PSYCHOMETRIKA—VOL. 45, NO. 2.
JUNE, 1980

INEQUALITIES AMONG LOWER BOUNDS TO RELIABILITY: WITH APPLICATIONS TO TEST CONSTRUCTION AND FACTOR ANALYSIS

P. M. BENTLER AND J. ARTHUR WOODWARD

UNIVERSITY OF CALIFORNIA, LOS ANGELES

A chain of lower-bound inequalities leading to the greatest lower bound to reliability is established for the internal consistency of a composite of unit-weighted components. The chain includes the maximum split-half coefficient, the lowest coefficient consistent with nonimaginary common factors, and the lowest coefficient consistent with nonimaginary common and unique factors. Optimization theory is utilized to determine the conditions that are requisite for the inequalities. Convergence proofs demonstrate that the coefficients can be attained. Rapid algorithms obtain estimates of the coefficients with sample data. The theory yields methods for splitting items into maximally similar sets and for exploratory factor analysis based on a theoretical solution to the communality problem.

Key words: reliability, internal consistency, test theory, factor analysis.

1. Introduction

Consider the well-known factor analytic model for an $n \times n$ population covariance matrix

$$\Sigma = AA' + \Omega^2 + E^2 = AA' + Z^2, \tag{1.1}$$

where A is an $n \times r$ common factor loading matrix and Ω^2 and E^2 are nonnegative definite diagonal matrices of specific and error variances, respectively. The reliability of a unit weighted composite of the n components may be defined as

$$\rho_{xx} = 1 - \frac{1'E^2 1}{1'\Sigma 1}, \tag{1.2}$$

where 1 is the unit vector of weights. Alternatively, this coefficient can be defined on the random vectors underlying (1.1), but we restrict ourselves to the level of second moments because of a functional equivalence to the random vector approach. In view of the equality $Z^2 = \Omega^2 + E^2$, it follows that $1'Z^2 1 \geq 1'E^2 1$. Consequently, upon defining internal consistency as

$$\rho_I = 1 - \frac{1'Z^2 1}{1'\Sigma 1}, \tag{1.3}$$

we obtain

$$\rho_I \leq \rho_{xx} \tag{1.4}$$

This investigation was supported in part by a Research Scientist Development Award (K02-DA00017) and a research grant (DA01070) from the U.S. Public Health Service. The manuscript profited substantially from the critiques of several anonymous reviewers, whose assistance is gratefully acknowledged.

Requests for reprints should be sent to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California 90024.

0033-3123/80/0600-2649$00.75/0

based on the unique variances Z^2 and the factor analytic model $\Sigma = AA' + Z^2$. This result is well-known, but of theoretical interest only since it represents a statement among unknown quantities. We assume that Σ is known, or at least well-estimated, but that the specific decomposition (1.1) is not known. Thus, we must turn to coefficients that can be defined abstractly as well as computationally. We shall develop new lower bounds to ρ_{xx} via an extension of the minimum trace factor analysis theory introduced by Bentler [1972], utilizing the positive semidefiniteness of AA' [Bentler & Woodward, Note 1]. No assumptions about the rank r of A are made. Traditional psychometric theory, in contrast, makes rather strong assumptions not only about A but also about E^2 [e.g., Lord & Novick, 1968].

Although the decomposition (1.1) is not known, useful lower bounds to reliability can be developed under the choice of diagonal matrix Ψ^2 in tautological expressions of the form $\Sigma = (\Sigma - \Psi^2) + \Psi^2$. Assuming Σ to be fixed, we define the scalar function $\rho = f(\Psi^2) = 1 - 1'\Psi^2 1/1'\Sigma 1$, the domain of which is the set of Ψ^2 matrices consistent with the previous tautology. Under certain restrictions on the domain of this function, useful minima can be identified that have the property that $\min \rho \leq \rho_I$. The first, following Bentler [1972], is the smallest possible $f(\Psi^2)$

$$\rho_I = \min f(\Psi^2) = 1 - \frac{1'\Psi_I^2 1}{1'\Sigma 1} \quad (\Sigma = \Lambda_I\Lambda_I' + \Psi_I^2) \tag{1.5}$$

having $\Lambda_I\Lambda_I' = (\Sigma - \Psi_I^2)$, i.e., nonimaginary common factors. The second, developed here and independently by Jackson and Agunwamba [1977] and Woodhouse and Jackson [1977], is the smallest possible $f(\Psi^2)$ having nonimaginary common and unique factors, that is,

$$\rho_+ = \min f(\Psi^2) = 1 - \frac{1'\Psi_+^2 1}{1'\Sigma 1} \quad (\Sigma = \Lambda_+\Lambda_+' + \Psi_+^2, \psi_p^2 \geq 0) \tag{1.6}$$

where $p = 1, \cdots, n$. We have purposely given the Λ and Ψ^2 matrices in (1.5) and (1.6) different subscripts to clarify that the matrices minimizing $f(\Psi^2)$ under different restrictions on the domain of possible Ψ^2 matrices need not be the same. The inequalities

$$\rho_I \leq \rho_+ \leq \rho_I \tag{1.7}$$

are easily shown to hold. Since A and Z^2 are possible candidates for Λ_+ and Ψ_+^2, $\rho_+ \leq \rho_I$. Coefficient ρ_+ may be called the greatest lower bound to internal consistency reliability since, without additional information about the model, the Λ_+ and Ψ_+^2 matrices of (1.6) cannot be rejected as candidates for the unknown parameters A and Z^2 of model (1.1). Furthermore, the domain of unrestricted diagonal matrices Ψ^2 is obviously greater than the domain of nonnegative definite Ψ^2 matrices. The greater range of $f(\Psi^2)$ in the former case leads to wider extrema and to $\rho_I \leq \rho_+$ in particular. In the body of this paper we develop a larger chain of inequalities that includes (1.7), show how to obtain the various coefficients computationally, and describe some implications of the results.

2. *Theoretical Results*

We shall develop a useful chain of inequalities that can be applied to any arbitrary decomposition of Σ yielding nonimaginary common factors. To fix ideas, let us consider the decomposition $\Sigma = \Lambda\Lambda' + \Psi^2$, where Λ and the diagonal matrix Ψ^2 are arbitrary. Corresponding to this decomposition, we may define the internal consistency coefficient

$$\rho = 1 - \frac{1'\Psi^2 1}{1'\Sigma 1}. \tag{2.1}$$

We shall develop a chain of coefficients, none of which exceed ρ. Note that ρ is not being

treated as a function since Ψ^2 does not vary. It represents an arbitrary value of the function $f(\Psi^2)$, under a particular restriction on its domain. Since there are in principle many potential coefficients ρ, depending on given decompositions $\Sigma = \Lambda\Lambda' + \Psi^2$, it will be useful later to identify ρ with ρ_l or ρ_+, thus yielding a chain of lower bounds to reliability. The computational theory associated with these bounds is left to another section.

We shall proceed by establishing a basic inequality that will be useful in generating still stronger results. Consider all members $\bar{N}_i$ of the set Π_i of n by n, unit diagonal, positive-semidefinite matrices having rank r_i ($r_i \leq i < n$). These matrices can be decomposed as $\bar{N}_i = \bar{T}_i\bar{T}_i'$, where diag $(\bar{N}_i) = I$. Any such matrix $\bar{T}_i$ of order $n \times i$ and rank $r_i \leq i$ can be used to obtain

$$\bar{T}_i'\Sigma\bar{T}_i = \bar{T}_i'\Lambda\Lambda'\bar{T}_i + \bar{T}_i'\Psi^2\bar{T}_i , \tag{2.2}$$

but, in view of the constraints on $\bar{T}_i$, the sum of diagonal elements yields

$$\mathrm{tr}\bar{T}_i'\Sigma\bar{T}_i = \mathrm{tr}\bar{T}_i'\Lambda\Lambda'\bar{T}_i + l'\Psi^2 l . \tag{2.3}$$

It follows from the positive semidefiniteness of $(\Sigma - \Psi^2)$ that

$$\mathrm{tr}\bar{T}_i'\Sigma\bar{T}_i \geq l'\Psi^2 l , \tag{2.4}$$

an inequality that serves as the basis for much of the following development. We may define a coefficient based on (2.4) as

$$\rho_{\bar{T}_i} = 1 - \frac{\mathrm{tr}\bar{T}_i'\Sigma\bar{T}_i}{l'\Sigma l} . \tag{2..5}$$

In view of (2.1) and (2.4), we obtain the inequality

$$\rho_{\bar{T}_i} \leq \rho \tag{2.6}$$

for all coefficients (2.5) based on matrices $\bar{T}_i$ such that $\bar{T}_i\bar{T}_i' \in \Pi_i$. We now make use of three theoretical principles to expand the basic inequality (2.6), not only in order to elucidate dimensions along which coefficients vary, but also in order to develop the greatest lower bound to reliability. For theoretical completeness we shall define a chain of coefficients that ranges from the least lower bound of the form (2.5) to the greatest in (2.6).

