correlated factors $\begin{pmatrix}\xi\\ \zeta\end{pmatrix}$ and $\begin{pmatrix}\xi^*\\ \zeta^*\end{pmatrix}$ is given by

$$C_{\min} = 2\begin{pmatrix}\psi A'\\ U\end{pmatrix}\Sigma^{-1}(A\psi, U) - \begin{pmatrix}\psi & \\ & I_p\end{pmatrix} \tag{3.1}$$

where $\eta = A\xi + U\zeta$; var $(\eta) = \Sigma$, var $(\xi) = \psi$, var $(\zeta) = I_p$, cov $(\xi, \zeta) = \phi$. This matrix has been discussed in some detail by Guttman [1955], while Heermann [1966] considered the special case for $\psi = I_m$.

Note that in the oblique case the diagonal elements are still the squared multiple correlations $\rho^2_{x_j \cdot y_1 \ldots y_p}$ and $\rho^2_{z_i \cdot y_1 \ldots y_p}$ as they were in the orthogonal case. But, in contrast to the orthogonal case, this covariance matrix no longer coincides with the transformation matrix $T_{\min}$ which carries one set of factors $\begin{pmatrix}\xi\\ \zeta\end{pmatrix}$ into another, equivalent set $\begin{pmatrix}\xi^*\\ \zeta^*\end{pmatrix} = T'\begin{pmatrix}\xi\\ \zeta\end{pmatrix}$ which is pairwise minimally correlated with the first. Rather, if the common factors x_i are correlated, one has

$$C_{\min} = \text{cov}\left\{\begin{pmatrix}\xi\\ \zeta\end{pmatrix}, \begin{pmatrix}\xi^*\\ \zeta^*\end{pmatrix}\right\} = \text{cov}\left\{\begin{pmatrix}\xi\\ \zeta\end{pmatrix}, T'\begin{pmatrix}\xi\\ \zeta\end{pmatrix}\right\} = \left[\text{var}\begin{pmatrix}\xi\\ \zeta\end{pmatrix}\right]T \tag{3.2}$$

i.e.,

$$T' = C_{\min}\begin{pmatrix}\psi & \\ & I_p\end{pmatrix}^{-1} = 2\begin{pmatrix}\psi A'\\ U\end{pmatrix}\Sigma^{-1}(A, U) - I_{p+m} \neq T. \tag{3.3}$$

This transformation matrix is no longer symmetric, as it was in the orthogonal case. And while the matrix

$$N = \begin{pmatrix}\psi A'\\ U\end{pmatrix}\Sigma^{-1}(A, U) = N^2 \neq N' \tag{3.4}$$

is idempotent, so that its trace equals its rank (which is p), $2N - I = T'$ is no longer orthogonal.

For this reason the fairly straightforward argument which Schönemann [1971] used to evaluate the minimum average correlation between equivalent sets of uncorrelated factors can no longer be employed in the correlated case. However, a somewhat different argument can be used to obtain bounds for the minimum correlation between equivalent sets of correlated factors. Let

$$WW' = \Sigma, \qquad W \text{ nonsingular} \tag{3.5}$$

be any complete factorization of Σ and $\psi^{1/2}$ a symmetric factor of ψ, so that

$$\psi^{1/2} = \psi^{1/2\prime} \qquad \psi^{1/2}\psi^{1/2\prime} = (\psi^{1/2})^2 = \psi. \tag{3.6}$$

Then from

$$\Sigma = WW' = (A\psi^{1/2}, U)\begin{pmatrix} \psi^{1/2}A' \\ U \end{pmatrix} \tag{3.7}$$

one knows that a $(p + m) \times (p + m)$ orthogonal matrix Q exists so that

$$(W, \phi)Q = (A\psi^{1/2}, U) \tag{3.8}$$

since the rows of both (W, ϕ) and $(A\psi^{1/2}, U)$ give rise to the same scalar products in Σ. Therefore,

$$\begin{aligned} C_{\min} &= 2\begin{pmatrix} \psi A' \\ U \end{pmatrix}\Sigma^{-1}(A\psi, U) - \begin{pmatrix} \psi & \\ & I_p \end{pmatrix} \\ &= 2\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix}\left\{\begin{pmatrix} \psi^{1/2}A' \\ U \end{pmatrix}\Sigma^{-1}(A\psi^{1/2}, U)\right\}\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix} - \begin{pmatrix} \psi & \\ & I_p \end{pmatrix} \\ &= 2\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix}\left\{Q'\begin{pmatrix} W' \\ \phi \end{pmatrix}(WW')^{-1}(W, \phi)Q\right\}\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix} - \begin{pmatrix} \psi & \\ & I_p \end{pmatrix} \\ &= 2\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix}\left\{Q'\begin{pmatrix} I_p & \\ & \phi \end{pmatrix}Q\right\}\begin{pmatrix} \psi^{1/2} & \\ & I_p \end{pmatrix} - \begin{pmatrix} \psi & \\ & I_p \end{pmatrix} \end{aligned} \tag{3.9}$$

and

$$\operatorname{tr} C_{\min} = 2 \operatorname{tr} Q\begin{pmatrix} \psi & \\ & I_p \end{pmatrix}Q'\begin{pmatrix} I_p & \\ & \phi \end{pmatrix} - (p + m). \tag{3.10}$$

