A Class of Population Covariance Matrices for Monte Carlo Simulation

Ke-Hai Yuan*, Kentaro Hayashi** and Hirokazu Yanagihara***

*Department of Psychology
University of Notre Dame
Notre Dame, Indiana 46556, USA

***Department of Psychology
University of Hawaii
Honolulu, Hawaii 96822-2294, USA

***Department of Social Systems and Management
Graduate School of Systems and Information Engineering
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

March 31, 2006

The research was supported by NSF grant DMS-0437167, the James McKeen Cattell Fund and the Japan Ministry of Education, Science, Sports and Culture Grant-in-Aid for Young Scientists (B) #17700274. Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

Abstract

Model evaluation in covariance structure analysis is critical before the results can be trusted. Due to finite sample sizes and unknown distributions of practical data, existing conclusion regarding a particular statistic may not be applicable in practice. The bootstrap procedure automatically takes care of the unknown distribution and, for a given sample size, also provides more accurate results than those based on standard asymptotics. But it needs a matrix to play the role of the population covariance matrix. The closer the matrix is to the true population covariance matrix, the more valid the bootstrap inference is. The current paper proposes a class of covariance matrices by combining theory and data. Thus, a proper matrix from this class is closer to the true population covariance matrix than those constructed by any existing methods. Each of the covariance matrices is easy to generate and also satisfies several desired properties. Examples verify the properties of the matrices and illustrate the details for creating a matrix with a given amount of misspecification.

Keywords: Bootstrap, model misspecification, noncentrality parameter.

1. Introduction

Structural equation modeling (SEM), covariance structure analysis (CSA) in particular, has been widely used in the social and behavioral sciences (Bentler & Dudgeon, 1996; Bollen, 2002; MacCallum & Austin, 2000). The advantage of CSA is that manifest variables, latent variables as well as measurement errors can be modeled and tested simultaneously. In applying CSA models, model evaluation is critical before the results can be trusted. Various procedures have been developed for such purposes (Bentler, 1983; Bentler & Dijkstra, 1985; Browne, 1984; Kano, Berkane, & Bentler, 1990; Jöreskog, 1969; Satorra, 1989; Satorra & Bentler, 1994; Shapiro & Browne, 1987; Yuan & Bentler, 1997, 1998, 1999a). The behaviors of several commonly used statistics for overall model evaluation are extensively studied by either asymptotics (Amemiya & Anderson, 1990; Browne & Shapiro, 1988; Shapiro, 1983; Yuan & Bentler, 1999b) or simulation (see Bentler & Yuan, 1999; Chou, Bentler & Satorra, 1991; Curran, West & Finch, 1996; Hu, Bentler, & Kano, 1992; Muthén & Kaplan, 1985; Yuan & Bentler, 1998). However, the conclusion obtained may not be valid when applying the statistics to a practical data. For example, the asymptotic robustness condition for the likelihood ratio statistic (Amemiya & Anderson, 1990; Browne & Shapiro, 1988) may not be satisfied in any real data analysis. Similarly, the data generation scheme in most simulation studies may rend the rescaled statistic (Satorra & Bentler, 1994) to asymptotically follow a chi-square distribution (Yuan & Bentler, 1999b), such obtained results may not apply to real data (see Yuan & Hayashi, 2003). Of course, the validity of any asymptotic properties is closely related to the sample size of the data, which is typically out of control in practice. Even when the sample is normally distributed, the behavior of a statistic is also closely related to the model and the population covariance matrix as well as their distance, as measured by a discrepancy function. When the model gradually departs from the population covariance matrix, the likelihood ratio statistic is better described by a noncentral chi-square distribution at beginning and then better described by a normal distribution (Yuan, Hayashi & Bentler, 2005).

When the distribution of the sample is unknown or when the sample size is not large enough, the bootstrap approach represents a promising alternative (Efron, 1979). The boot-

strap has demonstrated its potential in dealing with various problems that challenge traditional statistical methods (Efron & Tibshirani, 1993; Davison & Hinkley, 1997). Bootstrap methods have been applied to covariance structure models by various authors. Beran and Srivastava (1985) set out the theoretical foundation for bootstrap inference about covariance matrices in general. Bollen and Stine (1993) introduced the bootstrap approach to studying the distribution of test statistics and fit indices and clearly demonstrated the importance of choosing the proper matrix in playing the role of bootstrap population. Chatterjee (1984), Boomsma (1986) and Bollen and Stine (1990) used bootstrap to study standard errors in covariance structure models. Yung and Bentler (1996) reviewed many applications of bootstrap in covariance structure analysis. Yung and Bentler (1996), Yuan and Hayashi (2003) and Yuan and Marshall (2004) used bootstrap to estimate power and lack of fit in CSA. Enders (2002) used bootstrap to study the goodness of fit in CSA with missing data. In contrast with inferences that are based on normal theory maximum likelihood (ML), the bootstrap approach does not assume normally distributed data. Even when data are normally distributed, at a given sample size, the bootstrap may give more accurate results than those based on standard asymptotics due to its second order accuracy (see Hall & Titterington, 1989).

