PSYCHOMETRIKA—VOL. 65, NO. 3, 413–425
SEPTEMBER 2000

THE ASYMPTOTIC BIAS OF MINIMUM TRACE FACTOR ANALYSIS, WITH APPLICATIONS TO THE GREATEST LOWER BOUND TO RELIABILITY

ALEXANDER SHAPIRO

SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING
GEORGIA INSTITUTE OF TECHNOLOGY

JOS M.F. TEN BERGE

HEYMANS INSTITUTE OF PSYCHOLOGICAL RESEARCH
UNIVERSITY OF GRONINGEN

In theory, the greatest lower bound (g.l.b.) to reliability is the best possible lower bound to the reliability based on single test administration. Yet the practical use of the g.l.b. has been severely hindered by sampling bias problems. It is well known that the g.l.b. based on small samples (even a sample of one thousand subjects is not generally enough) may severely overestimate the population value, and statistical treatment of the bias has been badly missing. The only results obtained so far are concerned with the asymptotic variance of the g.l.b. and of its numerator (the maximum possible error variance of a test), based on first order derivatives and the asumption of multivariate normality. The present paper extends these results by offering explicit expressions for the second order derivatives. This yields a closed form expression for the asymptotic bias of both the g.l.b. and its numerator, under the assumptions that the rank of the reduced covariance matrix is at or above the Ledermann bound, and that the nonnegativity constraints on the diagonal elements of the matrix of unique variances are inactive. It is also shown that, when the reduced rank is at its highest possible value (i.e., the number of variables minus one), the numerator of the g.l.b. is asymptotically unbiased, and the asymptotic bias of the g.l.b. is negative. The latter results are contrary to common belief, but apply only to cases where the number of variables is small. The asymptotic results are illustrated by numerical examples.

Key words: asymptotic bias, asymptotic normality, reliability, minimum trace factor analysis, large sample theory, semidefinite programming

1. Introduction

The problem of how to assess the reliability of a test from a single test administration has been a key issue in classical test theory. When strong assumptions about the equivalence of the items, like congeneric or essential tau-equivalence measurements, seem unwarranted, and parallel tests are not available, one has to resort to lower bounds to the reliability. Guttman (1945) proposed 6 lower bounds, one of which, λ_3, has become famous as coefficient alpha (Cronbach, 1951). Jackson and Agunwamba (1977) and ten Berge and Zegers (1978) introduced additional lower bounds. More importantly, Jackson and Agunwamba also demonstrated that there exists a greatest lower bound (g.l.b.) to the reliability of a test.

The g.l.b. can only be evaluated by an iterative procedure. A first numerical solution was offered by Woodhouse and Jackson (1977), and more efficient procedures were soon to follow (Bentler & Woodward, 1980; also see ten Berge, Snijders & Zegers, 1981). More recently, interior point algorithms were designed for such type of problems, which became known as semidefinite programming problems (see, e.g., a survey paper of Vandenberghe & Boyd, 1996). An interior points algorithm can be downloaded from the internet, see Alizadeh, Haeberly, Nayakkankuppam, Overton & Schmieta (1997).

Jackson and Agunwamba (1977) observed that the greatest lower bound to the reliability is

This research was supported by grant DMI-9713878 from the National Science Foundation. Requests for reprints should be sent to Alexander Shapiro, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205. E-Mail: ashapiro@isye.gatech.edu

0033-3123/2000-3/1999-0712-A $00.75/0

the worst possible value of the ratio of true to observed variance, for any given covariance matrix, which is still compatible with the two standard assumptions of classical test theory in the context of lower bounds to reliability: zero covariance between the error parts of the items of the test, and zero covariance between the errors and the true scores. The g.l.b. has a direct link to Constrained Minimum Trace Factor Analysis (CMTFA, see Bentler & Woodward, 1980; Shapiro, 1982; ten Berge et al., 1981).

Specifically, when $\mathbf{\Sigma}$ is a $p \times p$ population covariance matrix, CMTFA is concerned with finding the diagonal matrix $\mathbf{\Psi} = \text{diag}(\Psi_{11}, \ldots, \Psi_{pp})$, of unique variances, which solves the following optimization problem:

$$\underset{\mathbf{\Psi}}{\text{Min}}\ \text{tr}(\mathbf{\Sigma} - \mathbf{\Psi}) \text{ subject to } \mathbf{\Sigma} - \mathbf{\Psi} \succeq 0,\ \mathbf{\Psi} \succeq 0. \tag{1}$$

The notation $\mathbf{\Sigma} - \mathbf{\Psi} \succeq 0$ means that the matrix $\mathbf{\Sigma} - \mathbf{\Psi}$ is positive semidefinite. When the constraint $\mathbf{\Psi} \succeq 0$ is ignored, we are dealing with Minimum Trace Factor Analysis (MTFA), first considered by Bentler (1972) under the name "Minimum Variance FactorAnalysis." The trace that is minimized in MTFA and CMTFA is the sum of the communalities, in the context of common factor analysis, and it is the sum of the true score variances, in the context of reliability theory.

Both MTFA and CMTFA imply a lower bound to the reliability. Specifically, the greatest lower bound can be written as

$$\rho := 1 - \frac{\text{tr}\,\bar{\mathbf{\Psi}}}{\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p}, \tag{2}$$

where $\mathbf{1}_p$ denotes the $p \times 1$ vector of ones, and $\bar{\mathbf{\Psi}} = \bar{\mathbf{\Psi}}(\mathbf{\Sigma})$ is the optimal solution of the problem (1). (The notation " := " means "equal by definition.") Note that if all diagonal elements of the matrix $\bar{\mathbf{\Psi}}$ are nonzero, and hence are positive (i.e., the constraints $\Psi_{ii} \geq 0, i = 1, \ldots, p$, are not active at the optimal solution), then MTFA and CMTFA have the same optimal solutions, and hence the same optimal values.

The g.l.b. would have gained tremendous popularity if it were not for the problem of sampling bias. For instance, Cronbach (1988) has pointed to the serious positive sampling bias when the g.l.b. is estimated in samples of less than several thousands of subjects. In the present paper, the size and the variance of the bias in MTFA (CMTFA) will be approximated by asymptotic arguments. Although this does not provide us with a directly useful unbiased estimate for the g.l.b., it may well represent a major step ahead in dealing with the sampling bias problem of the g.l.b.

In practice, the population covariance matrix $\mathbf{\Sigma}$ is unknown, and instead, we shall have to resort to a *sample* covariance matrix $\mathbf{S}$. Clearly, MTFA and CMTFA for samples can be defined analogously to those of the population, and the same goes for the statistics. That is, the population value ρ of the g.l.b. can be estimated by its sample counterpart

$$\hat{\rho} := 1 - \frac{\text{tr}\,\bar{\mathbf{\Psi}}(\mathbf{S})}{\mathbf{1}_p'\mathbf{S}\mathbf{1}_p}, \tag{3}$$

where $\bar{\mathbf{\Psi}}(\mathbf{S})$ is the optimal solution of CMTFA with $\mathbf{\Sigma}$ replaced by $\mathbf{S}$. We study in this paper sampling properties of the estimator $\hat{\rho}$. In particular, we discuss the bias of $\hat{\rho}$.