The diagonal matrix $\begin{pmatrix} I_p & \\ & \phi \end{pmatrix}$ simply suppresses the last m diagonal elements in

$$B = Q\begin{pmatrix} \psi & \\ & I_p \end{pmatrix}Q' \tag{3.11}$$

so that

$$\operatorname{tr} Q\begin{pmatrix} \psi & \\ & I_p \end{pmatrix}Q'\begin{pmatrix} I_p & \\ & \phi \end{pmatrix} = \sum_{i=1}^{p} b_{ii}$$

can be called a "partial trace."

That such partial traces are maximized under choice of orthogonal matrices Q if the first p rows of Q are the eigenvectors corresponding to the p largest roots of B, and are minimized if the first p rows of Q are the eigenvectors corresponding to the p smallest roots of B again follows from the well-known variational properties of eigendecompositions [*e.g.*, Bellman, 1960, p. 111, or Householder, 1963, p. 76f]. Thus, if

$$\psi = KD_d^2K' \qquad KK' = K'K = I_m\,,\; D_d^2 = \text{diagonal} \tag{3.12}$$

has eigenvalues

$$d_1^2 \geq d_2^2 \geq \cdots \geq d_m^2 \tag{3.13}$$

so that $\begin{pmatrix} \psi & \\ & I_p \end{pmatrix}$ has eigenvalues

$$\underbrace{d_1^2 \geq d_2^2 \geq \cdots \geq d_k^2}_{\substack{k \text{ largest roots of } \psi \\ \text{which exceed unity}}} > 1, \cdots, 1, \geq \underbrace{d_{k+1}^2 \cdots d_m^2}_{\substack{m - k \text{ smallest roots of } \psi \\ \text{which are less or equal unity}}} \tag{3.14}$$

then

$$p - k + \sum_i^k d_i^2 \qquad (d_i^2 > 1) \tag{3.15}$$

will be the maximum of the partial trace of (3.11) and

$$p - (m - k) + \sum_{j=k+1}^{m} d_j^2 = p + k - \sum_{i=1}^{k} d_i^2\,, \qquad (d_j^2 < 1)(d_i^2 > 1) \tag{3.16}$$

will be its minimum. If we now define

$$\delta^2/2 = \sum_i^k (d_i^2 - 1),\; (d_i^2 > 1),\; \tau^* = \frac{1}{p + m}\,\mathrm{tr}\, C_{\min} \tag{3.17}$$

and

$$\tau = (p - m)/(p + m),$$

then we obtain

$$\tau - \delta^2/(p + m) \leq \tau^* \leq \tau + \delta^2/(p + m). \tag{3.18}$$

These bounds are weak (*i.e.*, the interval $2\,\delta^2/(p + m)$ is wider than necessary) because we assumed for their derivation that we are free to choose among all orthogonal matrices Q of order $(p + m) \times (p + m)$. In fact, however, Q is a function of A and ψ (and thus U and Σ):

$$Q = \begin{pmatrix} I_p & \\ & S \end{pmatrix}\begin{pmatrix} W^{-1} & \\ & P' \end{pmatrix}\begin{pmatrix} A\psi^{1/2} & U \\ \psi^{-1/2} & -A'U^{-1} \end{pmatrix} \tag{3.19}$$

where P is defined by

$$PP' = PSS'P' = \psi - \psi A'\Sigma^{-1}A\psi \tag{3.20}$$

and S is an arbitrary orthogonal matrix of order $m \times m$. (3.19), (3.20) are straightforward generalizations of Heermann's [1966] explicit representation of Q to the oblique case. It is easily verified that $(W, \phi)Q = (A\psi^{1/2}, U)$. To see that Q is also orthogonal it is convenient to use $(PP')^{-1} = \psi^{-1} - A'U^{-2}A$ [Guttman, 1955].

Since "oblique rotation" does not affect the unique factors at all, (3.18) can be transformed into a set of bounds for the minimum average correlation between equivalent correlated common factors, say τ_x^*, in terms of the minimum average correlation between equivalent uncorrelated common factors, say $\tau_x (= 1 - 2 \sum_i^m b_i^{-2}/m$, from (2.24), where the sum extends over the retained latent roots of $U^{-1}CU^{-1}$ in (2.7)):

$$\tau_x - \delta^2/m \leq \tau_x^* \leq \tau_x + \delta^2/m. \tag{3.21}$$

But since these bounds differ from those in (3.18) only in the smaller denominator m (in place of $p + m$), they are much wider and, therefore, as our experience showed, in practice rather useless. For that reason we shall not discuss them any further.