Let $\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_n$ be a sample with a sample covariance matrix $\mathbf{S}$ whose population counterpart is $\mathbf{\Sigma}_0$. Conditional on the sample, the bootstrap repeatedly draw samples from a known empirical distribution function. This empirical distribution plays the role of the "population" for the bootstrap samples. Because $\mathbf{\Sigma}_0$ is generally unknown, one has to find an alternative $\mathbf{S}_a$ to play the role of $\mathbf{\Sigma}_0$. With a given $\mathbf{S}_a$, the remainder of the bootstrap is standard. For example, $\mathbf{x}_i$ can be transformed to $\mathbf{x}_i^{(a)}$ by

$$\mathbf{x}_i^{(a)} = \mathbf{S}_a^{1/2}\mathbf{S}^{-1/2}\mathbf{x}_i,\ i = 1, 2, \cdots, n, \tag{1}$$

where $\mathbf{S}_a^{1/2}$ is a $p\times p$ matrix satisfying $(\mathbf{S}_a^{1/2})(\mathbf{S}_a^{1/2})' = \mathbf{S}_a$. The following steps are to obtain the bootstrap samples by sampling with replacement from $(\mathbf{x}_1^{(a)}, \mathbf{x}_2^{(a)}, \ldots, \mathbf{x}_n^{(a)})$ and to calculate the sample covariance matrices $\mathbf{S}^*$ of the bootstrap samples. Interesting statistics will be generated when fitting the substantive model to $\mathbf{S}^*$. Of course, the form of $\mathbf{S}_a$ is decided by the purpose of the study. When studying the behavior of model statistics under the

null hypothesis $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, one needs to choose $\mathbf{S}_a = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ for an admissible $\hat{\boldsymbol{\theta}}$ (Beran & Srivastava, 1985; Bollen & Stine, 1993). When studying the power of a statistic one needs to choose a $\mathbf{S}_a$ to represent the interesting alternative hypothesis (Beran, 1986; Yuan & Hayashi, 2003; Yuan & Marshall, 2004).

Different $\mathbf{S}_a$'s correspond to different populations. Because any interesting model in practice is inevitably misspecified, there is a lot of interest in the property of test statistics with misspecified models. In particular, the population covariance matrix $\boldsymbol{\Sigma}_0 = E(\mathbf{S})$ is generally unknown while the behavior of a test statistic at $\boldsymbol{\Sigma}_0$ is of fundamental interest. It would be nice if a $\mathbf{S}_a$ that is close to $\boldsymbol{\Sigma}_0$ can be obtained. With unavoidable model misspecification, formal or informal cutoff values have been established to judge a model based on fit indices. However, the distribution of almost all the fit indices are generally unknown (Yuan, 2005), and bootstrap can be used to study their distributions or confidence intervals. In such a study, alternative covariance matrices are needed. For example, if one needs to understand the behavior of the sample RMSEA (Steiger & Lind, 1980) when its population value is 0.05 or 0.08, alternative covariance matrices are needed to achieve the desired population values.

Parallel to obtaining a $\mathbf{S}_a$ that plays the role of $\boldsymbol{\Sigma}_0$ in the bootstrap simulation, obtaining a covariance matrix that is at a given distance from a model structure is of interest in Monte Carlo studies, where there is generally no particular $\boldsymbol{\Sigma}_0$ to consider. In that direction, Satorra and Saris (1985) obtained a $\boldsymbol{\Sigma}$ through setting a fixed parameter in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ at an incorrect value. Cudeck and Browne (1992) provided a procedure for obtaining a $\boldsymbol{\Sigma}$ that is at a given distance from a model structure. Curran, West and Finch (1996) generated a $\boldsymbol{\Sigma}$ that contains loadings in the population but not in the model. Yuan and Bentler (1997) generated a $\boldsymbol{\Sigma}$ by a structure with three factors while the null hypothesis is a two-factor model. Fouladi (2000) created a $\boldsymbol{\Sigma}$ by properly perturbing the structured covariance matrix. These approaches of generating $\boldsymbol{\Sigma}$'s might also be used in constructing a population covariance matrix in the bootstrap simulation, but the resulting $\boldsymbol{\Sigma}$'s are generally not as desired as the one to be described in this paper.