2. Background Material

Consider the following optimization problem

$$\underset{\mathbf{X}}{\text{Min}}\ \text{tr}\,(\mathbf{Z} - \mathbf{X}) \text{ subject to } \mathbf{Z} - \mathbf{X} \succeq 0,\ \mathbf{X} \succeq 0, \tag{4}$$

where $\mathbf{X}$ is a $p \times p$ diagonal matrix and $\mathbf{Z}$ is a given $p \times p$ symmetric matrix. Clearly the above problem is equivalent to the CMTFA problem (1) with $\mathbf{X} = \boldsymbol{\Psi}$ and $\mathbf{Z} := \boldsymbol{\Sigma}$. The problem (4) can be also written in the following equivalent form

$$\underset{\mathbf{X}}{\text{Min}} \ \text{tr}\,(-\mathbf{X}) \text{ subject to } \mathbf{Z} - \mathbf{X} \succeq 0, \ \mathbf{X} \succeq 0. \tag{5}$$

It is known (Della Riccia & Shapiro, 1982, Theorem 4; ten Berge et al., 1981, Lemma 2.4) that, for any $\mathbf{Z}$, problem (4), and its equivalent (5), has a unique optimal solution, denoted subsequently by $\bar{\mathbf{X}}$ or $\bar{\mathbf{X}}(\mathbf{Z})$. We also consider the optimal value $v(\mathbf{Z}) := -\,\text{tr}\,\bar{\mathbf{X}}(\mathbf{Z})$ of the problem (5) and

$$g(\mathbf{Z}) := 1 + \frac{v(\mathbf{Z})}{\mathbf{1}_p' \mathbf{Z} \mathbf{1}_p}, \tag{6}$$

viewed as functions of the matrix $\mathbf{Z}$. Clearly the population g.l.b. ρ is equal to $g(\boldsymbol{\Sigma})$, and its sample estimate $\hat{\rho} = g(\mathbf{S})$. Therefore it should be not surprising that the sampling behavior of $\hat{\rho}$ is closely related to analytical properties of these functions $v(\cdot)$ and $g(\cdot)$.

It has been shown (Shapiro, 1982) that the function $v(\cdot)$ is continuous, on the space of $p \times p$ symmetric matrices, and hence $g(\cdot)$ is continuous on the set of positive definite matrices. Since the sample covariance matrix $\mathbf{S}$ is a consistent estimator of $\boldsymbol{\Sigma}$, it follows that $\hat{\rho}$ is a consistent estimator of ρ.

By $\mathbf{s} = \text{vec}\,(\mathbf{S})$ we denote the $p^2 \times 1$ vector obtained by stacking columns of $\mathbf{S}$, and similarly $\boldsymbol{\sigma} = \text{vec}\,(\boldsymbol{\Sigma})$. We have that $\mathbf{S}$ is an unbiased estimator of $\boldsymbol{\Sigma}$. Moreover, if the population distribution has fourth order moments, then by the Central Limit Theorem, $n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to multivariate normal $N(0, \boldsymbol{\Gamma})$. Note that since $\mathbf{s}$ has at most $p(p+1)/2$ distinct elements, the rank of the $p^2 \times p^2$ covariance matrix $\boldsymbol{\Gamma}$ is less than or equal to $p(p+1)/2$. In particular, if the population distribution is normal, then $\boldsymbol{\Gamma} = 2\mathbf{M}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$, where $\mathbf{M}_p$ is a symmetric idempotent matrix of rank $p(p+1)/2$ with element in row ij and column kl given by

$$M_p(ij, kl) = \tfrac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{7}$$

where $\delta_{ik} = 1$ if $i = k$, and $\delta_{ik} = 0$ if $i \neq k$ (Browne, 1974).

Let us discuss now the structure of the problem (5). Recently such type of problems became a subject of intensive investigation in optimization theory under the name semidefinite programming problems (see, e.g., Vandenberghe & Boyd, 1996). Let us summarize some basic facts about problem (5) (see Shapiro, 1982).

The (Lagrangian) dual of (5), for $\mathbf{Z} = \boldsymbol{\Sigma}$, can be written in the form

$$\underset{\boldsymbol{\Omega}}{\text{max}} \ -\,\text{tr}\ \boldsymbol{\Omega}\boldsymbol{\Sigma} \text{ subject to } \text{diag}(\boldsymbol{\Omega}) \succeq \mathbf{I}_p, \ \boldsymbol{\Omega} \succeq 0. \tag{8}$$

The dual problem (8) always has an optimal solution (which may not be unique), and the optimal values of the problems (5) and (8) are finite and equal to each other. Moreover, a $p \times p$ symmetric matrix $\bar{\boldsymbol{\Omega}}$ is an optimal solution of the dual problem (8) if and only if it is feasible, that is, $\text{diag}(\bar{\boldsymbol{\Omega}}) \succeq \mathbf{I}_p$ and $\bar{\boldsymbol{\Omega}} \succeq 0$, and the following complementarity condition holds

$$\bar{\boldsymbol{\Omega}}(\boldsymbol{\Sigma} - \bar{\mathbf{X}}) = 0. \tag{9}$$

We denote by r the rank of the matrix $\boldsymbol{\Sigma} - \bar{\mathbf{X}} (= \boldsymbol{\Sigma} - \bar{\boldsymbol{\Psi}})$, and refer to it as the *reduced* rank of $\boldsymbol{\Sigma}$. It is said that the optimal solution $\bar{\mathbf{X}}$ is *nondegenerate* if vectors $\mathbf{f}_i * \mathbf{f}_j$, $1 \leq i \leq j \leq p - r$, are linearly independent. Here $\mathbf{f}_1, \ldots, \mathbf{f}_{p-r}$ are vectors forming a basis of the null space of the matrix $\boldsymbol{\Sigma} - \bar{\mathbf{X}}$, and $\mathbf{f}_i * \mathbf{f}_j$ denotes the Hadamard (term by term) product. This nondegeneracy condition is a specification of a so-called transversality condition to the considered problem. Note that the nondegeneracy condition can only hold if $(p - r)(p - r + 1)/2 \leq p$. In a certain sense this

nondegeneracy condition is generic (see Shapiro, 1982, 1997). Note also that it follows from (9) that

$$\operatorname{rank}(\bar{\mathbf{\Omega}}) + \operatorname{rank}(\mathbf{\Sigma} - \bar{\mathbf{X}}) \leq p.$$

It is said that the *strict complementarity* condition holds if

$$\operatorname{rank}(\bar{\mathbf{\Omega}}) + \operatorname{rank}(\mathbf{\Sigma} - \bar{\mathbf{X}}) = p. \tag{10}$$

If the nondegeneracy condition holds, then the dual problem (8) has a unique optimal solution $\bar{\mathbf{\Omega}}$. Moreover, if the strict complementarity condition holds and $\bar{\mathbf{X}}$ does not have zero diagonal elements, then $\bar{\mathbf{\Omega}}$ is unique, if and only if the nondegeneracy condition holds.

The optimal value function $v(\mathbf{z})$ is convex (we can view $v(\cdot)$ also as a function of $\mathbf{z} := \operatorname{vec}(\mathbf{Z})$). By the theory of convex analysis we have that the subdifferential $\partial v(\boldsymbol{\sigma})$, of the optimal value function at $\mathbf{z} = \boldsymbol{\sigma}$, coincides with the set of optimal solutions with the minus sign of the dual problem (8), and that $v(\mathbf{z})$ is differentiable at $\mathbf{z} = \boldsymbol{\sigma}$ iff $\partial v(\boldsymbol{\sigma})$ is a singleton, that is, the dual problem (8) has unique optimal solution $\bar{\mathbf{\Omega}}$ (see Rockafellar, 1970). In the last case $\nabla v(\boldsymbol{\sigma}) = -\operatorname{vec}(\bar{\mathbf{\Omega}})$, where $\nabla v(\boldsymbol{\sigma})$ denotes the gradient vector, that is, the $p^2 \times 1$ vector of first order partial derivatives. We obtain that if the above nondegeneracy condition holds, then the optimal value function is differentiable at $\boldsymbol{\sigma}$ and $\nabla v(\boldsymbol{\sigma}) = -\operatorname{vec}(\bar{\mathbf{\Omega}})$. Since $g(\mathbf{z})$ is a smooth function of $v(\mathbf{z})$, it follows then that $g(\mathbf{z})$ is also differentiable at $\boldsymbol{\sigma}$ and $\nabla g(\boldsymbol{\sigma})$ can be calculated by the chain rule, that is

$$\nabla g(\boldsymbol{\sigma}) = (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-1}\nabla v(\boldsymbol{\sigma}) - (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-2}v(\boldsymbol{\sigma})\mathbf{1}_{p^2}. \tag{11}$$

Recall that the nondegeneracy condition can only hold if $(p-r)(p-r+1)/2 \leq p$, or equivalently if

$$r \geq \frac{2p + 1 - (8p+1)^{1/2}}{2}. \tag{12}$$

The above lower bound for the rank r is called the Ledermann bound. If the reduced rank r is greater than or equal to the Ledermann bound, then typically the dual problem has one optimal solution, and consequently in such cases the functions $v(\mathbf{z})$ and $g(\mathbf{z})$ are differentiable at $\mathbf{z} = \boldsymbol{\sigma}$.