4. Some Empirical Illustrations

Although the basic facts about factor indeterminacy are as old as factor analysis itself [Wilson, 1928, Camp, 1932, Piaggio 1933] one rarely finds an empirical study which actually reports the indeterminacy measures for each common factor.† In an attempt to gain some perspective about the practical relevance of this issue we thought it worthwhile to reanalyze a number of studies with the method of maximum likelihood and compute some of the indeterminacy measures both before and after rotation.

We reanalyzed 13 different studies, the first 11 of which are the same and in the same order as in a paper by Jöreskog [1967] who used them to illustrate the efficiency of his MLFA algorithm. We are indebted to Dr. M. Browne for a very efficient MLFA algorithm which utilizes the Gauss–Seidel method [Browne, 1968]. This particular algorithm does not employ the identifiability constraint (2.5) which therefore had to be imposed upon convergence. At this stage we also computed all latent roots b_i^2, rather than just the "retained roots" which could have been obtained directly from (2.5). Our results are presented in Tables 1, 2, and 3.

For each data set we computed the "exact" likelihood ratio test $T(X)$ in (2.9) as well as an "approximate" test $T'(X)$ which is given by

$$T'(x) = k \sum_{i<j} \sum e_{ij}^2 = -k \sum_{i=m+1}^{p} \{(b_i^2 - 1) - (b_i^2 - 1)^2/2\} \tag{4.1}$$

(with k, e_{ij}^2, b_i^2 defined as in (2.9)).

† A notable exception, according to one of our reviewers, is a recent Ph.D. thesis by E. P. Meyer, "Some Results Concerning Choice of Uniqueness Estimates, Number of Factors, and Determinacy of Factor Score Matrices," University of Wisconsin, 1969, which contains "extensive empirical results."

TABLE 1

Summary of Empirical Results

N	p	m	var. omit-ted	df	T(X)	π	T'(X)	π	T''(X)	π	τ	τ_x	τ^*	τ^*_x	τ_L	τ_U	min ρ_{xx^*}
Data 1: Emmett (1949)																	
211	9	2		19	27.53	.09	29.00	.07	27.53	.09	.636	.643	.655	.795	(.531	.741)	.516
	9	3		12	7.15	.85	7.23	.84	7.15	.85	.500	.445	.539	.599	(.360	.640)	-.082
	9	4		6	2.81	.83	2.77	.84	2.81	.83	.385	.226	.442	.413	(.215	.555)	-.466
Data 2: Maxwell (1961)																	
810	10	3		18	78.55	.00	76.77	.00	78.55	.00	.539	.319	.592	.549	(.398	.679)	-.193
	9	3	8	12	14.48	.27	14.00	.30	18.31	.11	.500	.361	.554	.577	(.343	.657)	-.084
	8	3	6,8	7	12.43	.09	12.05	.10	13.29	.07	.455	.356	.513	.572	(.286	.623)	-.098
Data 3: Bechtoldt, S1 (1961)																	
212	17	5		61	104.04	.00	96.32	.00	104.09	.00	.545	.617	.581	.772	(.384	.707)	.300
	16	5	2	50	49.44	.50	49.36	.50	54.35	.31	.524	.622	.560	.774	(.356	.692)	.306
	15	5	2,3	40	29.02	.90	29.93	.88	35.98	.65	.500	.603	.540	.761	(.322	.678)	.304
Data 4: Bechtoldt, S2 (1961)																	
213	17	5		61	100.66	.00	99.83	.00	100.41	.00	.545	.599	.576	.754	(.387	.704)	.280
	16	5	1	50	53.89	.33	51.58	.41	52.34	.38	.524	.579	.562	.739	(.360	.688)	.211
	*17	7		38	33.44	.68	32.05	.74	33.55	.68	.417	.475	.471	.661	(.226	.607)	-.170

TABLE 1 (continued)

N	p	m	var. omit-ted	df	T(X)	π	T'(X)	π	T''(X)	π	τ	τ_x	τ^*	τ^*_x	τ_L	τ_U	min ρ_{xx^*}
Data 5: Lord (1956)																	
649	33	9		267	345.80	.00	346.63	.00	345.70	.00	.571	.404	.607	.569	(.450	.693)	-.234
	32	10	27	221	229.85	.33	224.38	.42	**		.524	.347	.563	.511	(.376	.671)	-.296
	33	11		220	225.91	.38	221.22	.46	225.91	.38	.500	.369	.547	.557	(.331	.670)	-.288
Data 6: Davis (1944), Rao (1955), m = 1 Omitted.																	
Data 7: Harman (1960, p. 82)																	
305	8	2		13	75.74	.00	75.44	.00	75.74	.00	.600	.865	.606	.893	(.508	.692)	.801
	7	2	2	8	20.95	.01	21.27	.01	22.67	.00	.556	.830	.564	.866	(.453	.658)	.749
	7	3	2	3	1.53	.68	1.52	.68	4.33	.33	.400	.551	.424	.632	(.281	.520)	-.011
Data 8: Harman (1960, p. 136)																	
145	12	3	11	33	46.54	.06	43.72	.10	47.77	.05	.600	.596	.618	.685	(.499	.702)	.352
	12	4	5	24	25.47	.38	24.21	.45	28.64	.33	.500	.583	.530	.703	(.367	.633)	.303
	10	3	3,5,11	18	10.13	.93	10.40	.92	14.33	.71	.539	.577	.562	.677	(.410	.667)	.331
Data 9: Hemmerle (1965). Omitted; no sample size.																	
Data 10: Browne, S1 (1965)																	
100	11	2	1	34	50.38	.04	46.38	.08	55.17	.01	.692	.619	.692	.617	(.689	.696)	.480
	12	4			Omitted; boundary case on var. 1.												
	12	5		16	7.29	.97	7.06	.97	9.32	.90	.412	.715	.421	.745	(.323	.500)	.371