We will provide a class of $\mathbf{S}_a$'s by combining empirical information represented by $\mathbf{S}$ and substantive theory represented by $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. In section 2 of this paper we give the form of the

$\mathbf{S}_a$ and study its property as an alternative covariance matrix. The details leading to these properties are given in the appendix. In section 3, we illustrate how to obtain a particular $\mathbf{S}_a$ using practical data. We conclude this paper with a discussion of issues related to the proposed $\mathbf{S}_a$ and its applications in other contexts.

2. A Class of Covariance Matrices for Bootstrap Simulation

This section proposes a class of $\mathbf{S}_a$'s and studies their analytical properties. These properties justify their ideal candidacy for the role of the population covariance matrix in a bootstrap simulation. To better understand these properties we give a brief review of the role of $\mathbf{\Sigma}_0 = E(\mathbf{S})$ in characterizing the distribution of the normal theory likelihood ratio statistic.

Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ be a sample from a p-dimensional population $\mathbf{x}$ with $E(\mathbf{x}) = \boldsymbol{\mu}_0$ and $\mathrm{Cov}(\mathbf{x}) = \mathbf{\Sigma}_0$. CSA involves fitting the sample covariance matrix $\mathbf{S}$ to a structural model represented by $\mathbf{\Sigma}(\boldsymbol{\theta}) = (\sigma_{ij}(\boldsymbol{\theta}))$, such that a discrepancy function $F(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta}))$ is minimized. Several discrepancy functions exist and the normal theory based likelihood function

$$F_{ML}(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta})) = \mathrm{tr}[\mathbf{S}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})] - \ln|\mathbf{S}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})| - p \tag{2}$$

is most commonly used in practice. Let $\hat{\boldsymbol{\theta}}$ be the minimizer of $F_{ML}(\mathbf{S}, \mathbf{\Sigma}(\boldsymbol{\theta}))$ and

$$T_{ML} = nF_{ML}(\mathbf{S}, \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})).$$

When data are normally distributed and the model structure is correct,

$$T_{ML} \xrightarrow{\mathcal{L}} \chi^2_{df},$$

where $df = p(p+1)/2 - q$ and q is the number of free parameters in $\boldsymbol{\theta}$. Because test statistics are almost always significant with empirical data, there is a great deal of interest in measuring how far the model $\mathbf{\Sigma}(\boldsymbol{\theta})$ is from $\mathbf{\Sigma}_0$, as characterized by

$$\tau = F_{ML}(\mathbf{\Sigma}_0, \mathbf{\Sigma}(\boldsymbol{\theta}_*)), \tag{3}$$

where $\boldsymbol{\theta}_*$ minimizes $F_{ML}(\mathbf{\Sigma}_0, \mathbf{\Sigma}(\boldsymbol{\theta}))$. Notice that the τ in (3) does not depend on the sample size, it is the distance from $\mathbf{\Sigma}(\boldsymbol{\theta})$ to the population covariance matrix $\mathbf{\Sigma}_0$. Under a sequence of

local alternatives and normally distributed data (see Satorra & Saris, 1985; Steiger, Shapiro, & Browne, 1985),

$$T_{ML} \xrightarrow{\mathcal{L}} \chi^2_{df}(\delta),$$

where $\delta = n\tau$. Thus

$$E(T_{ML}) \approx n\tau + df, \tag{4}$$

where the approximation sign is due to a finite sample size. When data are not normally distributed or when $\boldsymbol{\Sigma}_0$ is fixed, the result in (4) can be improved (Shapiro, 1983; Yuan et al., 2005). For a $p \times p$ matrix $\mathbf{A}$, let vech($\mathbf{A}$) be the vector that stacks the nonduplicated elements of $\mathbf{A}$ by leaving out those above the diagonal, and denote $\boldsymbol{\sigma}(\boldsymbol{\theta}) = \text{vech}[\boldsymbol{\Sigma}(\boldsymbol{\theta})]$. Let $\mathbf{D}_p$ be the duplication matrix as defined by Magnus and Neudecker (1999, p. 49),

$$\mathbf{W}_0 = \frac{1}{2}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p, \quad \mathbf{W}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{D}_p'[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})]\mathbf{D}_p,$$

$$\boldsymbol{\Sigma}_* = \boldsymbol{\Sigma}(\boldsymbol{\theta}_*), \quad \mathbf{W}_* = \mathbf{W}(\boldsymbol{\theta}_*), \quad \dot{\boldsymbol{\sigma}}_* = \partial\boldsymbol{\sigma}(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}_*', \quad \boldsymbol{\Delta}_{ij*} = \partial^2\sigma_{ij}(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}_*\partial\boldsymbol{\theta}_*',$$

$$\mathbf{G} = \boldsymbol{\Sigma}_*^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_*^{-1} - \frac{1}{2}\boldsymbol{\Sigma}_*^{-1}, \quad \mathbf{H} = (h_{ij}) = \boldsymbol{\Sigma}_* - \boldsymbol{\Sigma}_*^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_*^{-1},$$

$$\mathbf{M} = \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} h_{ij}\boldsymbol{\Delta}_{ij*} \text{ and } \boldsymbol{\Pi} = \text{Cov}\{\text{vech}[(\mathbf{x} - \boldsymbol{\mu}_0)(\mathbf{x} - \boldsymbol{\mu}_0)']\}.$$