Consider now a situation where the set of optimal solutions of the dual problem is not a singleton. By the above discussion this typically happens if the reduced rank r is less than the Ledermann bound. Let D be the set of optimal solutions of the dual problem (8) and denote $\mathcal{D} := -D$. The directional derivative of $v(\cdot)$, at $\boldsymbol{\sigma}$ in a direction $\boldsymbol{\delta}$, is defined as follows

$$v'(\boldsymbol{\sigma}, \boldsymbol{\delta}) := \lim_{t \downarrow 0} \frac{v(\boldsymbol{\sigma} + t\boldsymbol{\delta}) - v(\boldsymbol{\sigma})}{t}.$$

Recall that $\partial v(\boldsymbol{\sigma}) = \mathcal{D}$. It follows then by convex analysis that $v'(\boldsymbol{\sigma}, \cdot)$ is the support function of the set $\mathcal{D}$, that is,

$$v'(\boldsymbol{\sigma}, \boldsymbol{\delta}) = \sup_{\boldsymbol{\omega} \in \mathcal{D}} \boldsymbol{\omega}'\boldsymbol{\delta}, \tag{13}$$

(see Rockafellar, 1970). In particular, if the set $\mathcal{D}$ is a singleton, that is, the dual problem has only one optimal solution $\bar{\mathbf{\Omega}}$, then the right hand side of (13) is linear in $\boldsymbol{\delta}$ and hence $v(\cdot)$ is differentiable at $\boldsymbol{\sigma}$, as it was mentioned earlier.

By the chain rule we have

$$g'(\boldsymbol{\sigma}, \boldsymbol{\delta}) = (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-1}v'(\boldsymbol{\sigma}, \boldsymbol{\delta}) - (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-2}v(\boldsymbol{\sigma})\mathbf{1}_{p^2}'\boldsymbol{\delta}, \tag{14}$$

and hence $g'(\boldsymbol{\sigma}, \cdot)$ is the support function of the set

$$\mathcal{D}^* := (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-1}\mathcal{D} - (\mathbf{1}_p'\mathbf{\Sigma}\mathbf{1}_p)^{-2}v(\boldsymbol{\sigma})\mathbf{1}_{p^2}, \tag{15}$$

that is

$$g'(\boldsymbol{\sigma}, \boldsymbol{\delta}) = \sup_{\boldsymbol{\omega} \in \mathcal{D}^*} \boldsymbol{\omega}' \boldsymbol{\delta}. \tag{16}$$

Clearly the set $\mathcal{D}^*$ is a singleton iff the set $\mathcal{D}$ is a singleton. Note that the directional derivatives functions $v'(\boldsymbol{\sigma}, \cdot)$ and $g'(\boldsymbol{\sigma}, \cdot)$ are positively homogeneous, that is, $v'(\boldsymbol{\sigma}, t\boldsymbol{\delta}) = tv'(\boldsymbol{\sigma}, \boldsymbol{\delta})$ for any $t \geq 0$, and that $v'(\boldsymbol{\sigma}, \cdot)$ and $g'(\boldsymbol{\sigma}, \cdot)$ are convex and continuous.

3. Asymptotics by First Order Approximations

In this section we discuss implications of the first order differential structure of the functions $v(\mathbf{z})$ and $g(\mathbf{z})$ for an asymptotic analysis of CMTFA. Since the function $v(\mathbf{z})$ is convex, we have that its directional derivative gives a first order approximation of $v(\mathbf{z})$, at $\boldsymbol{\sigma}$, uniformly in $\boldsymbol{\delta} = \mathbf{z} - \boldsymbol{\sigma}$. That is

$$v(\boldsymbol{\sigma} + \boldsymbol{\delta}) = v(\boldsymbol{\sigma}) + v'(\boldsymbol{\sigma}, \boldsymbol{\delta}) + o(\|\boldsymbol{\delta}\|). \tag{17}$$

By the chain rule a similar approximation holds for the function $g(\mathbf{z})$.

Consider the random vector $\mathbf{y}_n := n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$, where, as before, $\mathbf{s} = \text{vec}(\mathbf{S})$ and $\mathbf{S}$ is the sample covariance matrix based on a sample of size n. We assume that the population distribution has fourth order moments, and hence $\mathbf{y}_n$ converges in distribution to a random vector $\mathbf{y}$ (denoted $\mathbf{y}_n \Rightarrow \mathbf{y}$) having a multivariate normal distribution , that is, $\mathbf{y} \sim N(0, \boldsymbol{\Gamma})$. It follows then that $\mathbf{y}_n$ is bounded in probability and hence, by (17),

$$v(\mathbf{s}) - v(\boldsymbol{\sigma}) = n^{-1/2} v'(\boldsymbol{\sigma}, \mathbf{y}_n) + o_p(n^{-1/2}). \tag{18}$$

A similar expansion can be written for $g(\mathbf{s})$ as well. That is, since $\hat{\rho} = g(\mathbf{s})$ and $\rho = g(\boldsymbol{\sigma})$,

$$\hat{\rho} - \rho = n^{-1/2} g'(\boldsymbol{\sigma}, \mathbf{y}_n) + o_p(n^{-1/2}). \tag{19}$$

Equations (18) and (19) imply the following asymptotics (such type of results are known as the Delta Theorem, e.g., Rao, 1973, p. 388).

Theorem 3.1. Suppose that $\mathbf{y}_n := n^{1/2}(\mathbf{s} - \boldsymbol{\sigma})$ converges in distribution to a random vector $\mathbf{y}$. Then

$$n^{1/2}[v(\mathbf{s}) - v(\boldsymbol{\sigma})] \Rightarrow v'(\boldsymbol{\sigma}, \mathbf{y}), \tag{20}$$

and

$$n^{1/2}(\hat{\rho} - \rho) \Rightarrow g'(\boldsymbol{\sigma}, \mathbf{y}). \tag{21}$$

In particular, if $\partial v(\boldsymbol{\sigma})$ is a singleton (i.e., the dual problem has a unique optimal solution), then $v'(\boldsymbol{\sigma}, \mathbf{y}) = [\nabla v(\boldsymbol{\sigma})]' \mathbf{y}$ is linear in $\mathbf{y}$. In that case $v'(\boldsymbol{\sigma}, \mathbf{y})$ has a normal distribution with zero mean and variance $[\nabla v(\boldsymbol{\sigma})]' \boldsymbol{\Gamma} [\nabla v(\boldsymbol{\sigma})]$. Consequently

$$n^{1/2}[v(\mathbf{s}) - v(\boldsymbol{\sigma})] \Rightarrow N\left(0, [\nabla v(\boldsymbol{\sigma})]' \boldsymbol{\Gamma} [\nabla v(\boldsymbol{\sigma})]\right). \tag{22}$$