As a first principle, we can develop criteria for choosing a particular $\bar{N}_i$ from the set Π_i. Different criteria will define coefficients that can be ordered relative to each other and to ρ. One particular matrix $\bar{N}_i$ can generate a least lower bound of the form (2.5), while another matrix $\bar{N}_i$ can generate the best possible inequality in (2.6), i.e, the latter matrix allows $\rho_{\bar{T}_i}$ to approach ρ as closely as possible. Our second principle involves selection of the set of matrices Π_i. By expanding the rank of Π_i to yield a set Π_j of rank r_j such that $r_i \leq r_j \leq j$ and $\Pi_i \subseteq \Pi_j$, we are able to assess the consequences of selecting a given matrix $N_j \in \Pi_j$ on a coefficient $\rho_{\bar{T}_j}$ defined comparably to $\rho_{\bar{T}_i}$ in (2.5). As before, different matrices $\bar{N}_j$ lead to different coefficients that also can be ordered relative to each other and to ρ. The combination of these two dimensions of choice defines a still larger field of internal consistency coefficients, some of which are known and some of which are new. In presenting these coefficients, we concentrate on a systematization that yields inequalities that include known coefficients such as Cronbach's [1951] α, Guttman's [1945] split-half coefficients λ_4, and Bentler's [1972] dimension-free lower-bound coefficient. Special attention will be paid to the endpoints of the inequalities, in order to clarify the least and greatest lower bounds. Finally, if Ψ^2 is singular, these inequalities will not lead to the greatest lower bound. Consequently, we introduce as a third principle a different type of expansion of the set of matrices Π_i to a new set $\tilde{\Pi}_i$ such that $\Pi_i \subseteq \tilde{\Pi}_i$. The matrices $\tilde{N}_i$ in this set

$\tilde{\Pi}_i$ only have partial unit diagonals. Choice of a particular matrix $\tilde{N}_i$ will allow (2.6) to become an equality that can, through (1.7), potentially yield the greatest lower bound.

We are now ready to invoke these three principles for expanding (2.6) into a more complete set of inequalities. In the first place, consider the problem of selecting matrices $N_{i'} = T_{i'}T_{i'}' \in \Pi_i$ and $N_i = T_iT_i' \in \Pi_i$ that maximize $\mathrm{tr}T_{i'}'\Sigma T_{i'}$ and minimize $\mathrm{tr}T_i'\Sigma T_i$, respectively. These optimizing matrices $T_{i'}$ and T_i, which may not be unique, have the property that

$$\mathrm{tr}T_{i'}'\Sigma T_{i'} \geq \mathrm{tr}\tilde{T}_i'\Sigma\tilde{T}_i \geq \mathrm{tr}T_i'\Sigma T_i \geq 1'\Psi^2 1 \ , \tag{2.7}$$

in view of (2.4). It follows immediately that

$$\rho_{i'} \leq \rho_{\tilde{T}_i} \leq \rho_i \leq \rho \ , \tag{2.8}$$

where $\rho_{i'}$ and ρ_i are defined by the use of $T_{i'}$ and T_i in (2.5). Equation (2.8) defines the least and greatest coefficients possible by considering only the method (2.2) through (2.6) based on the set Π_i. In order to expand the chain (2.8), we shall have to expand the set of matrices leading to (2.8).

Consider next the set Π_j, $\Pi_i \subseteq \Pi_j$, such that $\tilde{N}_j = \tilde{T}_j\tilde{T}_j' \in \Pi_j$ is a matrix with unit diagonal elements, of rank r_j where $n > j > i, j \geq r_j \geq r_i$; $\tilde{T}_j$ is of order $n \times j$. Again we select the matrices $N_{j'} = T_{j'}T_{j'}' \in \Pi_j$ and $N_j = T_jT_j' \in \Pi_j$ that maximize $\mathrm{tr}T_{j'}'\Sigma T_{j'}$ and minimize $\mathrm{tr}T_j'\Sigma T_j$, respectively. Matrices $T_{j'}$ and T_j need not be unique. Functions defined on the set $[\Pi_j,\Sigma]$ can obviously take on more extreme values than functions defined on $[\Pi_i,\Sigma]$, considering that $[\Pi_i,\Sigma] \subset [\Pi_j,\Sigma]$. Thus,

$$\mathrm{tr}T_{j'}'\Sigma T_{j'} \geq \mathrm{tr}T_{i'}'\Sigma T_{i'} \geq \mathrm{tr}T_i'\Sigma T_i \geq \mathrm{tr}T_j'\Sigma T_j \geq 1'\Psi^2 1 \ , \tag{2.9}$$

utilizing (2.4) and (2.7). To see (2.9) more graphically, note that there exist matrices $\tilde{T}_j$ that can always be partitioned as $[\tilde{T}_i 0]$, unless $\Pi_i = \Pi_j$. For example, if one takes T_i and T_j to be uniquely specified in partial lower triangular form, the nonnull elements of these minimizing matrices are identical unless there exists a matrix T_j of rank $r_j > r_i$ such that $\mathrm{tr}T_j'\Sigma T_j < \mathrm{tr}T_i'\Sigma T_i$. By a similar argument we may consider the set of matrices Π_k, $\Pi_k \supseteq \Pi_j$, such that $\tilde{N}_k = \tilde{T}_k\tilde{T}_k' \in \Pi_k$, with $\tilde{T}_k$ of order $n \times k$ being row-normalized with rank r_k, where $n > k > j > i$ and $k \geq r_k \geq r_j \geq r_i$. For uniqueness of specification, we shall define the set Π_k to be that set for which $\mathrm{tr}\tilde{T}.'\Sigma\tilde{T}.$ reaches the absolute extreme values for a $\tilde{T}.$ such that $\tilde{T}.\tilde{T}.' = \tilde{N}. \in \Pi.$ with $\mathrm{diag}(\tilde{N}.) = I$. Thus, we need not consider the set Π_{k+1}. Using a notation comparable to that given in (2.9), it follows that for $1 < i < j < k < n$,

$$\begin{aligned}\mathrm{tr}T_{k'}'\Sigma T_{k'} \geq \mathrm{tr}T_{j'}'\Sigma T_{j'} \geq \mathrm{tr}T_{i'}'\Sigma T_{i'} \geq \mathrm{tr}T_{1'}'\Sigma T_{1'} \geq \\ \mathrm{tr}T_1'\Sigma T_1 \geq \mathrm{tr}T_i'\Sigma T_i \geq \mathrm{tr}T_j'\Sigma T_j \geq \mathrm{tr}T_k'\Sigma T_k \geq 1'\Psi^2 1 \ .\end{aligned} \tag{2.10}$$

The corresponding internal consistency inequalities are given by

$$\rho_{k'} \leq \rho_{j'} \leq \rho_{i'} \leq \rho_{1'} \leq \rho_1 \leq \rho_i \leq \rho_j \leq \rho_k \leq \rho. \tag{2.11}$$

If Ψ^2 is not singular, ρ_k is the greatest lower bound to ρ obtainable in this manner, since there exists no $T.$ such that $\mathrm{tr}T_k'\Sigma T_k > \mathrm{tr}T.'\Sigma T. \geq 1'\Psi^2 1$.

If the unique variance matrix Ψ^2 is singular, however, we can use our third principle to obtain inequalities sharper than (2.11); although applicable to any coefficient (2.11), it is of greatest interest in improving ρ_k. Suppose Ψ^2 consists of the submatrices Ψ_1^2 and Ψ_2^2, where $\Psi_1^2 \neq 0$ and $\Psi_2^2 = 0$, i.e, that the inequality $1 \leq \mathrm{rank}\ (\Psi_1^2) < n$ holds. Now consider given members $\tilde{N}_k = \tilde{T}_k\tilde{T}_k'$ of the set $\tilde{\Pi}_k$ of rank $r_k \leq k$ matrices that are of the form $\tilde{T}_k' = [\tilde{T}_{k1}', \tilde{T}_{k2}']$, where $\mathrm{diag}(\tilde{T}_{k1}\tilde{T}_{k1}') = I$ and there is no constraint on $\mathrm{diag}(\tilde{T}_{k2}\tilde{T}_{k2}')$. Among such matrices we are particularly concerned with the (possibly nonunique) matrix $T_K' = [T_{k1}', T_{k2}']$ that minimizes $\mathrm{tr}T_K'\Sigma T_K$. Since the set $\tilde{\Pi}_k$ is less restricted than the set Π_k, i.e., $\Pi_k \subseteq \tilde{\Pi}_k$, it follows from standard optimization theory that

$$\text{tr}T_k'\Sigma T_k \geq \text{tr}T_K'\Sigma T_K . \tag{2.12}$$

Then, by cyclic permutation of matrices under the trace operator,

$$\text{tr}T_K'\Sigma T_K = \text{tr}T_K'\Lambda\Lambda' T_K + \text{tr}\Psi^2 T_K T_K' . \tag{2.13}$$

However,

$$\text{tr}\Psi^2 T_K T_K' = \text{tr}\Psi_1^2 T_{k1} T_{k1}' + \text{tr}\Psi_2^2 T_{k2} T_{k2}' = 1'\Psi^2 1 , \tag{2.14}$$

so that (2.12), (2.13) and (2.14) yield the result

$$\text{tr}T_k'\Sigma T_k \geq \text{tr}T_K'\Sigma T_K \geq 1'\Psi^2 1 . \tag{2.15}$$

Consequently, if we define

$$\rho_K = 1 - \frac{\text{tr}T_K'\Sigma T_K}{1'\Sigma 1} , \tag{2.16}$$

we obtain, along with (2.11), the inequalities

$$\rho_k \leq \rho_K \leq \rho . \tag{2.17}$$

We adopt the convention that Ψ_2^2 can be nonexistent, i.e., that Ψ^2 can be equivalent to Ψ_1^2. Consequently, $\rho_k = \rho_K$ when $\Psi^2 = \Psi_1^2$.