TABLE 1 (continued)

N	p	m	var. omit-ted	df	T(X)	π	T'(X)	π	T''(X)	π	τ	τ_x	τ^*	τ_x^*	τ_L	τ_U	min ρ_{xx^*}
Data 11: Browne, S19 (1965)																	
100	12	3		33	31.43	.55	29.19	.66	31.53	.54	.600	.776	.600	.777	(.546	.654)	.644
	11	3	12	25	21.60	.66	20.75	.71	22.99	.53	.571	.783	.572	.785	(.520	.623)	.652
	10	3	2,4	18	19.47	.36	18.86	.40	11.97	.85	.539	.425	.548	.466	(.485	.592)	.138
Data 12: Harman (1967, p. 166)																	
147	7	2	8	8	84.87	.00	101.80	.00	--	--	.556	.927	.556	.930	(.472	.639)	.865
Data 13: Thurstone (1951)																	
213	13	4		32	107.51	.00	111.29	.00	--	--	.529	.648	.545	.714	(.434	.625)	.277
Data 14: Harman (1967, p. 125)																	
145	24	4		186	226.68	.02	213.89	.08	--	--	.714	.568	.736	.719	(.615	.814)	.228
	24	5		166	186.82	.13	172.16	.36	--	--	.655	.510	.676	.632	(.556	.755)	.134
	23	6	3	130	131.59	.45	123.24	.65	--	--	.586	.408	.611	.530	(.470	.702)	-.050

* Failed to converge after 100 iterations.

** Not included in Jöreskog (1967)

TABLE 2

Minimum Correlations Between Equivalent Common Factors

Legend: CP - "Canonical position," $\hat{A}$ identified by $\hat{A}'\hat{U}^{-2}\hat{A}$ = diagonal

VX - Varimax position

PX - (Oblique) Promax position

Study	p	m	π	Position	Common Factors 1	2	3	4	5	6	7	8	9	10	11
Emmett	9	2	.09	CP	.87	.52									
(1949)				VX	.71	.68									
				PX	.80	.79									
	9	3	.35	CP	.88	.54	-.08								
				VX	.72	.53	.09								
				PX	.81	.76	.23								
	9	4	.83	CP	.88	.57	-.09	-.47							
				VX	.73	.51	.11	-.45							
				PX	.82	.74	.28	-.19							

TABLE 2 (continued)

Study	p	m	π	Position	1	2	3	4	5	6	7	8	9	10	11
Maxwell	10	3	.00	CP	.77	.38	-.19								
(1961)				VX	.28	.50	.17								
				PX	.51	.60	.54								
	9	3	.27	CP	.77	.39	-.08								
				VX	.31	.51	.26								
				PX	.52	.61	.60								
	8	3	.09	CP	.77	.40	-.10								
				VX	.31	.51	.25								
				PX	.53	.60	.58								
Bechtoldt	17	5	.00	CP	.92	.76	.70	.42	.30						
S1 (1961)				VX	.75	.75	.66	.51	.42						
				PX	.87	.77	.79	.68	.75						
	16	5	.50	CP	.92	.76	.72	.42	.31						
				VX	.76	.75	.69	.49	.42						
				PX	.87	.77	.81	.67	.75						
	15	5	.90	CP	.90	.74	.66	.41	.30						
				VX	.64	.75	.66	.50	.47						
				PX	.81	.77	.79	.67	.76						

TABLE 2 (continued)