Then (4) should be replaced by

$$E(T_{ML}) \approx n\tau + \text{tr}(\mathbf{U}\boldsymbol{\Pi}), \tag{5}$$

where

$$\mathbf{U} = \mathbf{W}_0 - \mathbf{W}_*\dot{\boldsymbol{\sigma}}_*[\dot{\boldsymbol{\sigma}}_*'\mathbf{D}_p'(\mathbf{G} \otimes \boldsymbol{\Sigma}_*^{-1})\mathbf{D}_p\dot{\boldsymbol{\sigma}}_* + \mathbf{M}]^{-1}\dot{\boldsymbol{\sigma}}_*'\mathbf{W}_*.$$

When $\mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, then $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ and $\text{tr}(\mathbf{U}\boldsymbol{\Pi}) = df$. When $\mathbf{x}$ has heavier tails than that of $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, $\text{tr}(\mathbf{U}\boldsymbol{\Pi}) > df$. But the behavior of the distribution of T_{ML} at a finite sample size, especially when the distribution of $\mathbf{x}$ is unknown, is not well-described by its asymptotic distribution. Actually, both the empirical mean and variance of T_{ML} are closely related to $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, as well as the distribution shape of $\mathbf{x}$ (Yuan et al., 2005).

When using the bootstrap to study the behavior of T_{ML}, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is already known. The distribution shape of $\mathbf{x}$ is approximated by the histogram decided by $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, one

probably cannot improve the approximation before further information of the distribution of $\mathbf{x}$ is known. So the quality of a bootstrap study is up to finding a substitute $\mathbf{S}_a$ for $\boldsymbol{\Sigma}_0$. The closer the $\mathbf{S}_a$ to $\boldsymbol{\Sigma}_0$ the more valid the bootstrap conclusion regarding T_{ML} is. The remainder of this section concentrates on finding a $\mathbf{S}_a$ that combines empirical information represented by $\mathbf{S}$ and substantive theory represented by $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. When the interest is in finding a $\boldsymbol{\Sigma}_0$ corresponding to a particular τ in (3), (4) and (5) facilitate the selection of a proper $\mathbf{S}_a$ to play the role of $\boldsymbol{\Sigma}_0$, as will be illustrated in the next section.

When $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is inadequate in modeling $\boldsymbol{\Sigma}_0$, it follows from (4) and (5) that the τ defined in (3) is most likely somewhere between 0 and T_{ML}/n. It is possible for $\tau > T_{ML}/n$ with gross sampling errors, but the probability for T_{ML}/n to be smaller than τ is tiny unless both n and *df* are rather small. For $\tau \in [0, T_{ML}/n]$, there is a simple way to search for a substitute $\mathbf{S}_a$ of $\boldsymbol{\Sigma}_0$ that satisfies (3). The following development uses the normal theory based discrepancy function (2) for finding such a substitute. Using other discrepancy functions will be discussed at the end of this paper.

Let

$$\mathbf{S}_a = a\mathbf{S} + (1-a)\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \quad a \in [0,1] \tag{6}$$

where $\hat{\boldsymbol{\theta}}$ is the minimizer of equation (2). The sample covariance matrix $\mathbf{S}$ represents current data. The covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a theoretical hypothesis, which generally represents prior information about the population covariance matrix $\boldsymbol{\Sigma}_0$. This prior information is usually accumulated through previous research. For example, in a confirmatory factor analysis the factor structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ may be based on an exploratory factor analysis of a previous sample. The $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is an estimator of the prior information using current data. Consequently, (6) is an empirical Bayes estimator of the population covariance matrix $\boldsymbol{\Sigma}_0$ (Haff, 1980; Hayashi & Sen, 2002; Ledoit & Wolf, 2004). Choosing $\mathbf{S}_a$ in the form of (6) not only incorporates the empirical Bayes point of view but also simplifies the model fitting procedure. Specifically, when $\hat{\boldsymbol{\theta}}$ is the minimizer of (2), we will show that for any given $a \in [0,1]$ the function $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ also attains its minimum at $\hat{\boldsymbol{\theta}}$. Furthermore,

$$g(a) = F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$$

is a strictly increasing function of a with $g(0) = 0$ and $g(1) = T_{ML}/n$. Below we will formally

establish the properties of $\mathbf{S}_a$ and $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. We will illustrate how to choose a in the next section.