By a similar argument, we could determine $\tilde{N}_{K'} = \tilde{T}_{K'}\tilde{T}_{K'}' \in \tilde{\Pi}_K$ such that $\text{tr}T_{K'}'\Sigma T_{K'}$ is maximized. Since $\Pi_k \subseteq \tilde{\Pi}_K$, we have

$$\underline{\text{tr}T_{K'}'\Sigma T_{K'}} \geq \text{tr}T_{k'}'\Sigma T_{k'} . \tag{2.18}$$

Using steps similar to (2.13)–(2.15), and defining

$$\rho_{K'} = 1 - \frac{\text{tr}T_{K'}'\Sigma T_{K'}}{1'\Sigma 1} , \tag{2.19}$$

we obtain the inequality yielding the least lower bound of the chain (2.11) as

$$\rho_{K'} \leq \rho_{k'} . \tag{2.20}$$

Collecting the major results (2.11), (2.17), and (2.20), we have

$$\rho_{K'} \leq \rho_{k'} \leq \underline{\rho_{1'} \leq \rho_1} \leq \rho_k \leq \rho_K \leq \rho . \tag{2.21}$$

Intermediate coefficients in (2.11) have been omitted from (2.21), and the self-evident steps (2.12)–(2.16) and (2.18)–(2.19) were not implemented for all possible sets Π_i when Ψ^2 is singular. The inequality $\rho_K \leq \rho$ will yield an equality when

$$\Lambda' T_K = 0 . \tag{2.22}$$

According to (2.13)–(2.15), one then obtains $\text{tr}T_K'\Sigma T_K = 1'\Psi^2 1$, so that $\rho_K = \rho$.

Although the chain (2.21) was developed by considering a given decomposition $\Sigma = \Lambda\Lambda' + \Psi^2$ consistent with (2.1), it will be noted that the various coefficients $\rho_{k'}$ to ρ_k are defined completely independently of Λ or Ψ^2. Thus, these lower bounds can be implemented by simply considering the relevant T_i matrices, as will be shown in the next section. Coefficients $\rho_{K'}$ and ρ_K do require knowledge of the variables associated with $\Psi_2^2 = 0$. While our primary interest lies in the right half of (2.21), $\rho_{K'}$ is likely to have some important applications. For example, one might expect the sampling distribution of estimated lower bound coefficients to differ when comparing two population matrices Σ_1 and Σ_2 having the same ρ_K but widely different $\rho_{K'}$ coefficients.

It may be observed that $\rho_{1'}$ and ρ_1 can be considered to be split-half coefficients, where the number of items in each "half" is not necessarily equal, and that the interval $\rho_{1'}$ to ρ_1 is, according to (2.8) and (2.21), covered by various split-half coefficients. Note that

$\bar{T}_1$ is a sign vector. Let v and w represent $n \times 1$ vectors containing only the elements having a given sign in $\bar{T}_1$ with zeros elsewhere. Then $v'w = 0$, $v + w = \bar{T}_1$, and v and $(-w)$ represent positive (0,1) weight vectors. The composites formed by the vectors $v = (1/2)(I + \bar{T}_1)$ and $(-w) = (1/2)(I - \bar{T}_1)$ have variances given by $v'\Sigma v = (1/4)(I'\Sigma I + 2\bar{T}_1'\Sigma I + \bar{T}_1'\Sigma\bar{T}_1)$ and $w'\Sigma w = (1/4)(I'\Sigma I - 2\bar{T}_1'\Sigma I + \bar{T}_1'\Sigma\bar{T}_1)$. Using Guttman's [1945] definition of $\lambda_4 = 2[1 - (\sigma_1^2 + \sigma_2^2)/\sigma_t^2]$, where σ_1^2 and σ_2^2 are component variances and σ_t^2 is the variance of the composite, and making appropriate substitutions, yields $\lambda_4 = \rho_{\bar{T}_1}$. Thus, all $\rho_{\bar{T}_1}$ coefficients are λ_4 coefficients, with $\rho_{1'}$ being the least such coefficient and ρ_1 being the best possible split-half coefficient. Jackson and Agunwamba [1977] and Woodhouse and Jackson [1977] use the more restricted definition $\lambda_4 = \rho_1$. It would seem that the importance of $\rho_{1'}$ is that, when it is large, any randomly chosen split of the components will yield a $\rho_{\bar{T}_1}$ that is large; however, the practical value of $\rho_{1'}$ and lesser coefficients is not known, particularly since they are negative.

We can also relate $\rho_{1'}$ and ρ_1 to Cronbach's [1951] α, which is defined as $\alpha = [n/(n-1)][1 - (I'D_\Sigma I)/(I'\Sigma I)]$, where $D_\Sigma = \text{diag}(\Sigma)$. When n is even, α is the average of all possible split-half coefficients formed from composites based on $n/2$ items; such an average is clearly neither smaller than $\rho_{1'}$ nor larger than ρ_1. However, it is more accurate to state that

$$\rho_{1'} \lesssim \alpha \lesssim \rho_1 , \tag{2.23}$$

where the symbol $\lesssim$ denotes an inequality that is generally, but not always, true [Bentler & Woodward, Note 1; Jackson, 1979]. The approximate inequality (2.23) can be replaced by $\rho_{1'} \leq \alpha \leq \rho_1$ whenever α is the average of all split-half coefficients based on $n/2$ item composites, namely, when n is an even number. A counterexample to the inequality $\alpha \leq \rho_1$ can simply be found: the three-variable equicorrelation matrix with positive correlations has the property that $\alpha > \rho_1$. The remainder of the paper will concentrate on lower-bounds (2.21) beginning with ρ_1. Although α cannot be placed into the chain (2.21), Bentler [1972] proved $\alpha \leq \rho_I$. It follows immediately that $\alpha \leq \rho_+$.

Since ρ_+ is one possible choice of ρ, we obtain the final result

$$\rho_{K'} \leq \cdots \leq \rho_k \leq \rho_K \leq \rho_+ \leq \rho_I \tag{2.24}$$

by considering (1.4), (1.7), and (2.21). These coefficients form a single continuum, unlike those reviewed by Jackson and Agunwamba [1977]. While most of their coefficients can be obtained algebraically, those in (2.24) must be obtained by iterative calculations. The computational theory for (2.24) is given in the next section; Woodhouse and Jackson [1977] utilize a different approach. We shall find coefficients of the form $\rho_k = \rho_I$ and $\rho_K = \rho_+$.

3. Optimization Theory

The problem of selecting a matrix T_i $(i = 1, \cdots, k)$ that minimizes the function $\text{tr}T_i'\Sigma T_i$ is illuminated by the calculus. A function that defines the problem analytically while taking account of the necessary constraints is

$$f_i = \text{tr}T_i'\Sigma T_i - \text{tr}\Delta_i[\text{diag}(T_iT_i') - I] , \tag{3.1}$$

where Δ_i is a diagonal matrix of Lagrangian multipliers. Differentiating (3.1) with respect to T_i, and equating the result to zero, yields

$$(\Sigma - \Delta_i)T_i = 0 . \tag{3.2}$$

Evidently, T_i is only defined up to an orthonormal transformation. To specify T_i uniquely (except for column sign), it can be taken in echelon form as partially lower triangular. If

the smallest possible rank of $(\Sigma - \Delta_i)$ under choice of Δ_i is r^*, the maximum rank r_i of T_i is $(n - r^*)$. If $i > (n - r^*)$, the T_i solving (3.2) will have some null columns. For the T_1 special case of (3.2), see Woodhouse and Jackson [1977], equation (6.4). The matrix T_r that maximizes $\mathrm{tr}T_r'\Sigma T_r$ must obviously meet the condition $(\Sigma - \Delta_r)T_r = 0$ analogous to (3.2). A more detailed study of the function (3.1) is needed to generate computational algorithms applicable to the general case.