If, moreover, the population distribution is normal, and hence $\boldsymbol{\Gamma} = 2\mathbf{M}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$, then the corresponding asymptotic variance can be calculated in the form

$$[\nabla v(\boldsymbol{\sigma})]' \boldsymbol{\Gamma} [\nabla v(\boldsymbol{\sigma})] = 2 \sum_{i,j=1}^{p-r} (\mathbf{f}_i' \bar{\boldsymbol{\Psi}} \mathbf{f}_j)^2, \tag{23}$$

where $\bar{\boldsymbol{\Psi}}$ is the optimal solution of the CMTFA problem (1), $\bar{\boldsymbol{\Omega}}$ is the (unique) optimal solution of the dual problem, and $\mathbf{f}_1, \ldots, \mathbf{f}_{p-r}$ are column vectors of a $p \times (p - r)$ matrix $\mathbf{F}$ such that

$\bar{\mathbf{\Omega}} = \mathbf{F}\mathbf{F}'$. Similar formulas can be derived, by the same arguments, for $\hat{\rho}$ (see Shapiro, 1982, Theorem 4.2).

The situation is quite different if the dual problem has more than one optimal solution. It follows from Theorem 3.1, and the analysis of the previous section, that $v(\mathbf{s})$ and $\hat{\rho}$ are asymptotically normal iff the dual problem has only one optimal solution. If the dual problem has multiple solutions, and hence the sets $\mathcal{D}$ and $\mathcal{D}^*$ are not singletons, then the expected values $E\{v'(\boldsymbol{\sigma}, \mathbf{y})\}$ and $E\{g'(\boldsymbol{\sigma}, \mathbf{y})\}$ are both nonzero and positive. We can view the quantities $n^{-1/2}E\{v'(\boldsymbol{\sigma}, \mathbf{y})\}$ and $n^{-1/2}E\{g'(\boldsymbol{\sigma}, \mathbf{y})\}$ as asymptotic biases of $v(\mathbf{s})$ and $\hat{\rho}$, respectively. Unfortunately it does not seem to be possible to give a closed form expression for these expectations. Nevertheless we can draw the following conclusions from the above discussion.

- If the dual problem (8) has a unique optimal solution, then the estimators $v(\mathbf{s})$ and $\hat{\rho}$ are asymptotically normal.
- If the dual problem (8) has more than one optimal solution, then the (asymptotic) biases of $v(\mathbf{s})$ and $\hat{\rho}$ are positive and of order $O(n^{-1/2})$.

In the next section we discuss situations where the dual problem (8) has a unique optimal solution. In such cases the (asymptotic) biases of the estimators $v(\mathbf{s})$ and $\hat{\rho}$ are of order $O(n^{-1})$.

4. Asymptotic Bias of the Sample g.l.b. by a Second Order Analysis

In this section we discuss a calculation of the bias of $\hat{\rho}$ in situations where the dual problem (8) has a unique optimal solution $\bar{\mathbf{\Omega}}$. Recall that typically this happens if the (reduced) rank $r := \text{rank}(\mathbf{\Sigma} - \bar{\mathbf{\Psi}})$ is greater than or equal to the Ledermann bound. We assume that $\mathbf{y}_n := n^{1/2}(\mathbf{s} - \boldsymbol{\sigma}) \Rightarrow N(0, \mathbf{\Gamma})$.

Suppose that the function $g(\mathbf{z})$ is twice continuously differentiable at $\mathbf{z} = \boldsymbol{\sigma}$. Then we have by a second order Taylor expansion that

$$g(\mathbf{z}) = g(\boldsymbol{\sigma}) + (\mathbf{z} - \boldsymbol{\sigma})'\nabla g(\boldsymbol{\sigma}) + \tfrac{1}{2}(\mathbf{z} - \boldsymbol{\sigma})'\nabla^2 g(\boldsymbol{\sigma})(\mathbf{z} - \boldsymbol{\sigma}) + o(\|\mathbf{z} - \boldsymbol{\sigma}\|^2), \tag{24}$$

where $\nabla^2 g(\boldsymbol{\sigma})$ is the $p^2 \times p^2$ Hessian matrix of second order partial derivatives. It follows that

$$g(\mathbf{s}) = g(\boldsymbol{\sigma}) + (\mathbf{s} - \boldsymbol{\sigma})'\nabla g(\boldsymbol{\sigma}) + \tfrac{1}{2}(\mathbf{s} - \boldsymbol{\sigma})'\nabla^2 g(\boldsymbol{\sigma})(\mathbf{s} - \boldsymbol{\sigma}) + o_p(n^{-1}). \tag{25}$$

If we assume further that the population distribution has moments higher than 4 (this is certainly true for the normal distribution), then the expected value of the remainder $o_p(n^{-1})$ in the above expansion is of order $o(n^{-1})$. Since $\mathbf{s}$ is an unbiased estimator of $\boldsymbol{\sigma}$, that is, $E(\mathbf{s}) = \boldsymbol{\sigma}$, it follows then that

$$\begin{aligned} E\{g(\mathbf{s})\} &= g(\boldsymbol{\sigma}) + \tfrac{1}{2}E\left\{(\mathbf{s} - \boldsymbol{\sigma})'\nabla^2 g(\boldsymbol{\sigma})(s - \boldsymbol{\sigma})\right\} + o(n^{-1}) \\ &= g(\boldsymbol{\sigma}) + \tfrac{1}{2}n^{-1}\ \text{tr}\left[\mathbf{\Gamma}\nabla^2 g(\boldsymbol{\sigma})\right] + o(n^{-1}). \end{aligned} \tag{26}$$

That is,

$$\tfrac{1}{2}n^{-1}\ \text{tr}[\mathbf{\Gamma}\nabla^2 g(\boldsymbol{\sigma})] \tag{27}$$

gives the asymptotic bias of the estimator $\hat{\rho}$, of order $O(n^{-1})$.

A similar analysis applies to the estimator $v(\mathbf{s})$, of the optimal value $v(\boldsymbol{\sigma})$, and hence the asymptotic bias of $v(\mathbf{s})$ is given by

$$\tfrac{1}{2}n^{-1}\ \text{tr}[\mathbf{\Gamma}\nabla^2 v(\boldsymbol{\sigma})], \tag{28}$$

provided $v(\cdot)$ is twice continuously differentiable at $\boldsymbol{\sigma}$. Note that by (6) we have that $g(\cdot)$ is twice differentiable at $\boldsymbol{\sigma}$ if $v(\cdot)$ is twice differentiable at $\boldsymbol{\sigma}$, and

$$\nabla^2 g(\boldsymbol{\sigma}) = (\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}\nabla^2 v(\boldsymbol{\sigma}) - (\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-2}\left(\nabla v(\boldsymbol{\sigma})\mathbf{1}_{p^2}' + \mathbf{1}_{p^2}\nabla v(\boldsymbol{\sigma})'\right) + 2v(\boldsymbol{\sigma})(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-3}\mathbf{1}_{p^2}\mathbf{1}_{p^2}'. \tag{29}$$

Unfortunately, calculation of the Hessian matrix $\nabla^2 v(\boldsymbol{\sigma})$, of second order partial derivatives of the optimal value function $v(\cdot)$, is not simple. Therefore we restrict our analysis in this section to the MTFA only. That is, we assume that the diagonal elements of the optimal solution $\bar{\boldsymbol{\Psi}}$, of the problem (1), are all positive (in case the constraint $\boldsymbol{\Psi} \succeq 0$ is ignored). This means that MTFA in the population does not display negative unique variances (so-called Heywood cases) and hence coincides with CMTFA. Also, in the present situation, the population g.l.b. coincides with a lower bound to reliability that was first proposed by Bentler (1972).