Study	p	m	π	Position	1	2	3	4	5	6	7	8	9	10	11
Bechtoldt	17	5	.00	CP	.91	.73	.61	.47	.28						
S2 (1961)				VX	.71	.69	.67	.53	.40						
				PX	.83	.73	.76	.74	.77						
	16	5	.41	CP	.91	.72	.59	.47	.21						
				VX	.73	.68	.63	.54	.31						
				PX	.86	.73	.73	.72	.66						
	17	7	.74	CP	.91	.73	.60	.49	.42	.35	-.17				
				VX	.74	.69	.61	.53	.47	.24	.05				
				PX	.84	.73	.73	.73	.56	.54	.50				
Lord	33	9	.00	CP	.95	.91	.82	.66	.60	.13	-.04	-.17	-.24		
(1956)				VX	.90	.90	.69	.72	.56	.03	-.05	.02	-.14		
				PX	.93	.92	.79	.84	.54	.28	.24	.32	.26		
	32	10	.42	CP	.95	.91	.85	.68	.56	.14	.06	-.15	-.23	-.30	
				VX	.90	.90	.77	.72	.52	.01	-.05	-.04	-.18	-.08	
				PX	.91	.92	.85	.84	.53	.28	.11	.31	-.02	.37	
	33	11	.46	CP	.96	.91	.88	.76	.68	.33	.14	.08	-.17	-.23	-.29
				VX	.90	.90	.80	.75	.72	.10	.18	-.06	-.04	-.16	-.03
				PX	.91	.92	.88	.72	.84	.32	.50	.25	.24	.03	.50

TABLE 2 (continued)

Study	p	m	π	Position	1	2	3	4	5	6	7	8	9	10	11
Harman	8	2	.00	CP	.93	.80									
(1960,				VX	.89	.84									
p. 82)				PX	.91	.87									
	7	2	.01	CP	.91	.75									
				VX	.84	.82									
				PX	.88	.86									
	7	3	.68	CP	.91	.75	-.01								
				VX	.85	.76	.04								
				PX	.89	.81	.20								
Harman	12	3	.06	CP	.85	.59	.35								
(1960,				VX	.74	.61	.44								
p. 136)				PX	.82	.65	.59								
	12	4	.38	CP	.88	.74	.41	.30							
				VX	.65	.72	.53	.43							
				PX	.78	.80	.65	.59							
	10	3	.93	CP	.83	.57	.33								
				VX	.72	.55	.47								
				PX	.80	.62	.62								

TABLE 2 (continued)

Study	p	m	π	Position	1	2	3	4	5	6	7	8	9	10	11
Browne	11	2	.08	CP	.76	.48									
S1 (1965)				VX	.74	.50									
				PX	.74	.50									
	12	5	.97	CP	.95	.89	.75	.62	.37						
				VX	.88	.88	.69	.70	.44						
				PX	.92	.91	.70	.75	.45						
Browne	12	3	.55	CP	.91	.78	.64								
S19 (1965)				VX	.83	.74	.71								
				PX	.83	.77	.73								
	11	3	.66	CP	.92	.78	.65								
				VX	.87	.77	.71								
				PX	.86	.77	.73								
	10	3	.36	CP	.79	.35	.14								
				VX	.70	.34	.23								
				PX	.71	.36	.33								
Harman	7	2	.00	CP	.99	.87									
(1967,				VX	.99	.87									
p. 166)				PX	.99	.87									

TABLE 2 (continued)

Study	p	m	π	Position	1	2	3	4	5	6	7	8	9	10	11
Thurstone	13	4	.00	CP	.96	.81	.55	.28							
(1951)				VX	.55	.81	.56	.68							
				PX	.52	.82	.66	.86							
Harman	24	4	.02	CP	.89	.66	.49	.23							
(1967,				VX	.74	.66	.52	.36							
p. 125)				PX	.83	.74	.71	.59							
	24	5	.13	CP	.90	.69	.51	.31	.13						
				VX	.74	.66	.54	.39	.23						
				PX	.83	.74	.72	.60	.27						
	23	6	.45	CP	.91	.69	.48	.30	.13	-.05					
				VX	.78	.60	.50	.40	.19	-.02					
				PX	.86	.71	.72	.60	.23	.07					

TABLE 3

Doublets or Specifics in Reference Vector Structures Upon Oblique Rotation

	1	2	3	4	5	6	7	8	9	10	11	12	13
Harman (1960, p. 136)													
p = 13, m = 4	.06	.02	-.04	-.11	.02	.11	-.10	-.04	.03	.02	.79	-.03	.15
var. 11 partialled out													
p = 13, m = 5	.07	.03	-.06	-.11	-.01	.16	-.10	-.07	.04	.06	.68	-.02	.18
var. 5 partialled out	-.01	.16	.04	-.14	.63	-.05	.00	.07	.16	.05	-.01	.00	-.03
p = 13, m = 6	-.05	.11	.83	-.01	-.00	.02	-.02	.17	-.09	-.10	.00	.02	.07
var. 3, 5, 11 partialled out	.03	.17	-.01	-.13	.63	-.04	-.01	.05	.17	.05	-.02	.00	-.03
p = 8, m = 3	-.05	.36	.16	-.07	-.09	-.01	.01	.17					
var. 2 partialled out													
Browne (1965), S 19													
p = 12, m = 4	.00	.01	-.01	-.05	-.05	.04	-.08	-.02	.21	.04	-.12	.94	
var. 12 partialled out													
p = 12, m = 5	.00	.02	-.02	-.07	-.02	.03	-.09	-.04	.20	.03	-.10	.94	
var. 2, 5, 12 partialled out													
Maxwell (1961)													
p = 10, m = 5	.07	.02	.03	-.06	-.03	.01	-.01	.83	.17	-.03			
var. 6, 8 partialled out	-.03	.00	-.00	.00	.02	.89	.00	.01	-.04	.01			