For a given $a \in [0, 1]$, let $\hat{\boldsymbol{\theta}}_a$ minimize $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, then it satisfies the normal estimating equation associated with this minimization

$$\dot{\boldsymbol{\sigma}}'(\hat{\boldsymbol{\theta}}_a)\mathbf{W}(\hat{\boldsymbol{\theta}}_a)[\mathbf{S}_a - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}}_a)] = \mathbf{0}, \tag{7}$$

where $\mathbf{S}_a = \text{vech}(\mathbf{S}_a)$. Let $\mathbf{s} = \text{vech}(\mathbf{S})$. Because $\mathbf{S}_a = a\mathbf{s} + (1-a)\boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$ satisfies

$$\dot{\boldsymbol{\sigma}}'(\hat{\boldsymbol{\theta}})\mathbf{W}(\hat{\boldsymbol{\theta}})[\mathbf{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})] = \mathbf{0},$$

$\hat{\boldsymbol{\theta}}_a = \hat{\boldsymbol{\theta}}$ also satisfies equation (7). Thus, $\hat{\boldsymbol{\theta}}$ is a stationary point of the function $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. A question that remains is whether $\hat{\boldsymbol{\theta}}$ also minimizes $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. The answer is given by the following theorem.

Theorem 1. For any given $\mathbf{S}$ and $a \in [0, 1]$, let $\mathbf{S}_a$ be given by (6). If $\hat{\boldsymbol{\theta}}$ is the vector that minimizes $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, then $\hat{\boldsymbol{\theta}}$ is the minimizer of the function $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

The above theorem depends on $\hat{\boldsymbol{\theta}}$ minimizing (2). When $\hat{\boldsymbol{\theta}}$ is not the global minimizer of (2) but only a local minimizer, then $\hat{\boldsymbol{\theta}}$ is also a local minimizer of $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. So even when the alternative hypothesis changes, the parameter estimate is still the same when choosing $\mathbf{S}_a$ in the form of (6). When using $\mathbf{S}_a$ as the bootstrap population covariance matrix in (1), the population value of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}$, which does not depend on a.

With $\hat{\boldsymbol{\theta}}$ minimizing $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ for all a we may wonder if the function $g(a)$ is also a constant. The next theorem clarifies this concern.

Theorem 2. The function $g(a)$ is a strictly increasing function on $[0, 1]$ unless $\mathbf{S} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$.

Note that $F_{ML}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ contains $p(p+1)/2 + q$ unknown elements, finding a $\boldsymbol{\Sigma}_0$ for a given τ is an optimization problem. Theorem 2 greatly simplifies the problem when constructing a $\mathbf{S}_a$ in the form of (6). Together with Theorem 1, searching for $\mathbf{S}_a$ becomes a line-search problem, and for any given $\tau \in [0, T_{ML}/n]$, there is a unique a such that $\mathbf{S}_a$ satisfies $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) = \tau$. Of course, one generally does not need to search for $\mathbf{S}_a$ such that $g(a) > T_{ML}/n$. If such a case is needed, then one cannot construct $\mathbf{S}_a$ through (6). Actually, even if one chooses $\mathbf{S}_a = a\mathbf{S}$ and a is large enough, $g(a) > T_{ML}/n$ may still not be reached. This can be illustrated through the class of LISREL models (Jöreskog & Sörbom, 1996).

In the LISREL setup, the measurement model that relates hypothetical latent variables $(\boldsymbol{\xi}, \boldsymbol{\eta})$ to their measured indicators $(\mathbf{x}, \mathbf{y})$ is

$$\mathbf{x} = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta}, \quad \mathbf{y} = \boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{8a}$$

where $\boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{x})$, $\boldsymbol{\mu}_{\mathbf{y}} = E(\mathbf{y})$, $E(\boldsymbol{\xi}) = \mathbf{0}$, $E(\boldsymbol{\eta}) = \mathbf{0}$, $\boldsymbol{\Lambda}_{\mathbf{x}}$ and $\boldsymbol{\Lambda}_{\mathbf{y}}$ are factor loading matrices, and $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are measurement errors with $\boldsymbol{\Theta}_{\boldsymbol{\delta}} = \mathrm{Cov}(\boldsymbol{\delta})$, $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} = \mathrm{Cov}(\boldsymbol{\varepsilon})$. The structural model that describes interrelations among latent variables is

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{8b}$$

where $\boldsymbol{\zeta}$ is a vector of prediction errors having a covariance matrix $\boldsymbol{\Psi} = \mathrm{Cov}(\boldsymbol{\zeta})$. Let $\boldsymbol{\Phi} = \mathrm{Cov}(\boldsymbol{\xi})$, the resulting covariance structure of the observed variables $(\mathbf{x}', \mathbf{y}')'$ is (see Jöreskog & Sörbom, 1996, p. 3)