By the discussion of the previous section we have that the nondegeneracy is a sufficient and "almost" necessary condition for differentiability of the function $v(\cdot)$ at $\boldsymbol{\sigma}$. Moreover, it is possible to show that $v(\cdot)$ is twice continuously differentiable at $\boldsymbol{\sigma}$ iff the nondegeneracy and the strict complementarity conditions hold. This is quite a delicate result and its rigorous derivation would lead outside the scope of this paper. We refer to Shapiro (1997) for such derivations, and to Bonnans and Shapiro (2000) where such type of results are discussed in detail.

So suppose that the nondegeneracy and strict complementarity conditions hold, for the problem (1) (or, equivalently, for the problem (5) with $\mathbf{Z} = \boldsymbol{\Sigma}$), and consider the following optimization problem

$$\begin{array}{ll} \min_{\mathbf{h}\in\mathbb{R}^p} & \left\{ f(\mathbf{h}) := \operatorname{tr}[\bar{\boldsymbol{\Omega}}(\mathbf{H}+\boldsymbol{\Delta})(\boldsymbol{\Sigma}-\bar{\mathbf{X}})^{\dagger}(\mathbf{H}+\boldsymbol{\Delta})] \right\} \\ \text{subject to} & \mathbf{F}'\mathbf{H}\mathbf{F} + \mathbf{F}'\boldsymbol{\Delta}\mathbf{F} = 0, \end{array} \tag{30}$$

depending on $p \times p$ symmetric matrix $\boldsymbol{\Delta}$. Here $\mathbf{H}$ is $p \times p$ diagonal matrix $\mathbf{H} = \operatorname{diag}(h_1, \ldots, h_p)$, $\mathbf{h} = \operatorname{diag}(\mathbf{H})$ is the corresponding vector, $\bar{\mathbf{X}}$ and $\bar{\boldsymbol{\Omega}}$ are optimal solutions of the problem (5) and its dual (8), respectively, $(\boldsymbol{\Sigma} - \bar{\mathbf{X}})^{\dagger}$ is the Moore-Penrose pseudoinverse of $\boldsymbol{\Sigma} - \bar{\mathbf{X}}$, $\mathbf{F}$ is $p \times (p - r)$ matrix of full column rank such that $\mathbf{F}\mathbf{F}' = \bar{\boldsymbol{\Omega}}$. Recall that $\bar{\boldsymbol{\Psi}} = \bar{\mathbf{X}}$. Note also that if $\boldsymbol{\Sigma} - \bar{\mathbf{X}} = \sum_{i=1}^{r} \lambda_i \mathbf{a}_i \mathbf{a}_i'$ is the spectral decomposition of $\boldsymbol{\Sigma} - \bar{\mathbf{X}}$, that is, λ_i, $i = 1, \ldots, r$, are nonzero eigenvalues and $\mathbf{a}_i$ are corresponding orthogonal eigenvectors of $\boldsymbol{\Sigma} - \bar{\mathbf{X}}$, then

$$(\boldsymbol{\Sigma} - \bar{\mathbf{X}})^{\dagger} = \sum_{i=1}^{r} \lambda_i^{-1} \mathbf{a}_i \mathbf{a}_i'.$$

The objective function of the optimization problem (30) is quadratic and the constraints are linear in $\mathbf{h}$. Let $\vartheta(\boldsymbol{\Delta})$ be the optimal value of (30), considered as a function of $\boldsymbol{\Delta}$. This is a quadratic function of $\boldsymbol{\delta} := \operatorname{vec}(\boldsymbol{\Delta})$, that is, $\vartheta(\boldsymbol{\Delta}) = \boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}$ for some $p^2 \times p^2$ symmetric matrix $\mathbf{Q}$, provided the matrix $\bar{\boldsymbol{\Omega}} * (\boldsymbol{\Sigma} - \bar{\mathbf{X}})^{\dagger}$ is positive definite on the linear space $\{\mathbf{h} : \mathbf{F}'\mathbf{H}\mathbf{F} = 0\}$. It follows then that $\nabla^2 v(\boldsymbol{\sigma}) = 2\mathbf{Q}$. This result can be proved by applying the Implicit Function Theorem to a certain system of equations (Shapiro, 1997).

Note that if $(p - r)(p - r + 1)/2 = p$, that is, the Ledermann bound is exact, then the number of equations in $\mathbf{F}'\mathbf{H}\mathbf{F} + \mathbf{F}'\boldsymbol{\Delta}\mathbf{F} = 0$ is equal to the number p of unknowns, and in that case these equations define a single solution $\mathbf{h} = \mathbf{h}(\boldsymbol{\Delta})$, which is also the optimal solution of (30). The function $\mathbf{h}(\boldsymbol{\Delta})$ is linear. By putting $\mathbf{h}(\boldsymbol{\Delta})$ into the objective function of (30), it is relatively easy to calculate the matrix $\mathbf{Q}$ in that case.

Let us calculate now the matrix $\mathbf{Q}$ in a closed form. We can write the objective function of the problem (30) as follows

$$f(\mathbf{h}) = \operatorname{tr}(\bar{\boldsymbol{\Omega}}\mathbf{H}\boldsymbol{\Phi}\mathbf{H}) + 2\operatorname{tr}(\bar{\boldsymbol{\Omega}}\mathbf{H}\boldsymbol{\Phi}\boldsymbol{\Delta}) + \operatorname{tr}(\bar{\boldsymbol{\Omega}}\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Delta}) = \mathbf{h}'(\boldsymbol{\Phi} * \bar{\boldsymbol{\Omega}})\mathbf{h} + 2\mathbf{h}'\mathbf{w} + \operatorname{tr}(\bar{\boldsymbol{\Omega}}\boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Delta}),$$

where $\boldsymbol{\Phi} := (\boldsymbol{\Sigma} - \bar{\mathbf{X}})^{\dagger}$ and $\mathbf{w} := \operatorname{vec}(\operatorname{diag}(\boldsymbol{\Phi}\boldsymbol{\Delta}\bar{\boldsymbol{\Omega}}))$. The constraint equations of (30) can be written in the form

$$(\mathbf{F}' \otimes \mathbf{F}')\operatorname{vec}(\mathbf{H}) + (\mathbf{F}' \otimes \mathbf{F}')\boldsymbol{\delta} = 0, \tag{31}$$

where $\boldsymbol{\delta} := \text{vec}(\boldsymbol{\Delta})$. Define the $p^2 \times p$ matrix $\mathbf{W} := [\mathbf{e}_1\mathbf{e}_1', \ldots, \mathbf{e}_p\mathbf{e}_p']'$, where $\mathbf{e}_i$ is the i-th column vector of the $p \times p$ identity matrix $\mathbf{I}_p$. The general solution of equations (31) can be written as

$$\mathbf{h} = \mathbf{N}\boldsymbol{\eta} - \mathbf{A}^{\dagger}(\mathbf{F}' \otimes \mathbf{F}')\text{vec}(\boldsymbol{\delta})$$

for some vector $\boldsymbol{\eta}$, where $\mathbf{A} := (\mathbf{F}' \otimes \mathbf{F}')\mathbf{W}$ is a $(p-r)^2 \times p$ matrix, and $\mathbf{N}$ is a $p \times [p - \frac{1}{2}(p-r)(p-r+1)]$ matrix whose column vectors generate the null space of the matrix $\mathbf{A}'$.