TABLE 3 (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Thurstone (1951)																	
p = 13, m = 6, var. 1,4,6 partialled out	-.07	-.01	.01	.05	.09	.71	.05	-.06	.04	-.11	.03	-.02	.02				
Harman (1960, p. 82)																	
p = 8, m = 4	-.06	.02	-.04	.07	.09	-.09	.51	.08									
var. 2 partialled out	-.05	.35	.14	-.13	-.09	-.01	.01	.15									
Bechtoldt (1961), S1																	
p = 17, m = 6	.36	.95	-.04	-.02	.01	-.02	.03	.00	-.06	.00	.09	.01	.02	-.11	.03	.01	-.01
p = 17, m = 7	.80	.42	-.05	.08	-.01	-.08	-.01	.05	-.04	.02	.01	.01	.04	-.02	.02	.02	.01
Bechtoldt (1961), S2*																	
p = 17, m = 6	.86	.36	-.04	-.00	-.05	-.02	.01	.05	.03	-.03	-.02	-.06	.02	.03	-.03	.16	-.05
p = 17, m = 7	.35	.75	.00	.02	-.01	.01	-.09	.08	.05	-.05	.06	.06	-.02	-.08	.02	.00	-.01

* For these data Jöreskog reports a boundary case for var. 1. Our reanalysis did not produce a boundary case.

This test can be obtained from (2.9) by replacing each ln b_i^2 by the first two terms of its Taylor expansion. Since $T'(x)$ is manifestly much easier to compute than $T(x)$ we were interested to see how both tests compare in practice. In Table 1 the values for both tests together with the degrees of freedom (df) and the associated (upper tail) probabilities (π) are given. It appears that the "approximate test" performs in nearly all cases as well as its supposedly more accurate rival. It is clear, of course, that strictly speaking both tests are "approximate" so far as the (chi-square) sampling distribution is concerned. $T(x)$ is the more appropriate test in the context of MLFA because the likelihood ratio involves determinants. In passing we note that $T'(x)$, on the other hand, would be the more appropriate test if one were to obtain a weighted least squares solution. Such an alternative was indeed proposed and programmed by M. Browne [1969]. It could be interpreted statistically as being based on the minimum chi square principle, instead of maximum likelihood.

Most data sets were analyzed more than once, partly to replicate the analyses in Jöreskog [1967] and partly because in most instances the more conservative m (number of common factors) was not supported statistically. For reasons to be given shortly we did not use Jöreskog's [1967] device to partial out variables when their unique variance converged on the boundary ($u^2 = 0$). Instead we eliminated such variables from the battery entirely before reanalyzing the remaining variables. In those cases, which are identified in Table 1, the values for the chi-square test and the associated degrees of freedom are different from Jöreskog's ($T''(x)$). Upon assessing the factorial indeterminacy, factor by factor, for the "canonical position" (Table 2), *i.e.* for $\hat{A}$ identified by (2.5) using (2.24), we proceeded to rotate $\hat{A}$ to simple structure using varimax (Table 2). As was pointed out in Schönemann [1971] such orthogonal rotation does not affect the average indeterminacy of the common factors (τ_x), it simply redistributes the total in accordance with (2.25). We then employed a rather crude imitation of the "Promax" algorithm proposed by Hendriksen and White [1964] which performed remarkably well as judged by intuitive standards. This we did in order to assess the change in factor indeterminacy upon "oblique rotation" (Table 2). The cols. labelled τ and τ_x in Table 1 give the minimum average correlation for equivalent factors for all factors (τ) and for the uncorrelated common factors (τ_x), respectively. The cols. τ^* and τ_x^* give the corresponding measures for the oblique solutions. One notes a slight increase in the averages in almost all cases. We tried to explain this fact on a somewhat intuitive basis in Sec. 3. Finally, in cols. τ_L , τ_U we present the bounds for the average minimum correlation in the oblique case which we computed according to (3.18). It is evident that most of the intervals are fairly wide so that this particular theoretical result is not likely to be very useful in practice. In the last line in Table 1 (min $\rho_{x_i x_i^*}$) we finally present the minimum correlation of the least determined common factor (eq. (2.24)). One finds that this correlation is often negative which means concretely that a given column of "factor scores" can be predicted with more success from a set of random numbers than from an equivalent column of "factor scores".