Extrema of function (3.1) can always be attained with any arbitrary starting matrix $\tilde{T}_i$. Since the convergence proof does not rely on any particular choice i of column dimensionality for T_i, nor upon the minimization or maximization of (3.1), we drop the subscript i in this development. It may be noted that function (3.1) can equivalently be written as

$$f = \sigma^2 t't + 2\sigma_0'Tt + \mathrm{tr}T^{0'}\Sigma^0 T^0 - \Delta(t't - 1) - \mathrm{tr}\Delta^0(T^0T^{0'} - I), \tag{3.3}$$

where σ^2 is the p^{th} diagonal element of Σ, t' is the p^{th} row of T, σ_0' is the p^{th} row of $(\Sigma - D_\Sigma)$, T^0 is a matrix of the $m = (n - 1)$ rows that remain after removing t' from T, Σ^0 and Δ^0 are the $(m \times m)$ matrices resulting when the p^{th} row and column are eliminated from Σ and Δ_i respectively, and Δ is the Lagrangian multiplier corresponding to the p^{th} constraint. Consequently, $f = g + z$, where

$$g = \sigma^2 t't + 2\sigma_0'Tt - \Delta(t't - 1), \tag{3.4}$$

and where z is constant with respect to t. Function (3.4) can be written for any arbitrary p^{th} row of T $(p = 1, \cdots, n)$. It can be shown that

$$\partial g/\partial t = 2T'\sigma_0 + 2\sigma^2 t - 2\Delta t. \tag{3.5}$$

Premultiplying (3.5) by t' and solving $\partial g/\partial t = 0$ under the constraint $t't = 1$ yields

$$(\sigma^2 - \Delta) = -(\sigma_0'Tt). \tag{3.6}$$

As a consequence, the function can be written as

$$g = \sigma^2 - 2(\sigma^2 - \Delta), \tag{3.7}$$

which is obviously minimized when $(\sigma^2 - \Delta) \geq 0$ and maximized when $(\sigma^2 - \Delta) \leq 0$. If $T'\sigma_0 = 0$, (3.6) shows that $(\sigma^2 - \Delta) = 0$, and $\partial g/\partial t = 0$ can be solved by an arbitrary t. The more general solution for t under $\partial g/\partial t = 0$ is

$$t = -T'\sigma_0(\sigma^2 - \Delta)^{-1}. \tag{3.8}$$

Since t must be normalized under (3.4), $(\sigma^2 - \Delta)$ simply represents a normalizing constant, and

$$t = \beta T'\sigma_0(\sigma_0'TT'\sigma_0)^{-1/2} \tag{3.9}$$

represents a general solution for t, with $\beta = -1$ when minimizing (3.4) and $\beta = +1$ when maximizing (3.4). The important feature of (3.9) is that it expresses t as a function of T^0 (i.e., all the other rows of T except t') and σ_0. Consequently, a converging algorithm is easily derived.

We have now laid the groundwork for the p^{th} row-replacement algorithm $(p = 1, \cdots, n)$

$$\begin{aligned} &\text{(a) } t^{(i+1)} = t^{(i)} && \text{if } \sigma_0'TT'\sigma_0 < \varepsilon \\ &\text{(b) } t^{(i+1)} = \beta T'\sigma_0(\sigma_0'TT'\sigma_0)^{-1/2} && \text{otherwise.} \end{aligned} \tag{3.10}$$

We prove that (3.10) must converge for any row-normalized starting matrix T and for a suitably small ε. The sign β determines the direction of change in the objective function.

We show that the function (3.4) cannot change in the undesired direction from iteration (i) to $(i+1)$. Since T and $t^{(i+1)}$ are normalized, (3.4) takes on the values

$$g^{(i)} = \sigma^2 + 2\sigma_0' Tt^{(i)} \text{ and } g^{(i+1)} = \sigma^2 + 2\sigma_0' Tt^{(i+1)}. \tag{3.11}$$

Assume first that $\sigma_0' TT'\sigma_0 < \varepsilon$; then $g^{(i)} = g^{(i+1)} = \sigma^2$ and the function is stable. Otherwise,

$$g^{(i+1)} = \sigma^2 + 2\beta(\sigma_0' TT'\sigma_0)^{1/2}. \tag{3.12}$$

Thus, the change in function can be represented by

$$\delta(g) = \tfrac{1}{2}(g^{(i)} - g^{(i+1)}) = \sigma_0' Tt^{(i)} - \beta(\sigma_0' TT'\sigma_0)^{1/2}. \tag{3.13}$$

which must be nonnegative in minimizing (3.4) and nonpositive in maximizing (3.4). Since $\sigma_0' TT'\sigma_0 = \text{tr}(T'\sigma_0\sigma_0' T)$, it follows from the singular value representation $T'\sigma_0\sigma_0' T = \Gamma\gamma^2\Gamma'$ ($\Gamma'\Gamma = 1$) that $(\sigma_0' TT'\sigma_0)^{1/2} = \gamma > 0$. Similarly, for Γ appropriately oriented, $\sigma_0' Tt^{(i)} = \gamma(t^{(i)\prime}\Gamma)$. Consequently,

$$\delta(g) = \gamma(t^{(i)\prime}\Gamma - \beta). \tag{3.14}$$

However, the inner product of two normal vectors yields $-1 \leq t^{(i)\prime}\Gamma \leq 1$. Hence

$$\begin{aligned} \delta(g) &\geq 0 \ (\beta = -1) \\ \delta(g) &\leq 0 \ (\beta = +1). \end{aligned} \tag{3.15}$$

Thus, application of algorithm (3.10) to any arbitrary row of T yields an improvement in the objective function unless (3.4) is already optimized. By cycling through all $p = 1, \cdots, n$ rows of T repetitively until $\delta(g) = 0$ for all n variables, the optimum of (3.3) is found. Obviously, in practice (3.10)–(3.15) are accurate only to the degree of calculational precision.

It can be shown that if certain elements of t are known to be zero, equations analogous to (3.4)–(3.9) are obtained. Specifically, if $t' = [t.', t..']$, where $t..' = 0$, and $T = [T., T..]$, $g = \sigma^2 t.'t. + 2\sigma_0' T.t. - \Delta(t.'t. - 1)$. In this case setting $\partial g/\partial t. = 0$ yields $t. = \beta T.'\sigma_0(\sigma_0' T.T.'\sigma_0)^{-1/2}$, equivalent to (3.9). The appropriate substitutions are, of course, made in the algorithm (3.10) and the convergence proof follows immediately. These considerations are relevant to any matrix T_i in (3.1) that is taken to be in partial lower triangular form for uniqueness of representation.

The optimization theory clarifies coefficients $\tilde{\rho}_i$ that are obtained as minimizing solutions to (3.2). Since under (3.2) $T_i'\Sigma T_i = T_i'\Delta T_i$ with $\text{diag}(T_i T_i') = I$, and since $\text{tr} T_i'\Sigma T_i = l'\Delta_i l$, the defining coefficient (2.5) can be written as

$$\tilde{\rho}_i = 1 - \frac{l'\Delta_i l}{l'\Sigma l}. \tag{3.16}$$

where we use the tilde notation to indicate a computational counterpart of (2.5) under the minimization of (3.1). If we were maximizing the function $\text{tr} T_r'\Sigma T_r$ under the constraint $\text{diag}(T_r T_r') = I$, analogously to (3.1), we would obtain

$$\tilde{\rho}_r = 1 - \frac{l'\Delta_r l}{l'\Sigma l}. \tag{3.17}$$

The difference between (3.16) and (3.17) stems from $\text{diag}(\Sigma - \Delta_i)$ versus $\text{diag}(\Sigma - \Delta_r)$, since $\text{diag}(\Sigma - \Delta_r) \leq 0 \leq \text{diag}(\Sigma - \Delta_i)$, as was shown above with each element $(\sigma^2 - \Delta)$. See Woodward and Bentler [1979] for other applications of the maximizing algorithm.

Coefficients (3.16) and (3.17) can be obtained as $i = 1, \cdots, k$. It is possible that local extrema can be encountered when i is small, so that several iterative sequences (3.10)

based on alternative initial T matrices might be entertained. At the global optimum to (3.3), coefficients $\tilde{\rho}_i$ and $\tilde{\rho}_r$ can be equated with coefficients ρ_i and ρ_r in (2.11) and (2.21).

The optimization theory developed above generates a matrix Δ_k that is a candidate for Ψ_i^2 in (1.5), thus possibly yielding ρ_i. The requirement for $\rho_i = \tilde{\rho}_k$ is that $(\Sigma - \Delta_k)$ be a minimal trace nonnegative definite matrix. We shall show that this condition is met. Consider the function defined by Bentler [1972], namely

$$\phi = \mathrm{tr}FF' - \mathrm{tr}M(\Sigma - FF' - U^2)\,, \tag{3.18}$$

which seeks a minimum of $\mathrm{tr}FF'$ with respect to F and a diagonal U^2, subject to the Lagrangian constraint $\Sigma = FF' + U^2$. The solution yields

$$(\Sigma - U^2)(I + M) = 0\,, \tag{3.19}$$

where M is symmetric with null diagonal and $(\Sigma - U^2) = FF'$. In view of the normalization of T_k, it is apparent that the matrix $T_kT_k'-I$ is symmetric with null diagonal as required for M, with $\Delta_k = U^2$, i.e., that (3.1) and (3.18), or (3.2) and (3.19), solve equivalent problems. Thus, not only are the diagonals of $(\Sigma - \Delta_k)$ positive at the solution (3.2), but $(\Sigma - \Delta_k)$ is positive semidefinite as well. As a consequence, by equating $\Psi_i^2 = U^2 = \Delta_k$ and $\Lambda_i = F$, we have $\tilde{\rho}_i = \tilde{\rho}_k$ with

$$\tilde{\rho}_i = 1 - \frac{I'U^2I}{I'\Sigma I} \tag{3.20}$$

at the minimum of (3.18). Obviously, (3.20) is a computational counterpart to the theoretical coefficient (1.5), as obtained by Bentler [1972], and $\tilde{\rho}_i = \rho_i$ at the global minimum to (3.18) or (3.1). An interesting feature of (3.20) is that $\tilde{\rho}_i$ is obtained as a minimizing solution to (3.18) while simultaneously being a maximized solution for ρ_k of (2.5) via the function (3.1) and a coefficient of the form (3.16). It is also useful to evaluate the matrix Δ_k. If it is nonnegative definite, we can take $\tilde{\rho}_i = \tilde{\rho}_+$, the computational greatest lower bound meeting the conditions for ρ_+ in (1.6). Thus, the theoretical development of this paper complements Bentler's previous work, and represents an alternative, more comprehensive, approach to obtaining the lower bound $\tilde{\rho}_i$. While Bentler appreciated the various rank solutions to (3.2), and developed an iterative routine to solve (3.2) and calculate (3.16), he did not study in detail their implications for reliability theory.