Then, after some algebra, we obtain

$$\mathbf{Q} = (\mathbf{WB} + \mathbf{I}_{p^2})'(\boldsymbol{\Phi} \otimes \bar{\boldsymbol{\Omega}})(\mathbf{WB} + \mathbf{I}_{p^2}), \tag{32}$$

where

$$\mathbf{C} := \mathbf{N}[\mathbf{N}'(\bar{\boldsymbol{\Omega}} * \boldsymbol{\Phi})\mathbf{N}]^{\dagger}\mathbf{N}'(\boldsymbol{\Phi} * \bar{\boldsymbol{\Omega}}),$$

$$\mathbf{B} := (\mathbf{C} - \mathbf{I}_p)\mathbf{A}^{\dagger}(\mathbf{F}' \otimes \mathbf{F}') - \mathbf{N}[\mathbf{N}'(\boldsymbol{\Phi} * \bar{\boldsymbol{\Omega}})\mathbf{N}]^{\dagger}\mathbf{N}'\mathbf{W}'(\boldsymbol{\Phi} \otimes \bar{\boldsymbol{\Omega}}).$$

Note that $\mathbf{B}$ is of order $p \times p^2$ and $\mathbf{C}$ is of order $p \times p$.

Let us make the following observations. If the Ledermann bound (12) holds exactly, that is, the reduced rank r is equal to the right hand side of (12), then $p - \frac{1}{2}(p-r)(p-r+1) = 0$, and hence the number of columns of the matrix $\mathbf{N}$ is zero. In that case the matrix $\mathbf{C}$ is a null matrix.

The constraints of the problem (30) can be written in the form

$$\mathbf{h}'(\mathbf{f}_i * \mathbf{f}_j) + \mathbf{f}_i'\boldsymbol{\Delta}\mathbf{f}_j = 0, \quad 1 \le i \le j \le p - r. \tag{33}$$

Therefore under the nondegeneracy condition, the linear equations (33) are linearly independent. Also we have that, for any $\mathbf{h} \in \mathbb{R}^p$,

$$\mathbf{h}'(\bar{\boldsymbol{\Omega}} * \boldsymbol{\Phi})\mathbf{h} = \text{tr}(\bar{\boldsymbol{\Omega}}\mathbf{H}\boldsymbol{\Phi}\mathbf{H}) = \text{tr}(\mathbf{F}'\mathbf{H}\boldsymbol{\Phi}\mathbf{H}\mathbf{F}).$$

Therefore, under the nondegeneracy and strict complementarity conditions, if a nonzero diagonal matrix $\mathbf{H}$ satisfies the equations $\mathbf{F}'\mathbf{H}\mathbf{F} = 0$, then columns of the matrix $\mathbf{H}\mathbf{F}$ belong to the range space of $\boldsymbol{\Phi}$, and hence $tr(\mathbf{F}'\mathbf{H}\boldsymbol{\Phi}\mathbf{H}\mathbf{F}) > 0$. That is, the nondegeneracy and strict complementarity conditions imply that the matrix $\bar{\boldsymbol{\Omega}} * \boldsymbol{\Phi}$ is positive definite on the linear space $\{\mathbf{h} : \mathbf{F}'\mathbf{H}\mathbf{F} = 0\}$.

If the reduced rank r is equal to $p - 1$, then the matrix $\bar{\boldsymbol{\Omega}} = \mathbf{F}\mathbf{F}'$ is of rank one, and hence $\mathbf{F}$ is of order $p \times 1$, that is, it is a column vector. Since $\text{diag}(\bar{\boldsymbol{\Omega}}) = \mathbf{I}_p$, it follows that the components of $\mathbf{F}$ are ± 1. In that case $\mathbf{F}$ is not changed by "small" perturbations of the elements of the matrix $\boldsymbol{\Sigma}$. Moreover, by the necessary condition (9) we have that $\boldsymbol{\Sigma}\mathbf{F} - \bar{\mathbf{X}}\mathbf{F} = 0$. Therefore under "small" perturbations of $\boldsymbol{\Sigma}$, the optimal solution $\bar{\mathbf{X}}$, and hence the function $v(\cdot)$, is a linear function of $\boldsymbol{\Sigma}$. Consequently, in that case, $\mathbf{Q}$ is a null matrix. This can be also derived directly from the optimization problem (30). Indeed, without loss of generality we can assume that all components of $\mathbf{F}$ are $+1$, that is, $\mathbf{F} = \mathbf{1}_p$. Then the optimization problem (30) can be written in the form

$$\begin{aligned} &\min_{\mathbf{h} \in \mathbb{R}^p} && \mathbf{h}'\boldsymbol{\Phi}\mathbf{h} + 2\mathbf{d}'\boldsymbol{\Phi}\mathbf{h} + \mathbf{d}'\boldsymbol{\Phi}\mathbf{d} \\ &\text{subject to} && \mathbf{1}_p'(\mathbf{h} + \mathbf{d}) = 0, \end{aligned} \tag{34}$$

where $\mathbf{d} := \boldsymbol{\Delta}\mathbf{1}_p$. Since $\mathbf{h}'\boldsymbol{\Phi}\mathbf{h} + 2\mathbf{d}'\boldsymbol{\Phi}\mathbf{h} + \mathbf{d}'\boldsymbol{\Phi}\mathbf{d} = (\mathbf{h}+\mathbf{d})'\boldsymbol{\Phi}(\mathbf{h}+\mathbf{d})$, and $\boldsymbol{\Phi}$ is positive definite over the linear space orthogonal to $\mathbf{1}_p$, it follows that the optimal value of the optimization problem (34) is zero for any $\mathbf{d}$, and hence $\mathbf{Q} = 0$. We obtain the interesting conclusion that if $r = p - 1$, then the asymptotic bias of $v(\mathbf{s})$ is zero.

In case the population distribution is normal, and hence $\boldsymbol{\Gamma} = 2\mathbf{M}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$, we can write the asymptotic bias of the g.l.b. estimator $\hat{\rho}$ in a more explicit form. Note that the matrix $\mathbf{M}_p$ has the following properties: $\mathbf{M}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) = (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{M}_p$, $\mathbf{M}_p\mathbf{1}_{p^2} = \mathbf{1}_{p^2}$ and $\mathbf{M}_p[\text{vec}(\bar{\boldsymbol{\Omega}})] = \text{vec}(\bar{\boldsymbol{\Omega}})$.

Therefore, by using (29), the asymptotic bias of $\hat{\rho}$, given in (27), can be written as follows

$$\tfrac{1}{2}n^{-1}\,\mathrm{tr}[\boldsymbol{\Gamma}\nabla^2 g(\boldsymbol{\sigma})] = \tfrac{1}{2}n^{-1}(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}\left[2tr(\boldsymbol{\Gamma}\mathbf{Q}) + 4(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}(\mathbf{1}_p'\boldsymbol{\Sigma}\bar{\boldsymbol{\Omega}}\boldsymbol{\Sigma}\mathbf{1}_p) + 4v(\boldsymbol{\sigma})\right].$$

Note that, since $\bar{\boldsymbol{\Omega}} = \mathbf{F}\mathbf{F}'$ and $(\boldsymbol{\Sigma} - \bar{\mathbf{X}})\mathbf{F} = 0$, it follows that

$$\mathbf{1}_p'\boldsymbol{\Sigma}\bar{\boldsymbol{\Omega}}\boldsymbol{\Sigma}\mathbf{1}_p = \mathbf{1}_p'\bar{\mathbf{X}}\mathbf{F}\mathbf{F}'\bar{\mathbf{X}}\mathbf{1}_p.$$

In particular, if $r = p - 1$, then $\mathbf{Q} = 0$ and $\mathbf{F}$ is a column vector and hence in that case the asymptotic bias of $\hat{\rho}$ takes the form

$$\tfrac{1}{2}n^{-1}\,\mathrm{tr}[\boldsymbol{\Gamma}\nabla^2 g(\boldsymbol{\sigma})] = 2n^{-1}(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}\left[(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}(\mathbf{1}_p'\bar{\mathbf{X}}\mathbf{F})^2 - \mathrm{tr}(\bar{\mathbf{X}})\right].$$