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{\mathbf{x}}' + \boldsymbol{\Theta}_{\boldsymbol{\delta}} & \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\Phi}\boldsymbol{\Gamma}'(\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}' \\ \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_{\mathbf{x}}' & \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}' + \boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} \end{pmatrix}. \tag{9}$$

In order for model (8) or (9) to be identified, at a minimum we have to fix the scales of the latent variables. There are two ways to fix these scales; one is to fix the variance of each exogenous latent construct at a given value and the other is to fix a path loading from each latent construct to one of its indicators. We will choose the latter approach, that is, a factor loading in $\boldsymbol{\Lambda}_{\mathbf{x}}$ or $\boldsymbol{\Lambda}_{\mathbf{y}}$ is fixed for each of the latent constructs. Enough zero elements will be specified in applications to identify the entire model. It is easy to see that (9) satisfies

$$a\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{x}}(a\boldsymbol{\Phi})\boldsymbol{\Lambda}_{\mathbf{x}}' + a\boldsymbol{\Theta}_{\boldsymbol{\delta}} & \boldsymbol{\Lambda}_{\mathbf{x}}(a\boldsymbol{\Phi})\boldsymbol{\Gamma}'(\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}' \\ \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}(a\boldsymbol{\Phi})\boldsymbol{\Lambda}_{\mathbf{x}}' & \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\Gamma}(a\boldsymbol{\Phi})\boldsymbol{\Gamma}' + a\boldsymbol{\Psi}](\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}_{\mathbf{y}}' + a\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}} \end{pmatrix}. \tag{10}$$

Because $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is a function of $\mathbf{S}_a\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$, unless $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}_{\boldsymbol{\delta}}$ or $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}$ contain at least one fixed nonzero element, using $\mathbf{S}_a = a\mathbf{S}$ and $\mathbf{S}_a = \mathbf{S}$ results in the same $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$. However, the parameter estimates in these matrices corresponding to $a\mathbf{S}$ will be proportional to those corresponding to $\mathbf{S}$. Equation (10) is closely related to the invariance with a constant scaling factor given in Browne (1984). On the other hand, if fixed nonzero elements in $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}_{\boldsymbol{\delta}}$, or $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon}}$ exist, then $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$ will increase as the a in $\mathbf{S}_a = a\mathbf{S}$ increases. However, most practical models generally do not contain fixed nonzero elements beyond those used for identifying the scales of latent variables.

More generally, when one chooses $\mathbf{S}_{\alpha\kappa} = \alpha\mathbf{S} + \kappa\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ for positive numbers α and κ, $\mathbf{S}_{\alpha\kappa}$ can be rewritten as

$$\mathbf{S}_{\alpha\kappa} = b[a\mathbf{S} + (1-a)\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})] = b\mathbf{S}_a,$$

where $b = (\alpha+\kappa)$ and $a = \alpha/(\alpha+\kappa)$. Suppose the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is represented by (9) and there are no fixed nonzero elements beyond those for identification purposes. It follows from (10) that, corresponding to each $\hat{\boldsymbol{\theta}}$ minimizing $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, there always exists a $\hat{\boldsymbol{\theta}}^*$ that minimizes $F_{ML}(\mathbf{S}_{\alpha\kappa}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ with the same minimum as that of $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. This implies that it is impossible to find a $\mathbf{S}_{\alpha\kappa}$ to reach a $\tau > T_{ML}/n$. Actually, a $\mathbf{S}_a$ corresponding to a $\tau > T_{ML}/n$ sits most likely farther than τ away from the model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

Theorems 1 and 2 showed that all the alternative hypotheses in the form of (6) share the same parameter estimate while the corresponding model fit, as measured by F_{ML} or T_{ML}, deteriorates as a increases. The same phenomenon occurs in CSA and can also occur in other statistical models. We have the following two simple examples to facilitate a better understanding of the two theorems.

Example 1. Let $p = 2$ and consider the classical spherical model $\boldsymbol{\Sigma}(\theta) = \theta\mathbf{I}_2$. We have

$$F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\theta)) = \frac{1}{\theta}\mathrm{tr}(\mathbf{S}) + 2\log\theta - \log|\mathbf{S}| - 2.$$

Let $\mathbf{S} = (s_{ij})$, it is easy to see that $\hat{\theta} = (s_{11} + s_{22})/2$ minimizes $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\theta))$. It follows from

$$\begin{aligned} F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\theta)) &= \frac{1}{\theta}\mathrm{tr}(\mathbf{S}) + 2\log\theta - \log|\mathbf{S}_a| - 2 \\ &= F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\theta)) + \log|\mathbf{S}| - \log|\mathbf{S}_a| \end{aligned}$$

that $\hat{\theta}$ also minimizes $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\theta))$ for any a. Because

$$|\mathbf{S}_a| = \frac{1}{4}\{(s_{11} + s_{22}) - a^2[(s_{11} - s_{22})^2 + s_{12}^2]\}$$

decreases with a, $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\theta}))$ is a strictly increasing function of a unless $s_{11} = s_{22}$ and $s_{12} = 0$, i.e., $\mathbf{S} = \boldsymbol{\Sigma}(\hat{\theta})$.