If the matrix of unique variances $\Psi_i^2 = \Delta_k$ of (3.20) is not nonnegative definite, $\tilde{\rho}_i$ will not be the greatest lower bound to reliability. As a consequence, it will be necessary to obtain a boundary solution to (3.1) in which one or more negative values ψ_p^2 are set to zero. (See Woodhouse and Jackson, 1977, for informative discussions of these boundary constraints.) The previously developed theoretical inequalities (2.21) verify that the resulting coefficient ρ_K will exceed ρ_k. This coefficient can be computed on the basis of two theoretical approaches, as will be described next.

Using notation previously developed, we define the function

$$f_K = \mathrm{tr}T_K'\Sigma T_K - \mathrm{tr}\Delta_{k1}[\mathrm{diag}(T_{k1}T_{k1}') - I]\,. \tag{3.21}$$

Note that (3.21) is equivalent to (3.1) if $T_K = T_k$, i.e., if Ψ^2 is not singular. In (3.21), however, we assume that Ψ_2^2 is null. Thus, $T_K' = [T_{k1}', T_{k2}']$, but only T_{k1} is constrained. Differentiating (3.21) with respect to T_{k1} and T_{k2} yields the partitioned matrix

$$\frac{\partial f_K}{\partial T_K} = 2\begin{bmatrix}(\Sigma_{11} - \Delta_{k1})T_{k1} + \Sigma_{12}T_{k2} \\ \Sigma_{21}T_{k1} + \Sigma_{22}T_{k2}\end{bmatrix}, \tag{3.22}$$

where Σ_{11}, Σ_{12}, Σ_{21} and Σ_{22} are the appropriate submatrices of Σ. Setting $\partial f_K/\partial T_{k2} = 0$ shows that T_{k2} can be expressed in terms of T_{k1} as

$$T_{k2} = -\Sigma_{22}^{-1}\Sigma_{21}T_{k1}\,. \tag{3.23}$$

Substituting (3.23) into $\partial f_K/\partial T_{k1}$, and equating this gradient to zero, yields

$$(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \Delta_{k1})T_{k1} = 0\,, \tag{3.24}$$

a form basically similar to (3.2). A solution to (3.24) can be obtained by the iterative process (3.10), under an obvious redefinition of terms. Instead of operating on the entire covariance matrix Σ, the iterations proceed on the partial covariance matrix $(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. Nonetheless, it can be shown that, as in (3.2), $(\Sigma - \Delta_K)T_K = 0$, where Δ_K is the supermatrix of Δ_{k1} and the null matrix Δ_{k2}.

It is possible to define a coefficient based on a solution to (3.24), but such a coefficient would not generally be useful unless the resulting individual elements of Δ_{k1} were also all nonnegative and all row sum of squares of T_{k2} were not less than one, as shown via (3.25) below. In order to achieve such a solution, steps (3.21)–(3.24) may have to be repeated, with Ψ_2^2 being redefined each time to include new negative elements ψ_p^2 and to exclude elements corresponding to rows of T_{k2} having sum of squares less than one. Finally, one will achieve a solution minimizing (3.21) and satisfying (3.24), with all nonnegative elements in Δ_{k1}.

A specialized feature of the optimization problem (3.3)–(3.4) can be utilized to provide an alternative computational algorithm for the greatest lower bound when negative values ψ_p^2 are encountered via algorithm (3.10). The theory (3.21)–(3.24) based on partial covariance matrices is not utilized in this approach. We describe the theory first, and then turn to the algorithm.

The solution for t given by (3.9) is associated with a Δ that can be expressed as $\Delta = \sigma^2 + \beta(\sigma_0'TT'\sigma_0)^{1/2}$, as substitution into (3.8) will show. These t and Δ solve (3.5), that is, $\partial g/\partial t = 0$. There exists, however, another minimizing solution to (3.5), namely

$$t = -\frac{T'\sigma_0}{\sigma^2}\,,\ (\beta = -1,\ \Delta = 0) \tag{3.25}$$

which is applicable only when minimizing (3.3). Substitution of (3.25) into (3.5) verifies that $\partial g/\partial t = 0$. Of course, (3.25) is based on a boundary solution for Δ, which is obtained without imposing the constraint $t't = 1$. While Δ associated with (3.9) can take on any arbitrary value, the interpretation of Δ as a variance as shown, for example, in (3.16), would require it to be nonnegative. If Δ associated with (3.9) is negative, (3.25) can be used in a minimization step instead, since it also can reduce the function (3.4). The resulting t must have $t't \geq 1$, since $\Delta = \sigma^2 - \sigma_0'TT'\sigma_0{}^{1/2} > 0$ if $t't < 1$ with (3.25); i.e., $t't < 1$ shows that a boundary solution is inappropriate.

A slight modification of algorithm (3.10), based on (3.25), is given in the minimization case by

$$\begin{aligned}
&\text{(a)}\ t^{(i+1)} = t^{(i)} && \text{if } \sigma_0'TT'\sigma_0 < \epsilon \\
&\text{(b)}\ t^{(i+1)} = -T'\sigma_0(\sigma_0'TT'\sigma_0)^{-1/2} && \text{if } \sigma^2 > (\sigma_0'TT'\sigma_0)^{1/2} \text{ and } g^{(i)} \geq g^{(i+1)} \\
&\text{(c)}\ t^{(i+1)} = -\frac{T'\sigma_0}{\sigma^2} && \text{otherwise.}
\end{aligned} \tag{3.26}$$

Alternatives (a) and (b) have been discussed in association with (3.10). Step (b) converges since it is taken only when the function decreases. Note that normalization of $t^{(i+1)}$ implies $\Delta^{(i+1)} > 0$, nonnormalization implies $\Delta^{(i+1)} = 0$, that the converse of both propositions is also true, and that it is not possible for $\Delta^{(i+1)}$ to be negative. Furthermore, (3.26) allows a previous boundary $\Delta^{(i)} = 0$ to leave the boundary via step (b). Convergence in case (c) can be shown as follows. The function (3.4) takes on the values

$$g^{(i)} = (\sigma^2 - \Delta^{(i)})t^{(i)\prime}t^{(i)} + 2\,\sigma_0'Tt^{(i)} + \Delta^{(i)}$$

(3.27)

$$g^{(i+1)} = -\frac{\sigma_0'TT'\sigma_0}{\sigma^2},$$

making no assumptions about normalization. Hence, the change in function is given by

$$g^{(i)} - g^{(i+1)} = \sigma^2(t^{(i)} - t^{(i+1)})'(t^{(i)} - t^{(i+1)}), \tag{3.28}$$

which is always nonnegative. More specifically,

$$\begin{aligned} g^{(i)} - g^{(i+1)} &> 0 \qquad \text{if } \Delta^{(i)} > 0 \\ g^{(i)} - g^{(i+1)} &\geq 0 \qquad \text{if } \Delta^{(i)} = 0. \end{aligned} \tag{3.29}$$

Apparently, the function decreases with certainty only if a previous nonboundary $\Delta^{(i)}$ becomes a boundary $\Delta^{(i+1)} = 0$.

While (3.26) represents an algorithm that monotonically reduces the function at every step, convergence may occur to a solution in which one or more boundary constraints are inappropriately imposed, i.e., one or more final $t't < 1$ with $\sigma^2 > \sigma_0'TT'\sigma_0^{1/2}$. As with the computations based on (3.21)-(3.24), algorithm (3.26) will then have to be restarted with any inappropriate boundary constraints removed; that is, any rows t' of T having $t't < 1$ should be normalized and (3.26) should be rerun.

Algorithm (3.26) will yield a matrix Δ_K with all nonnegative elements. According to the convexity and uniqueness arguments of Woodhouse and Jackson [1977] and Della Riccia and Shapiro [Note 2], this matrix will be the same as that obtainable via algorithm (3.10) and the partial covariance procedure (3.24). At this solution, one obtains the largest possible coefficient

$$\tilde{\rho}_K = 1 - \frac{1'\Delta_K 1}{1'\Sigma 1}, \tag{3.30}$$

which, by virtue of the previous theory (2.21) exceeds $\tilde{\rho}_k$. In order to complete the theory, we must also approach the coefficient (3.30) from above by seeking an appropriate absolute minimum to

$$\tilde{\rho}_+ = 1 - \frac{1'D^2 1}{1'\Sigma 1}, \tag{3.31}$$

where the diagonal matrix D^2 has some fixed zero elements. This is done by minimizing $\mathrm{tr}(\Sigma - D^2)$ while maintaining nonnegative definiteness of $(\Sigma - D^2)$ and D^2, via a minimization of the function

$$\phi_K = \mathrm{tr}BB' - \mathrm{tr}M(\Sigma - BB' - D^2), \tag{3.32}$$

while setting $D_2^2 = 0$. The solution, possibly after repeated redefinition of D^2, yields (3.30). That is, in consistency with (2.22), $\tilde{\rho}_K = \tilde{\rho}_+$, based on the equivalences $\Lambda_+ = B$ and $\Psi_+^2 = D^2 = \Delta_K$. Of course, at the global minimum to (3.31), $\tilde{\rho}_+ = \rho_+$.