Since all diagonal elements of the (diagonal) matrix $\bar{\mathbf{X}}$ are positive and the components of $\mathbf{F}$ are ± 1, we obtain that $(\mathbf{1}_p'\bar{\mathbf{X}}\mathbf{F})^2 \leq [\mathrm{tr}(\bar{\mathbf{X}})]^2$ and hence

$$(\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p)^{-1}(\mathbf{1}_p'\bar{\mathbf{X}}\mathbf{F})^2 - \mathrm{tr}(\bar{\mathbf{X}}) \leq \mathrm{tr}(\bar{\mathbf{X}})\left[\frac{\mathrm{tr}(\bar{\mathbf{X}})}{\mathbf{1}_p'\boldsymbol{\Sigma}\mathbf{1}_p} - 1\right] \leq 0.$$

It follows that if $r = p - 1$ and the population distribution is normal, then the asymptotic bias of the g.l.b. estimator is less than or equal to zero.

5. Numerical Examples

The asymptotic results of the present paper crystallize in the bias estimate (27), with associated asymptotic variance given in Shapiro (1982, Theorem 4.2). In the present section, some empirical results for these estimates will be obtained. We also examine the reduced trace function $v(\mathbf{S})$ with the corresponding asymptotic bias given in (28) and associated asymptotic variance given by (23).

The formulas for first and second order derivatives of the functions $v(\cdot)$ and $g(\cdot)$ were verified numerically, by calculating finite differences, and were found to be accurate. In order to examine the behavior of the estimates $v(\mathbf{S})$ and $g(\mathbf{S})$ in samples we start from a given population covariance matrix $\boldsymbol{\Sigma}$, with reduced rank r. From this population, we draw 500 samples of sizes $n = 100$, $n = 500$ and $n = 2000$, respectively. Increasing the number of samples beyond 500 has also been tried but did not seem to further affect the results. Each of the 500 samples is generated in the following manner. We sample n cases (row vectors of p elements) from the multivariate normal $N(0, \mathbf{I}_p)$ distribution, and postmultiply each vector by $\boldsymbol{\Sigma}^{1/2}$. This yields n realizations from the multivariate $N(0, \boldsymbol{\Sigma})$ distribution. Next, we compute the sample covariance matrix $\mathbf{S}$ by subtracting the sample means, taking sums of squares and cross-products, and dividing by $n - 1$. Repeating this process 500 times gives us 500 sample covariance matrices from the multivariate normal $N(0, \boldsymbol{\Sigma})$ distribution, based on a fixed sample size n.

For each sample, we take $\mathbf{S}$ to be an estimate of $\boldsymbol{\Sigma}$, and evaluate all matrices by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$. This yields in particular:

- the sample g.l.b. value $g(\mathbf{S})$,
- the bias estimate (27) for $g(\mathbf{S})$,
- the sample reduced trace value $v(\mathbf{S})$,
- the bias estimate (28) for $v(\mathbf{S})$.

We inspect the correlations between these values, and report the observed averages of the four statistics over 500 samples, and, in addition:

- the observed variance of the sample g.l.b. $g(\mathbf{S})$,
- the observed variance of the sample reduced trace $v(\mathbf{S})$.

The above values are reported in the first rows of the tables. Tables 1 through 3 refer to a population covariance matrix $\boldsymbol{\Sigma}$ with $p = 5$ variables and reduced rank $r = 3$. Tables 4–6 refer to a $\boldsymbol{\Sigma}$ with $p = 10$ variables and reduced rank $r = 8$.

The first four statistics are also evaluated for $\boldsymbol{\Sigma}$, yielding four comparable population parameters, reported in the first four cells of the second rows of the tables below. These parameters can be interpreted as the sample statistics that would be observed if the sample covariance matrix $\mathbf{S}$ (based on a particular sample size n) would happen to coincide with $\boldsymbol{\Sigma}$. Also, we evaluate the asymptotic variances of $g(\mathbf{S})$ and $v(\mathbf{S})$, using expressions from Shapiro (1982, Theorem 4.2), and (23), respectively. Note that both of these expressions have to be divided by n to give the desired asymptotic variances. These variances are reported in the last two places of the second row of the tables.

Finally, we evaluate the "real g.l.b. bias" and "real reduced trace bias" as the sample mean of $g(\mathbf{S})$ minus $g(\boldsymbol{\Sigma})$, and as the sample mean of $v(\mathbf{S})$ minus $v(\boldsymbol{\Sigma})$, respectively, see the third rows of the tables. The simulations were carried out with two covariance matrices based on $p = 5$ (Example 1) and $p = 10$ (Example 2) items, respectively of a scholastic achievement test. Heywood case were not encountered (except for a few cases with $n = 100$), which means that MTFA here coincides with CMTFA. Throughout the simulations, the correlations (over 500 samples) between g-bias and v-bias were close to 1.0.

Example 1. Five items from a scholastic achievement test. Results for three different sample sizes are given in Tables 1, 2 and 3. In general the asymptotic results based on samples tend to underestimate the real bias in the g.l.b. and its numerator. However, it can also be seen that the approximations become better as sample sizes increase. Because this example has only five

TABLE 1.
Simulation Results for $p = 5$ and $n = 100$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.5467	0.0370	−0.6084	0.0615	0.0061	0.0037
population	0.4799	0.0674	−0.7190	0.1075	0.0097	0.0082
real bias	0.0669		0.1005			

TABLE 2.
Simulation Results for $p = 5$ and $n = 500$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.5015	0.0143	−0.6836	0.0222	0.0014	0.0009
population	0.4799	0.0135	−0.7190	0.0215	0.0019	0.0016
real bias	0.0216		0.0318			

TABLE 3.
Simulation Results for $p = 5$ and $n = 2000$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.4883	0.0059	−0.7073	0.0088	0.0004	0.0003
population	0.4799	0.0034	−0.7190	0.0054	0.0005	0.0004
real bias	0.0085		0.0117			

variables, solutions with reduced rank 4 occur quite frequently, namely, 152, 61 and 23 times, respectively. When the sample statistics are evaluated separately for these cases, we find zero v-bias and negative g-bias throughout (as proven above). However, we also find higher g.l.b.-values and higher v-bias for these cases. It should be emphasized that these results are not incompatible with theory. It is the population reduced rank which affects the asymptotic bias, rather than the reduced rank we happen to find in samples. Inspection of the correlations reveals that v-bias and g-bias correlate near 1, but the correlation between g.l.b. and g-bias is negative, around $-.20$. However, for the specific cases with reduced rank 4, it appears that the latter correlations are extremely high (between .93 and .97), whereas they are negative (between $-.34$ and $-.18$) for the cases with reduced rank 3. This suggests that bias correction may be possible for those samples where the reduced rank happens to be $p - 1$.

Example 2. Ten (different) items from a scholastic achievement test. Tables 4 through 6 give the results for three different sample sizes. The asymptotic results look better than in example 1, especially for $n = 2000$. The reduced rank in the population is 8, and it varies from 7 to 9 in samples. Again, it has been verified that the samples display less bias as the reduced rank is lower. The correlations between the g.l.b. and the bias vary from $-.29$ to $-.47$. Again, looking at cases with reduced rank $p - 1$ reveals a totally different picture, they are .99 or 1.00 for each of the three tables. The frequencies of these cases were 8, 10 and 13, respectively, for Tables 4, 5 and 6.