Our general conclusion is that the factor model has serious limitations. If the sample size is large, which, in principle, it always should be, one finds that the model does not fit very well for only a small number of factors. If one then raises m in order to achieve a better fit, one is left with more and more poorly determined common factors. This dilemma is very acute in the Maxwell [1961] study and the Lord [1956] study which were the only two studies with respectable sample sizes. The same effect emerges, to a lesser extent, in the Emmett [1949] study, the Bechtoldt S2-study [1961], the Harman study with 8 physical variables [1960], and his classic 24 variable study. The only two instances where the minimum correlations were acceptable when the model fits are the studies by Bechtoldt [S1, 1961] and M. Browne [1968]. The latter uses artificial data.

An incidental finding of these reanalyses was the discovery that oblique rotation often produced doublets in the factor patterns, once m was raised to improve the fit. Such doublets, as is well known [Anderson and Rubin, 1956], correspond to unidentifiable factor patterns, in the sense that the communalities between the two variables which load nonzero on the doublet are arbitrary within certain limits. This is clearly an undesirable situation and it appears to arise with greater frequency than might have been suspected once m is raised so as to satisfy statistical standards. Some instances of such doublets (or sometimes, "specifics") are reproduced in Table 3.

The frequency of their occurrence rose markedly once we reanalyzed some of the studies containing "boundary cases" (*i.e.*, unique variances converging on zero) with the device described in Jöreskog [1967]. To obtain a solution for such boundary cases, he partials out a variable with zero uniqueness after having rotated one of the "common factors" (actually, now, components) into collinearity with that variable. We are skeptical about the utility of this procedure for two reasons: (i) if A is unidentifiable, as we often found it to be (Table 3), then the selection of the variable to be partialled out is essentially arbitrary, because its communality can be traded off against that of the other variable on the doublet, and (ii) it strikes us as inconsistent. If one entertains the factor model, which is erected on the theory that all error is uncorrelated, then one is driven to conclude that any variable with zero unique variance is perfectly reliable. In psychology such variables are very rare indeed. To switch boats in midstream, so to speak, simply because one encounters computational difficulties, has the effect of mixing two logically incompatible models, one with correlated and one with uncorrelated error. If one is willing to accept the notion of correlated error, one is probably better off with component analysis from the start. Not only is such an approach technically simpler and computationally more expedient, it also dispenses with all the mathematical and semantical problems which accompany the built-in indeterminacies of the factor model. Component models which may serve as suitable alternatives are discussed in Guttman [1953] and Harris [1962].

5. Discussion

The issue of factor indeterminacy [Wilson, 1928; Guttman, 1955] is not only a mathematical but also a semantical problem. Consider, for example, the widely held belief that "factor scores cannot be computed directly, they can only be estimated" (verbatim in Pawlik, [1968, p. 163]. Remarks to the same effect in Horst, [1969, p. 7–8]; Kaiser, [1963]; MacDonald and Burr, [1967, p. 382]; and elsewhere). What do such statements mean?

They evidently mean hardly anything as long as we are not told in clear and unambiguous terms what is meant by "factor scores" (as distinct from "factor score estimates"). Upon checking, one finds that the exact meaning of this term is a closely guarded secret. There are good reasons for not defining it: if by "factor scores" one means, as one sometimes does by implication, observations on random variables, then "factor scores" cannot be *defined* uniquely for the simple reason that the underlying random variables, the "factors," cannot be defined uniquely. This is quite different from saying that they cannot be "calculated uniquely" [Horst, 1969, p. 7–8], which is a minor matter, by comparison.

This, in turn, raises the question what, exactly, it is that is being estimated when "factor scores" are "estimated by the regression method" [*e.g.*, Lawley and Maxwell, 1963, p. 89]. Among the more recent authors only L. Guttman appears to have appreciated the full implications of the factor indeterminacy issue for the problems of "estimating factor scores": "··· it raises the question what is being estimated in the first place; instead of only one primary trait there are many widely different variables associated with a given profile of loadings [Guttman, 1955]." In the thirties several other authors thought about this question, among them Thomson [1935], who devoted a paper to "a comparison between Spearman's g technique and the ordinary method of the regression equation." He concluded that "Spearman's case is exactly the same as (the regression case) except for the important fact that he has no 'criterion,' no measure of g except through the team of tests themselves (p. 94)." "This distinction between the two cases may appear to be subtle, but it seems a proper distinction to draw (p. 97)." From his point of view "factor scores" and "factor score estimates" are one and the same thing. If one adopts it, there is no need to speak of "least squares estimates." It is well known, however, that "factor scores," so defined, would satisfy none of the strictures of the factor model. For example, they would be correlated for factors which, according to the model, should be uncorrelated.

If, on the other hand, by "factor scores" one means a set of numbers which have all the properties the factor model prescribes once the observed covariance matrix permits an exact resolution into $C = \hat{A}\hat{A}' + \hat{U}^2$, then one will find, upon reading Kestelman [1952] or Guttman [1955], that such numbers always exist. Indeed, infinitely many different sets of such numbers can be computed, which therefore need not be estimated, by the "regression method" or any other method. Other definitions of "factor scores" may be possible. But as long as they have not been made explicit the sentence "factor scores cannot be computed, but only estimated" remains perfectly vacuous.

