FITTING THE FACTOR ANALYSIS MODEL*

MICHAEL W. BROWNE

EDUCATIONAL TESTING SERVICE†

When the covariance matrix \( \Sigma(p \times p) \) does not satisfy the formal factor analysis model for \( m \) factors, there will be no factor matrix \( \Lambda(p \times m) \) such that \( T = (\Sigma - \Lambda\Lambda') \) is diagonal. The factor analysis model may then be replaced by a tautology where \( T \) is regarded as the covariance matrix of a set of "residual variates." These residual variates are linear combinations of "discarded" common factors and unique factors and are correlated. Maximum likelihood, alpha and iterated principal factor analysis are compared in terms of the manner in which \( T \) is defined, a "maximum determinant" derivation for alpha factor analysis being given. Weighted least squares solutions using residual variances and common variances as weights are derived for comparison with the maximum likelihood and alpha solutions. It is shown that the covariance matrix \( T \) defined by maximum likelihood factor analysis is Gramian, provided that all diagonal elements are nonnegative. Other methods can define a \( T \) which is nonGramian even when all diagonal elements are nonnegative.

The fundamental postulate in factor analysis is

\[
(1) \quad x = \mu + \Lambda^*z^* + e^*
\]

where \( x(p \times 1) \) is a real vector variate representing \( p \) observable variates, \( \mu(p \times 1) \) is the mean of the distribution of \( x \), \( \Lambda^*(p \times k) \) is a real matrix of factor loadings, \( z^*(k \times 1) \) is a real vector variate representing common factors, and \( e^*(p \times 1) \) is a real vector variate representing unique factors. It is assumed that

\[
(2) \quad \varepsilon(z^*) = 0 \\
(3) \quad \varepsilon(e^*) = 0 \\
(4) \quad \varepsilon(z^*e^{**}) = 0 \\
(5) \quad \varepsilon(e^*e^{**}) = D^*_e
\]

where \( D^*_e \) is a diagonal matrix of order \( p \). This paper will be primarily concerned with the case where the factor variates are defined as being uncorrelated and having unit variances so that

\[
(6) \quad \varepsilon(z^*z^{**}) = I.
\]

*A modified version of this paper forms part of a Ph.D. thesis submitted to the University of South Africa.
†Presently at the National Institute for Personnel Research, South Africa.
From these assumptions it follows that the population dispersion matrix, $\Sigma$, of $x$ has the form

$$
\Sigma = \Lambda^*\Lambda^* + D^*.
$$

It is well known that any $\Lambda^*(p \times k)$ satisfying (7) is not unique. If $\Lambda^*$ satisfies (7), then so does $(\Lambda^*\theta^*)$ where $\theta^*(k \times k)$ is any orthonormal matrix. The diagonal matrix $D^*$, where $(\Sigma - D^*)$ is positive semidefinite of rank $k$, is not necessarily unique. Conditions for $D^*$ to be unique are given by Anderson and Rubin [1956].

If the number, $k$, of factors satisfies Lederman's [1937] inequality

$$
k < \frac{2p + 1 - \sqrt{8p + 1}}{2},
$$
it will not be possible to find a $\Lambda^*(p \times k)$ such that $(\Sigma - \Lambda^*\Lambda^*)$ is diagonal unless the elements of $\Sigma$ satisfy certain equalities. In other words, when $k$ satisfies Lederman’s inequality, (7) represents a model which is capable of being contradicted by actual data as opposed to a tautology which will always be satisfied.

On the grounds that practical instances where a population covariance matrix, $\Sigma$, satisfies (7), with $k$ small, have a “personal probability equal to zero,” Kaiser and Caffrey [1965] suggest that it is more realistic to allow $k$ to be arbitrarily large. If $k$ is greater than $(p - 2)$, it is always possible to find a $D^*$ and a $\Lambda^*(p \times k)$ which satisfy (7). Therefore, when $k$ is permitted to be arbitrarily large, (7) represents a tautology. It can provide a description of the observed variables, but it does not represent a testable model. When $k$ is large, one is concerned with the “definition” of $\Lambda^*$ rather than with the “estimation” of $\Lambda^*$.

Kaiser and Caffrey [1965] suggest that a small subset of common factors be retained and that the remaining common factors be discarded. The number of factors to be retained will be denoted by $m$, while $k$ will stand for a number of factors sufficiently large for (7) to hold. It will be assumed that $m$ is small relative to $p$ while $k \geq m$ need not be small. It is possible to regard $k$ as the number of common factors of the “universe of variables of interest” [Kaiser & Caffrey, 1965].

Let $\Lambda^*$ and $z^*$ be partitioned as follows:

$$
\Lambda^*(p \times k) = [\Lambda(p \times m); \Lambda^{(d)}(p \times [k - m])]
$$

$$
z^*(1 \times p) = [z'(1 \times m); z^{(d)}'(1 \times [k - m])]
$$

$$
= [z_1, z_2, \ldots, z_m; z_{m+1}, \ldots, z_k].
$$

The “retained” common factors will be denoted by $z(m \times 1)$, and the “discarded” common factors by $z^{(d)}$ while $\Lambda$ and $\Lambda^{(d)}$ will represent the cor-