Example 3. In addition to the simulations shown above, we also looked at a contrived case with $p = 8$, and reduced rank $r = 1$ in the population (Table 7). The asymptotic bias results of section 4 of this paper do not apply in that example. However, the real bias in the g.l.b. is not much higher than usual with $n = 500$.

TABLE 4.
Simulation Results for $p = 10$ and $n = 100$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.7722	0.0398	−1.1811	0.2290	0.0014	0.0099
population	0.6801	0.2126	−1.7012	1.1646	0.0032	0.0335
real bias	0.0921		0.5201			

TABLE 5.
Simulation Results for $p = 10$ and $n = 500$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.7153	0.0221	−1.5096	0.1231	0.0005	0.0032
population	0.6801	0.0425	−1.7012	0.2329	0.0006	0.0067
real bias	0.0352		0.1916			

TABLE 6.
Simulation Results for $p = 10$ and $n = 2000$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.6923	0.0098	−1.6368	0.0538	0.0001	0.0010
population	0.6801	0.0106	−1.7012	0.0582	0.0002	0.0017
real bias	0.0122		0.0644			

TABLE 7.
Reduced Rank $r = 1$ in Population, $p = 8$ and $n = 500$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.6045	0.0270	−13.9140	1.0037	0.0008	0.3099
population	0.5478	0.0006	−16.0000	0.0859	0.0010	0.2865
real bias	0.0568		2.0860			

TABLE 8.
Reduced Rank $r = 3$ in Population, $p = 4$ and $n = 100$

	g.l.b.	g-bias	$v(\mathbf{S})$	v-bias	var(g.l.b.)	var($v(\mathbf{S})$)
sample	0.4452	0.0337	−0.5226	0.0414	0.0096	0.0031
population	0.3980	−0.0120	−0.5808	0.0000	0.0143	0.0067
real bias	0.0472		0.0582			

Example 4. Finally, we also examined a case with population reduced rank $p - 1$. Here we have asymptotic v-bias zero, and negative asymptotic g-bias. Nevertheless, Table 8 shows that, for a sample of $n = 100$, the average sample g.l.b. is still .047 higher than the population value. Interestingly, this cannot be attributed to the occurrence of reduced rank 2 (in 268 of the 500 cases): Like in Tables 1,...,6, samples with lower reduced rank perform better in terms of bias. This is not to say that these results contradict the asymptotic theory. As we let n increase, the real bias rapidly goes to zero, and it is slightly negative for $n = 5000$. For the data of Tables 1–7, negative bias in the g.l.b. does not arise, not even when a sample size of $n = 5000$ is used. The correlations between g.l.b. and g-bias follow the same pattern as before: When the cases with reduced rank 3 are considered separately, we find a value near 1, whereas for the rank 2 cases, the correlation is negative.

6. Discussion

The above tables reflect the well-known fact that the real bias in the g.l.b. and in the reduced trace decrease with sample size, and it can be seen that the asymptotic bias estimates and their variances follow the same pattern. As sample sizes increase, the sample statistics behave more and more according to the asymptotic theory.

The simulation study suggests that the asymptotic results cannot be used in small samples. In particular, the values of Table 4 are completely off the mark. The simulation results confirm the theoretical results, but also imply directions for future research. In particular, the asymptotic theory indicates that the bias should be bigger when the population reduced rank is relatively low. Still this was not always the case in simulation experiments. On the other hand, the prospects for a reliable bias correction seem remote for all cases considered, except those with reduced rank $p - 1$. This is another aspect that requires further study.

The examples above represent cases with low reliability. Since it is well-known that the bias in the g.l.b. tends to be higher when the g.l.b. is small, we have been dealing with large bias situations. The particular reason for taking such cases is that they do not lead to Heywood cases (indeed we found very few). Because the theory of the present paper does not apply to CMTFA, it has been imperative to have cases where MTFA happens to coincide with CMTFA.

The value of the variance of $v(\mathbf{S})$, calculated according to the asymptotic formula (23), is given in the last column of the tables under the name "population" var($v(\mathbf{S})$). The corresponding "true" variance of $v(\mathbf{S})$, estimated by the simulation, is given in the same column under the name "sample" var($v(\mathbf{S})$). We can see from the tables that the theoretical value typically overestimates the corresponding "true" variance (except in Table 7, where formula (23) does not

apply). This indicates that the first order approximation, on which formula (23) is based, is not accurate enough for the sample sizes considered. It is possible to get a better approximation by incorporating the corresponding second order term of the Taylor expansion in such a variance evaluation. This, however, although theoretically possible, would involve the calculations of 8-th order moments and could be very messy.

A key assumption of the theoretical developments of section 4 of the present paper is that the reduced rank associated with MTFA in the population is at or above the Ledermann bound (see (12)). This has the following practical implications. When the population covariance matrix $\mathbf{\Sigma}$ is constructed on the basis of a small number of hypothetical latent factors, as was done by Cronbach (1988), and also for the data reported in Table 7, the asymptotic results derived in section 4 do not apply. As it was argued in section 3, in such cases the asymptotic bias of the sample estimate of g.l.b. is of order $O(n^{-1/2})$.

On the other hand, it has been shown by Shapiro (1982) that the reduced rank is almost surely at or above the Ledermann bound, when the covariance matrix is obtained on the basis of a random sample from a continuous distribution. This means that the results of section 4 of this paper do apply when the population is defined empirically, either as the result of "sampling the entire population," or when an empirically obtained sample is treated as if it were the population, as we did above. Simulation studies on the same basis have recently been carried out by Socan (1999) and Verhelst (1998). They compared the g.l.b. to other lower bounds in terms of sampling bias. The g.l.b. bias they report is in line with the results of the present paper.

References

Alizadeh, F., Haeberly, J.–P., Nayakkankuppam, M.V., Overton, M.L., & Schmieta, S. (1997). SDPPACK (Version 0.9 BETA) [Computer software]. New York: Authors. (http://cs.nyu.edu/cs/faculty/overton/sdppack/sdppack.html)

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

Bentler, P.M., & Woodward, J.A. (1980). Inequalities among lower bounds to reliability: with applications to test construction and factor analysis. *Psychometrika, 45*, 249–267.

Bonnans, J.F., & Shapiro, A. (2000). *Perturbation analysis of optimization problems*. New York: Springer-Verlag.

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

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

Cronbach, L.J. (1988). Internal consistency of tests: Analyses old and new. *Psychometrika, 53*, 63–70.

Della Riccia, G., & Shapiro, A. (1982). Minimum rank and minimum trace of covariance matrices. *Psychometrika, 47*, 443–448.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 225–282.

Jackson, P.H., & Agunwamba, C.C. (1977). Lower bounds for the reliability of total scores on a test composed of nonhomogeneous items: I: Algebraic lower bounds. *Psychometrika, 42*, 567–578.

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

Rockafellar, R.T. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press.

Shapiro, A. (1982). Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis, *Psychometrika, 47*, 187–199.

Shapiro, A. (1997). First and second order analysis of nonlinear semidefinite programs, *Mathematical Programming, Series B, 77*, 301–320.

Socan, G. (1999, July). *Bias in some lower bounds to reliability in medium-sized samples* (poster presentation). 11th European Meeting of the Psychometric Society, Lüneburg, Germany.

ten Berge, J.M.F., & Zegers, F.E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika, 43*, 575–579.

ten Berge, J.M.F., Snijders, T.A.B., & Zegers, F.E. (1981). Computational aspects of the greatest lower bound to reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 201–213.

Vandenberghe, L., & Boyd, S. (1996). Semidefinite programming, *SIAM Review, 38*, 49–95.

Verhelst, N.D. (1998). *Estimating the reliability of a test from a single test administration* (Measurement and Research Department Report 98–2). Arnhem: CITO.

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

Manuscript received 22 FEB 1999
Final version received 20 SEP 1999