One of our reviewers pointed out quite correctly that whatever is obscure about the concept of factor scores should be equally osbcure about the concept of true scores in classical test theory. This follows simply from the fact that classical true score theory is a special case of two-factor theory. This is no coincidence since both go back to the same man, C. Spearman.

The point can be made that none of these problems bear on the practical utility of factor analysis as a research tool as long as it is used to study the structure of variables without attempting to estimate a person's factor scores. The factor model could be stated and studied solely at the covariance level as a model which stipulates a particular form for the observed covariance matrix. This is a perfectly legitimate interpretation of the factor model, and many statisticians seem to view it this way. But to be consistent with this interpretation one would, of course, have to refrain from recommending how to "estimate factor scores." It is a fact that most texts on factor analysis do include discussions and recommendations on this subject, as do, surprisingly, some papers and books by statisticians. We feel it betrays an inconsistency to treat the factor model in this ambiguous way. Once the factor variables are introduced as part of the model, the issue of factor indeterminacy has to be faced and resolved in some manner, however arbitrarily, and the exact meaning of the term "factor scores," as distinct from "factor score estimates," has to be specified. Thought-provoking, and sometimes amusing, discussions of these matters can be found in some of the older literature, which may well merit re-reading by today's students of factor analysis.

REFERENCES

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Anderson, T. W., & Rubin, H. Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of third Berkeley symposium on mathematical statistics and probability,* Berkeley: Univer. California Press, 1956, **5**, 111–150.

Bargmann, R. A study of independence and dependence in multivariate normal analysis. University of North Carolina, Institute of Statistics Mimeo Series No. 186, 1957.

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

Bellman, R. *Introduction to matrix analysis.* New York: McGraw-Hill, 1960.

Browne, Michael, W. A comparison of factor analytic techniques. *Psychometrika,* 1968, **33**, 267–334.

Browne, Michael W. Fitting the factor analysis model. *Psychometrika,* 1969, **34,** 375–394.

Camp, B. H. The converse of Spearman's two-factor theorem. *Biometrika,* 1932, **24,** 418–427.

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

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

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

Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika,* 1953, **18,** 277–296.

Harman, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

Harman, H. H. *Modern factor analysis.* Second Edition (Revised). Chicago: University of Chicago Press, 1967.

Harris, C. W. On factors and factor scores. *Psychometrika,* 1967. **32,** 363–379,

Harris, C. W. Some Rao-Guttman relationships. *Psychometrika,* 1962, **27,** 247–263.

Heermann, E. F. The geometry of factorial indeterminacy. *Psychometrika,* 1964, **29,** 371–381.

Heermann, E. F. The algebra of factorial indeterminacy. *Psychometrika,* 1966, **31,** 539–543.

Hendrickson, A. E., and White, P. O. PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology,* 1964, **17,** 65–70.

Horst, Paul. Generalized Factor Analysis. Part I. Technical Report. Seattle, Washington, 1969.

Householder, A. S. *Theory of matrices in numerical analysis.* New York: Ginn, 1963.

Howe, W. G. Some contributions to factor analysis. Report No. ORNL-1919, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1955.

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

Jöreskog, K. G. *Statistical estimation in factor analysis.* Stockholm: Almqvist & Wiksell, 1963.

Kaiser, H. F. Image analysis. In: Harris, C. W. (Ed.), *Problems in measuring change.* University of Wisconsin Press, 1963.

Kestelman, H. The fundamental equation of factor analysis. *British Journal of Psychology, Statistical Section,* 1952, **5,** 1–6.

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

Lawley, D. N. Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh,* Section A, 1942, **61,** 176–185.

Lawley, D. N. The application of the maximum likelihood method to factor analysis. *British Journal of Psychology,* 1943, **33,** 172–175.

Lawley, D. N., and Maxwell, A. E. *Factor analysis as a statistical method.* London: Butterworths, 1963.

Ledermann, W. The orthogonal transformation of a factorial matrix into itself. *Psychometrika,* 1938, **3,** 181–187.

Lord, F. M. A study of speed factors in tests and academic grades. *Psychometrika,* 1956, **21,** 31–50.

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

McDonald, R. P., and Burr, E. J. A comparison of four methods of construction factor scores. *Psychometrika,* 1967, **32,** 381–401.

Pawlik, K. *Dimensionen des verhaltens.* Bern: H. Huber, 1968.

Piaggio, H. T. H. Three sets of conditions necessary for the existence of a g that is real and unique except in sign. *British Journal of Psychology*, 1933, 24, 88–105.

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

Schönemann, Peter H. The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, 1971, **36**, 21–30.

Thomson, G. H. The definition and measurement of "g" (general intelligence). *Journal of Educational Psychology*, 1935, 26, 241–262.

Thurstone, L. L. The dimensions of temperament. *Psychometrika*, 1951, **16**, 11–20.

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

Manuscript received 6/16/70
Revised manuscript received 2/1/71