We now collect the optimization results into a chain of inequalities. Upon using the result on upper and lower bounds, varying the index i appropriately, and recognizing the ability to obtain stronger results for singular Δ_k, we can generate computable coefficients entering the chain of inequalities

$$\tilde{\rho}_{K'} \leq \cdots \leq \tilde{\rho}_l \leq \tilde{\rho}_+ \leq \rho_1, \tag{3.33}$$

in a manner analogous to (2.24).

4. Sampling Considerations

It was assumed above that the population covariance matrix Σ was known, or at least well-estimated, but the problem of entity sampling needs to be mentioned. The sampling issue is closely related to the problem of parameter estimation for the basic model (1.1). We shall use the notation $\hat{\rho}_\alpha$, $\hat{\rho}_l$, and $\hat{\rho}_+$ to represent sample statistics for ρ_α, ρ_l, and ρ_+. A reason for considering the effects of entity sampling on lower-bound coefficients can be illustrated with coefficient α. While it is correct to say that α is a lower-bound to reliability, it is often assumed that the sample-computed coefficient $\hat{\alpha}$ is a lower-bound to population reliability. It is not. Woodward and Bentler [1978] present a statistical lower-bound version of $\hat{\alpha}$ that is a lower-bound in both the sample and the population. Unfortunately, no sampling theory for coefficients $\hat{\rho}_\alpha$, $\hat{\rho}_l$, or $\hat{\rho}_+$ is yet available.

In the absence of an available population covariance matrix, or a fixed or known covariance matrix to be studied, the question of independence of the variables should be routinely investigated. When the random vectors underlying (1.1) have a multivariate normal distribution, a test of the hypothesis that $\rho_l = 0$ is equivalent to a test of complete multivariate independence, which can be made using standard methods [Bentler, 1972]. If the hypothesis of independence cannot be rejected, it makes no sense to obtain lower-bound estimates of reliability. Even though $\hat{\alpha}$ or $\hat{\rho}_l$ may be nonzero in a given sample, their nonzero value would then be attributed to sampling variation. This observation is of some importance, for it demonstrates that the inequalities $\hat{\alpha} \leq \hat{\rho}_l \leq \hat{\rho}_+ \leq \rho_l$ relating sample statistics to population parameters cannot be guaranteed. One could, of course, use the sample covariance matrix S in place of Σ to obtain a chain of inequalities based on the data in a purely descriptive way, but such a result would have meaning only if there is no problem of statistical inference to a population and the sample covariance matrix is taken to be fixed.

The covariance matrix S represents an acceptable substitute for the population matrix in arbitrarily large samples. While the inequalities $\hat{\alpha} \leq \hat{\rho}_l \leq \hat{\rho}_+ \leq \rho_l$ hold asymptotically, they should become reasonably accurate as the sample size becomes large (in the absence of a mathematical result, we would propose that 2–300 entities might suffice). In such a case one can rely on the best possible lower-bounds to reliability, $\hat{\rho}_l$ or $\hat{\rho}_+$, rather than their alternatives; the lesser coefficients can be ignored. Furthermore, as the sample becomes very large it typically becomes difficult to fit any statistically acceptable covariance structure model of the form (1.1) to data; consequently, it is then not useful to obtain estimates $\hat{\Sigma}$ of Σ other than S for purposes of reliability estimation. Where $\hat{\Sigma}$ is obtained in small samples from confirmatory factor or covariance structure analysis, the previous methods cannot assure that $\hat{\rho}_+ \leq \rho_l$.

5. Computational Practice

The computational theory and two specific iterative methods for solving the inequalities (3.33) in any sample have been given in Section 3. In practice, of course, the sample covariance matrix S replaces the matrix Σ utilized in the algorithms (3.10) and (3.26). It is probably advisable to place the matrix T in echelon or partial lower triangular form, so that its elements are uniquely specified. Specifically, if the p^{th} row of t' of T has element $1, \cdots, q, \cdots, i$, then if $p < q$, $t_q = 0$. This condition is imposed during the iterations. Since (3.10) and (3.26) are row replacement algorithms, one iteration can be conceived of as consisting of a cycle of n row replacements, as the p^{th} row is replaced ($p = 1, \cdots, n$). Iterations are continued until the desired coefficient has stabilized to a given degree of accuracy. As an initial estimate of t' we use $t_q = -c$ if $p > q$, $+c$ if $p = q$, and 0 if $p < q$, where c is chosen to normalize t'. Experience verifies that the iterations typically converge quickly from this starting point.

A solution to (3.2) based on the iterations (3.10) produces matrices T_i and Δ_i that represent stationary points of the function f_i (3.1) in the sample. (To be strictly accurate, these optimizing matrices should have a "hat", which has been omitted in order to simplify notation.) Based on such a solution one can follow (3.16) to obtain $\hat{\rho}_i = 1 - 1'\Delta_i 1/1'S1$, which estimates ρ_i if T_i is consistent with the global minimum of the function f_i. If T_i is trapped at a local minimum, $\hat{\rho}_i$ will be a poor estimate of ρ_i. When $\hat{\rho}_1$ is of particular interest, local minima seem more likely because of the combinatorial nature of the problem; restarts from randomly selected initial T_1 are then highly desirable. If only the coefficient ρ_l is of interest, it is possible simply to increase the number of columns by one, and to repeat the procedure until $\hat{\alpha} \leq \hat{\rho}_i$ for some value of i, since it is known that $\alpha \leq \rho_l$. Thereafter, the main concern will be to determine indices j and k such that $\hat{\rho}_j = \hat{\rho}_k$, i.e., that the maximal rank r_k has been determined sufficiently accurately. In this case one estimates ρ_k and ρ_l by $\hat{\rho}_l = 1 - 1'\Delta_k 1/1'S1$. Since any obtained $\hat{\rho}_l$ should lie below ρ_l in large samples, one may be satisfied with the solution at this point; if not, one could try several randomly initialized T_k on the iterations (3.10) to evaluate potential local minimum problems. Similarly, one could evaluate whether $(S - \Delta_k)$ is Gramian. In our experience, these steps are not necessary.

In the applications section below, we give an example of a specific interest in the sample f_1 with its sign vector T_1. In this case the computational procedure (3.10) also provides a very rapid method for calculating $\hat{\rho}_1$ as an estimate of maximum λ_4. (Previous writers have not proposed an efficient method for actually obtaining max $\hat{\lambda}_4$ from the class of λ_4 coefficients; the problem may not have a better algorithmic solution due to its apparent NP-completeness character. See Garey & Johnson, 1979.) Often, however, only ρ_l and ρ_+ are of interest. Consequently, computations can typically begin with T_i of column dimensionality no greater than $n/3$, as suggested by Bentler. In view of (3.2), the index i must mirror the degree of rank reduction in $(S - \Delta_i)$; there is no harm in overestimating this value since excess columns simply become null during the iterations. In fact, a useful check on the calculation lies in this nullity. Such a check is particularly important when working with matrices whose rank can be extensively reduced by appropriate communality estimates. A large value for i would be necessary when $S \equiv \hat{\Lambda}\hat{\Lambda}' + \Psi^2$, for example, as obtained from a factor analytic procedure; it appears then to be generally true that $\Psi^2 = \Delta_k$ [Della Riccia & Shapiro, Note 2].

Once $\hat{\rho}_l$ obtained via (3.10) is accepted as a reasonable estimate of ρ_l, the possibility of obtaining $\hat{\rho}_+ \geq \hat{\rho}_l$ must be evaluated. This can be done by investigating the values Δ_k to determine if they are all nonnegative. If so, $\hat{\rho}_+ = \hat{\rho}_l$. If not, two alternative procedures can be utilized. First, one can identify the variables having positive entries in Δ_k with those of Ψ_1^2 in (3.21)–(3.24), and the other variables with those of $\Psi_2^2 = 0$. If desired, this can be done by permuting S. Assume that S has been permuted such that S_{11} is consistent with Ψ_1^2. Then, according to (3.24), one calculates $(S_{11} - S_{12}S_{22}^{-1}S_{21})$ and uses this matrix in place of S in the iterations (3.10). The column space of T_{k1} is known as k. This partial covariance procedure may have to be repeated after reassigning resulting negative Δ_{k1} variables to Ψ_2^2, and variables having row sum of squares in T_{k2} less than unity to Ψ_1^2. The computations must ultimately yield the supermatrix estimate Δ_K composed of Δ_{k1} and $\Delta_{k2} = 0$, yielding $\hat{\rho}_+ = 1 - 1'\Delta_K 1/1'S1$. If $\hat{\rho}_+ \geq \hat{\rho}_l$, $\hat{\rho}_+$ may be taken as an estimate of ρ_+. Instead of using the partial covariance method, the second, simpler computational procedure based on (3.26) can be used to obtain $\hat{\rho}_+$ directly. Of course, these two methods should yield identical results.