Example 2. Let $(\mathbf{x}_1', y_1)$, $(\mathbf{x}_2', y_2)$, $\cdots$, $(\mathbf{x}_n', y_n)$ be a sample for which the regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{11}$$

is appropriate, where $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)'$ is a $n \times k$ matrix, $\boldsymbol{\beta}$ is a $k \times 1$ vector, $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$.

Then the least squares estimates of $\boldsymbol{\beta}$ and σ^2 are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

and

$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-k).$$

Let $\mathcal{R}(\mathbf{X})$ be the space spanned by the columns of $\mathbf{X}$, and $\mathbf{X}^{\perp}$ be a $n \times (n-k)$ matrix whose columns form the basis of the space orthogonal to $\mathcal{R}(\mathbf{X})$. Instead of using the vector $\mathbf{y}$ in (11), we use

$$\mathbf{y}_a = \mathbf{y} + \mathbf{X}^{\perp}\mathbf{a}, \tag{12}$$

where $\mathbf{a}$ is a $(n-k) \times 1$ vector. When fitting $\mathbf{y}_a$ by (11), the estimates of $\boldsymbol{\beta}$ and σ^2 corresponding to $\mathbf{y}_a$ are

$$\hat{\boldsymbol{\beta}}_a = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_a$$

and

$$\hat{\sigma}_a^2 = (\mathbf{y}_a - \mathbf{X}\hat{\boldsymbol{\beta}}_a)'(\mathbf{y}_a - \mathbf{X}\hat{\boldsymbol{\beta}}_a)/(n-k).$$

It is easy to see that $\hat{\boldsymbol{\beta}}_a = \hat{\boldsymbol{\beta}}$ while $\hat{\sigma}_a^2$ does not equal $\hat{\sigma}^2$ in general. To compare their systematic difference, we have

$$E(\hat{\sigma}^2) = \sigma^2$$

and

$$E(\hat{\sigma}_a^2) = \sigma^2 + \mathbf{a}'(\mathbf{X}^{\perp\prime}\mathbf{X}^{\perp})\mathbf{a}/(n-k).$$

So for any $\mathbf{a} \neq \mathbf{0}$, $E(\hat{\sigma}_a^2) > \sigma^2$. While the parameter estimate $\hat{\boldsymbol{\beta}}_a$ remains the same, the residual sum of squares are different. If letting $\mathbf{a} = a\mathbf{b}$ with $\mathbf{b}$ being a fixed vector, then $E(\hat{\sigma}_a^2)$ will increase with a.

In the context of covariance structure analysis, the model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ corresponds to $\mathbf{X}\boldsymbol{\beta}$ in (11). The data matrices $\mathbf{S}$ and $\mathbf{S}_a$ correspond to the data vectors $\mathbf{y}$ in (11) and $\mathbf{y}_a$ in (12), respectively. The transformation (1) changes the observed data in a similar way as in (12). Actually, mean and covariance structure analysis can be formulated as a nonlinear regression

model (Browne, 1982; Lee & Jennrich, 1984; Yuan & Bentler, 1997). Unfortunately, in the context of nonlinear models there does not exist a clean formula as in the context of linear regression models.

3. Illustrations

We will illustrate how to choose a proper $\mathbf{S}_a$ in (6). There are two circumstances that particular covariance matrices are needed. One is to study the behavior of T_{ML} or a fit index for a given sample under certain degree of misspecification. The degree of misspecification is typically quantified by the τ in (3). The other is to study the behavior of T_{ML} or a fit index for a given sample when $\boldsymbol{\Sigma}_0$ or the τ in (3) is unknown. Because τ can only be estimated, we can only obtain a substitute for $\boldsymbol{\Sigma}_0$ in the later case. We will illustrate the two cases by the following example.