The procedures outlined here will yield the sequence of inequalities $\hat{\rho}_1 \leq \hat{\rho}_l \leq \hat{\rho}_+$ for all computed estimates meeting the optimization criteria, as well as $\hat{\alpha} \leq \hat{\rho}_l \leq \hat{\rho}_+$. When the sample size is large one can utilize $\hat{\rho}_+$ as the optimal internal consistency estimator, since

it is then possible to rely upon the inequality $\hat{\rho}_+ \leq \rho_I$, which holds asymptotically. In this case, $\hat{\rho}_1 = \tilde{\rho}_1$, $\hat{\rho}_I = \tilde{\rho}_I$, and $\hat{\rho}_+ = \tilde{\rho}_+$.

6. *Examples*

The first example is taken from Lord and Novick [1968, p. 91], and represents the matrix of covariances among four measures of English as a foreign language. The covariance matrix, based upon a sample size of 1416, is presented in the left part of Table 1. Coefficient $\hat{\alpha}$ for this matrix was computed to be .891. The computational routine (3.10) leading to $\hat{\rho}_1$ yielded as the final T_1 vector the sign vector presented in the middle of Table 1. Coefficient $\hat{\rho}_1 = .898$, based upon the vector of diagonal elements Δ_1 presented in column form (i.e., $\Delta_1 I$) in the table. Ten random restarts were taken, and 4 of these converged to the correct value of $\hat{\rho}_1$ after the minimum number (three) of specified iterations. Stepping up the column space from T_1 to T_2 led to the solution vectors presented in the right part of Table 1. Further increment of the column space from T_2 to T_3 led to no further changes, so that $\hat{\rho}_k = \hat{\rho}_2 = \hat{\rho}_I = .920$, as reported previously by Bentler [1972]. It can be verified that $(S - \Delta_2)$ is positive semidefinite of rank two as it should be, and, since the uniquenesses are all positive, $\hat{\rho}_I = \hat{\rho}_+$, the greatest lower bound. The improvement possible by lower-bounding reliability via $\hat{\rho}_I$ rather than $\hat{\alpha}$ was quite small in this example. In many situations, the improvement is quite substantial [e.g., Jackson, 1977].

The second example provides an illustration of a Heywood case, i.e., negative uniquenesses under (3.10). Consequently, the previous theory would predict that $\hat{\rho}_I$ is not the greatest possible lower bound. The data from this example represent the intercorrelations among indicators of social class [Warner, Meeker, & Eels, 1960, p. 1972]. As can be anticipated from the high intercorrelations, coefficients $\hat{\alpha}$ and $\hat{\rho}_I$ do not differ by

TABLE 1

ETS Test of English as a Foreign Language

	Covariance Matrix				T_1	$\Delta_1 1$	T_2		$\Delta_2 1$
	1	2	3	4			1	2	
1	94.7				-1	12.9	1.000	0	47.58
2	87.3	212.0			1	114.2	-.996	.092	36.99
3	63.9	138.7	160.5		-1	-24.1	.262	-.965	42.77
4	58.4	128.2	109.8	115.4	1	75.4	.395	.919	13.28

$\hat{\alpha} = .891$, $\hat{\rho}_1 = .898$, $\hat{\rho}_2 = \hat{\rho}_\ell = \hat{\rho}_+ = .920$

much. In this example $\hat{\rho}_1 = \hat{\rho}_l$, i.e., its computation is based upon a vector T_1. Increasing the column space from T_1 to T_2 has, to the chosen degree of computational accuracy, no effect upon the internal consistency. The vector T_1 that produces $\hat{\rho}_l$ is given in the middle of Table 1. As reported by Bentler [1972, p. 351], Variable 2 has a negative uniqueness in this solution. Consequently, the partialing procedure of (3.24) may be utilized. Since T_1 consists of a single column, so does the orthogonal complement in (3.24); we have denoted it here as $T_{11(1)}$, to indicate the final values of T_{11} after partialing. The excluded variable (two), of course, does not have a sign element in the solution $T_{11(1)}$. At the solution of (3.24), Variable 6 had a negative uniqueness estimate. Consequently, the partialing procedure was reapplied, with Variables 2 and 6 being eliminated, and the vector $T_{11(2)}$ was obtained at the solution. The solution again yielded a Heywood variable, necessitating reapplication of the partialing procedure. The final solution vector $T_{11(3)}$, and the corresponding final uniquenesses $\Delta_{11(3)}$, are presented in the right part of Table 2. As an alternative to the repeated use of (3.24), (3.26) may be used. As expected, this algorithm obtained the same $\Delta_{11(3)}$ reported in Table 2. We find $\hat{\rho}_l < \hat{\rho}_+$ so that $\hat{\rho}_l$ cannot be taken as the greatest lower bound. As pointed out by Bentler [1972, p. 351], this example is quite unusual, so inferences to a population should not be drawn.

7. *Applications*

The theory of internal consistency developed here can be utilized in areas other than reliability estimation. We shall describe two applications: test construction and factor analysis.

A problem of some practical importance in the development of tests involves the construction of alternate forms. For example, given a pool of 100 items that adequately repre-

TABLE 2

Warner, Meeker, and Eels' [1960] Social Class Data

	Correlation Matrix						Final Solution Vectors				
	1	2	3	4	5	6	T_1	$T_{11(1)}$	$T_{11(2)}$	$T_{11(3)}$	$\Delta_{11(3)}$[1]
1	1.00						1	1	1	-	0
2	.87	1.00					-1	-	-	-	0
3	.76	.82	1.00				-1	1	1	1	.250
4	.71	.81	.71	1.00			1	1	1	1	.188
5	.70	.81	.69	.74	1.00		1	1	1	1	.222
6	.77	.59	.64	.70	.65	1.00	-1	-1	-	-	0

$\hat{\alpha} = .942$, $\hat{\rho}_1 = \hat{\rho}_\ell = .969$, $\hat{\rho}_+ = .976$

sent a construct being assessed, it may be desirable to assign each item to one of two alternate forms according to some criterion of optimal similarity. These alternate forms could then be used interchangeably.

We shall approach the problem of optimal assignment of items via reliability theory. If there are two alternate forms to be constructed, we propose to construct these forms such that a lower bound to reliability of the forms is maximized. We use the word "reliability" for simplification in place of "lower bound to reliability." This problem can be analyzed in terms of the typical relations between component and composite reliability: each alternate form, considered as a sum across a set of items, is a component, and the composite consists of the sum of the two component forms, i.e., it remains the original composite based on the pool of all items. The problem of maximizing component reliability can be approached via a step-down of composite reliability. Since there are many lower bounds to composite reliability, we shall use the one that can be expressed as a function of split-half component reliability, namely, ρ_1. The reliability of the component forms can be expressed as $\rho_{11} = \rho_1/(2 - \rho_1)$, which expresses ρ_{11} as the intraclass coefficient that represents the ratio of component covariance to the average of the components' variance. Thus, ρ_1 is a simple Spearman-Brown function of ρ_{11}; of course, when the two forms have equal variance, ρ_{11} is simply the correlation between the two forms. As such, ρ_{11} describes the reliability of a "half-length" composite. Since $\rho_1 \leq \rho_+$, $\rho_{11} \leq \rho_{11+}$, where $\rho_{11+} = \rho_+/(2 - \rho_+)$. It follows that partitioning the item set to maximize ρ_{11} is equivalent to maximizing ρ_1. This problem has been solved previously, both theoretically and computationally. The sample sign vector T_1 that maximizes $\hat{\rho}_1$ provides the information needed to classify the items; an item is assigned into one or the other form depending on whether its associated sign in T_1 is positive or negative.

The theory of optimal assignment under reliability theory can be illustrated with Lord's [1956] 15 intellectual measures, as summarized by Jöreskog [1974, p. 13]. Each variable can be classified according to its substantive content (three types: vocabulary, intersections, and arithmetical reasoning), as well as to the conditions of testing (two types: low speeded and speeded). If these variables were to be assigned to two sets designed to be maximally similar, one would expect that each set should contain similar content/condition variables. Specifically, each set should contain at least one low and one highly speeded variable in each content domain. In this way, the content/condition combinations would be distributed across the two forms. The computational method (3.10) associated with obtaining coefficient $\hat{\rho}_1$ was applied to the data reported by Jöreskog. Thirty random initial T_1 vectors were used, and the best solution was retained. This solution presumably yielded the global minimum of the function $T_1'ST_1$, as well as the associated maximum $\hat{\rho}_1$. We obtained $\hat{\rho}_1 = .944$, so that $\hat{\rho}_{11} = .894$. The solution classified the variables according to content and condition combinations as expected. Further, when the 15 variables were reduced to 12 so that there were two lowly and highly speeded tests per content domain, the procedure perfectly split the variables into sets of six, with each content area having one test from each of the speeding conditions. With some data, of course, there may be several splits that yield almost the same index of similarity for the alternate forms.

As a second area of application of the theory developed in this paper, we consider the problem of exploratory factor analysis where (a) there are no interesting statistical hypotheses to test, or (b) the data do not meet the basic assumptions of the relevant statistical procedures, or (c) the data set is so large that statistical factor analytic procedures become too costly.

Harman [1976, Ch. 5–6] pointed out that methods of factor analysis can be classified according to whether they assume a given dimensionality of the common space, and then

estimate the parameters of the model under that assumption, or whether they obtain communalities, and then determine the significant sources of variance in the common-factor space. The more elegant methods of factor analysis use the procedure of conditional estimation, based on an assumed rank of A in (1.1). If the assumption is correct, such procedures produce better estimates of the parameters in (1.1) than is possible with communality-based methods because they use valuable additional information. The only way to evaluate the rank assumption is to examine the fit of the model (e.g., by chi-square), and to reestimate the parameters under another assumed rank if the fit is not adequate. Obviously, this procedure may have to be repeated many times to yield an acceptable solution (which is expensive). There are advantages to estimating communality, and then factoring the correlation matrix with communalities. Once the communalities are known, a single principal components analysis of the uniqueness-reduced correlation matrix reveals the major and minor sources of variance. Since all solutions for A of a given rank r_i are subsets of solutions at rank $r_j (i < j)$, it is inexpensive to vary the dimensionality of the solution. Of course, a decision regarding the number of factors to retain for rotation must still be made. The reliability theory developed above represents a specific theoretical solution to the communality problem involving communalities whose sum is minimal, while assuring that the reduced correlation matrix and the matrix of uniquenesses are both Gramian to within computational accuracy. This approach would be particularly relevant to the analysis of sets of congeneric tests (cf., Jöreskog, 1974), in which the uniquenesses are assumed to represent error variances, since minimizing the sum of communalities maximizes the sum of error variances.

We have analyzed the famous five socioeconomic and eight physical variables [Harman, 1976, p. 14, 22] to illustrate the factor-analytic application. The socioeconomic variables yielded a solution for $\hat{\rho}_l$ involving two columns in T_2. At this solution, Variables 1 and 3 were Heywood cases. Consequently, the methods associated with $\hat{\rho}_+$ were utilized. The final communalities of 1.0, .831, 1.0, .795, and .982 reduced the correlation matrix to rank 3 with eigenvalues of 2.777, 1.764, .068, .0000, and .0000. The first two unrotated factors were virtually the same as Harman's minres and maximum likelihood solutions. (Minres also has a Heywood variable.) The $\hat{\rho}_l$-based factor solution for the eight physical variables yielded an orthogonal complement T_2, and consequently the rank of the reduced correlation matrix was six. The communalities of .936, .978, .879, .871, .907, .705, .664, and .555 yielded a reduced correlation matrix with eigenvalues of 4.517, 1.572, .175, .129, .075, .027, and two eigenvalues of zero to five decimal places. The solution for the first three factors was again very similar to Harman's minres and maximum likelihood results. However, there was a tendency for the equivalent dimensions to be associated with marginally higher communalities when based upon the reliability methods of the present paper. It should also be noted that Harman reported three maximum-likelihood solutions. While all three are virtually identical, two of these solutions found Variable 2 to be improper. Nonetheless, the methods of the current paper did not encounter a negative uniqueness.

8. Discussion

In this paper we have developed a new, reasonably general theoretical approach to developing lower bounds for reliability. The approach seems to be quite powerful, yielding a fairly extensive chain of inequalities (2.24) including such previously-defined coefficients as Guttman [1945] λ_4, Bentler's [1972] ρ_l, and the greatest lower bound ρ_+ obtained here and independently by Woodhouse and Jackson [1977]. Since the lower coefficients in (2.24) will generally be negative, attention will typically focus only on the positive coefficients. A condition necessary to obtaining the greatest lower bound was developed. The major failure of the chain of inequalities appears to be the inability to include Cronbach's

[1951] popular α into the chain under general conditions (because α is not generally a mean of split-half coefficients), and the inability to include all important known lower bounds into the chain. For example, Guttman [1945] and Jackson and Agunwamba [1977] have presented various lower-bound coefficients that appear to be difficult if not impossible to include into the inequalities (2.24). An explication of the inequalities via optimization theory proved to be useful, and a computational chain (3.33) of inequalities was developed. The optimization theory verified that the greatest lower bound can be attained via two algorithms with proven convergence properties. One computational technique that implements the theory is based on Bentler's [1972] method, but our modification allows one to obtain estimates of several coefficients in the chain of inequalities. In addition, the computational methods enable one to avoid negative uniquenesses, which is an important consideration primarily in association with the greatest lower bound.

A different approach to obtaining only the greatest lower bound was recently developed independently by Woodhouse and Jackson [1977]. They proposed a search procedure that involves the repeated calculation of determinants and cycling through steps designed to stay within the admissible region. This approach generally appears to yield the same greatest lower bound that we obtain, but further research comparing the computational methods is clearly needed. In regards to the improvement in lower-bound estimation possible via $\hat{\rho}_+$ and its competitors, Woodhouse and Jackson report that $\hat{\rho}_+$ is often only marginally greater than $\hat{\rho}_1$, while Jackson and Agunwamba [1977] indicate that $\hat{\rho}_l = \hat{\rho}_+$ in most practical applications. Our own experience agrees with the former observation, though we tend to find $\hat{\rho}_l < \hat{\rho}_+$ by a small amount. Systematic Monte Carlo work is needed to evaluate the generality of these observations.

The entire development of the current paper hinges upon the basic model (1.1) considered as an exploratory factor analysis model, with the factor loading matrix A being treated as an exploratory, unstructured matrix, and with specific and error variables being well-behaved (e.g., error variables not being correlated). If the model cannot be assumed, the results do not follow; Rozeboom [Note 3], among others, has questioned the general applicability of the model. A development of lower-bound coefficients could proceed in the same way when the factor loading matrix has a specific confirmatory factor structure, but such a structure would have to be determined by methods other than those of the present paper. Tying lower-bounds to reliability to a given structure may be quite useful, particularly where statistical estimation in small samples is of concern, since the current theory and methods are ideally applied only in very large samples. They are exact only in the population, or in a situation with a fixed or known covariance matrix. Clearly, the statistical theory associated with estimating the bounds (2.24) under various sampling conditions represents an important focus of future research.

Although the improvement in estimation made possible by $\hat{\rho}_+$ over Bentler's [1972] $\hat{\rho}_l$ is not necessarily large in practice, the main implication of the current results for applied psychometrics is that the greatest lower bound $\hat{\rho}_+$ should replace its predecessors [i.e., Bentler, 1972; Cronbach, 1951; Guttman, 1945; Jackson & Agunwamba, 1977; Lord & Novick, 1968] as the preferred estimator of internal consistency in large samples. However, the psychometric theory developed here also appears to have implications for psychometric practice outside the narrow area of reliability estimation. Only two applications were proposed, namely, assignment of variables to test forms so as to maximize an index of similarity of the forms, and estimation of communalities for a dimension-free approach to exploratory factor analysis; of course, the pragmatic value of these and other applications remains to be determined. Nonetheless, theoretical work related to minimum trace factor analysis [Bentler, 1972] is continuing as shown by the studies of Della Riccia and Shapiro [Note 2], and Bentler [Note 4]. The Della Riccia-Shapiro work is more general

than the approach considered here though its computational implementation is more complicated; Bentler [Note 4] developed an inexpensive method related to the current approach that is scale free and avoids zero as well as negative unique variances. These methods, however, are not necessarily tied to the concept of lower bounds to reliability, which is the focus of the current report.

REFERENCE NOTES

1. Bentler, P. M. & Woodward, J. A. *Inequalities among lower bounds to reliability.* Paper presented at the annual meeting of the Psychometric Society, Chapel Hill, June, 1977.
2. Della Riccia, G. & Shapiro, L. *Minimum rank and minimum trace of correlation matrices.* Department of Mathematics, Ben Gurion University of the Negev, Beer-Sheva, Israel, 1979.
3. Rozeboom, W. W. Personal communication, March, 1979.
4. Bentler, P. M. *Fast, legitimate, rank-independent factor analytic procedures.* Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Fort Worth, October, 1972.

REFERENCES

Bentler, P. M. A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research,* 1972, *1,* 343–357.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika,* 1951, *16,* 297–334.

Garey, M. R. & Johnson, D. S. *Computers and intractability: A guide to the theory of NP-completeness.* San Francisco: Freeman, 1979.

Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika,* 1945, *10,* 255–282.

Harman, H. *Modern factor analysis.* Chicago: University of Chicago, 1976.

Jackson, D. N. Reliability of the Jackson Personality Inventory. *Psychological Reports,* 1977, *40,* 613–614.

Jackson, P. H. A note on the relation between coefficient alpha and Guttman's "split-half" lower bounds. *Psychometrika,* 1979, *44,* 251–252.

Jackson, P. H. & Agunwamba, C. C. Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika,* 1977, *42,* 567–578.

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. II.* San Francisco: Freeman, 1974, pp. 1–56.

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

Lord, F. M. & Novick, M. R. *Statistical theories of mental test scores.* Reading, Mass.: Addison-Wesley, 1968.

Warner, W. L., Meeker, M. & Eels, K. *Social class in America.* New York: Harper & Row, 1960.

Woodhouse, B. & Jackson, P. H. Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika,* 1977, *42,* 579–591.

Woodward, J. A. & Bentler, P. M. A statistical lower bound to population reliability. *Psychological Bulletin,* 1978, *85,* 1323–1326.

Woodward, J. A. & Bentler, P. M. Application of optimal sign vectors to reliability and cluster analysis. *Psychometrika,* 1979, *44,* 337–341.

Manuscript received 8/10/77
First revision received 12/1/77
Second revision received 7/27/78
Final version received 2/5/80