Example 3. Holzinger and Swineford (1939) provided a data set consisting of 24 test scores from 145 seventh- and eighth-graders. Jöreskog (1969) used 9 of the 24 variables and studied their correlation structures with the normal theory ML method. We will also use these 9 variables for our illustration. The 9 variables are: Visual Perception, Cubes, Lozenges, Paragraph Comprehension, Sentence Completion, Word Meaning, Addition, Counting Dots, and Straight-Curved Capitals. In the original design of Holzinger and Swineford's study, the first three variables were to measure spatial ability, the next three variables were to measure verbal ability, and the last three variables were tested with a limited time period to measure a speed factor in performing the tests. Let $\mathbf{x}$ be the vector of the 9 observed variables, then the confirmatory factor model

$$\mathbf{x} = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{e} \quad \text{and} \quad \text{Cov}(\mathbf{x}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \tag{13a}$$

with

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{42} & \lambda_{52} & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{73} & \lambda_{83} & \lambda_{93} \end{pmatrix}', \quad \boldsymbol{\Phi} = \begin{pmatrix} 1.0 & \phi_{12} & \phi_{13} \\ \phi_{21} & 1.0 & \phi_{23} \\ \phi_{31} & \phi_{32} & 1.0 \end{pmatrix}, \tag{13b}$$

represents Holzinger and Swineford's hypothesis. We assume that the elements of the $\mathbf{e}$ are uncorrelated with $\boldsymbol{\Psi} = \text{Cov}(\mathbf{e})$ being a diagonal matrix. There are $q = 21$ unknown parameters in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, and the model degrees of freedom are 24.

For Holzinger and Swineford's data and model (13), suppose one has the interest to use the bootstrap to study the behavior of T_{ML} for certain τ. We can easily find a bootstrap population covariance matrix $\mathbf{S}_a$ in the form of (6) that satisfies $\tau = F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$. Interesting values of τ's can be decided by the cutoff values of many fit indices (see Hu & Bentler, 1999). A commonly used one is

$$\text{RMSEA} = \sqrt{\tau/df} \;\text{ or }\; \tau = df \times (\text{RMSEA})^2.$$

The τ's corresponding to the cutoff values of RMSEA = 0.05, 0.08, and 0.10 are 0.06, 0.1536, and 0.24, respectively. To find a proper a for $\tau_0 = 0.06$, we may start with $a = 0.5$, which results in $\tau = 0.0846284 > \tau_0$. A smaller a is necessary, the following are our successive selections of a according to the value of τ obtained in the previous step.

a	$F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$
0.5	0.0846284
0.25	0.0210088
0.4	0.0539408
0.45	0.0683939
0.425	0.0609466
0.422	0.0600826
0.4215	0.0599393
0.4216	0.0599679
0.4217	0.0599966
0.42171	0.0599994
0.421712	0.0600000

The $\mathbf{S}_a$ corresponding to $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) = 0.0600000$ is obtained at $a = 0.421712$, which agrees with the target $\tau_0 = 0.06$ down to the 7th decimal place. It is obvious that a a with many decimals is necessary in order for $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$ to be more close to τ_0. Using the same procedures, we also have $a = 0.6699$ corresponding to $\tau = 0.1536004$, agreeing with the target $\tau_0 = 0.1536$ down to the 6th decimal place; $a = 0.83089$ corresponding to $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) = 0.2400000$, agreeing with the target $\tau_0 = 0.24$ down to the 7th decimal place. Whichever $\mathbf{S}_a$ in the form of (6) is used as the bootstrap population covariance matrix in (1) the population value of $\boldsymbol{\theta}$ remains $\hat{\boldsymbol{\theta}}$.

When the interest is in the behavior of T_{ML} with the unknown population $\boldsymbol{\Sigma}_0$, we can also approximate the $\boldsymbol{\Sigma}_0$ by $\mathbf{S}_a$. The normal theory likelihood ratio statistic for model (13)

is $T_{ML} = 51.543$ and is highly significant when referring to χ^2_{24}. If assuming $T_{ML} \sim \chi^2_{24}(\delta)$ with $\delta = n\tau$ and $\tau = F_{ML}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}^*))$, then the commonly used estimate of τ is given by $\hat{\tau}_1 = (T_{ML} - 24)/n = 0.1899509$. Using essentially the same procedure for finding a $\mathbf{S}_a$ corresponding to a given RMSEA, we have $a = 0.74259$ corresponding to $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) = 0.1899506$, which agrees with $\hat{\tau}_1$ down to the 6th decimal place.

The statistic T_{ML} is derived from $\mathbf{x} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. Other test statistics that do not assume the normal distribution of the data also imply that the model does not fit the data well. It is more natural to regard the $\boldsymbol{\Sigma}_0$ as a fixed covariance matrix for Holzinger and Swineford's data rather than changing with n. Thus, an estimate of τ based on (5) should be more reasonable. Using $\mathbf{S}$ to estimate $\boldsymbol{\Sigma}_0$, $\hat{\boldsymbol{\theta}}$ to estimate $\boldsymbol{\theta}_*$, and denoting $\mathbf{y}_i = \text{vech}[(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})']$ and using the sample covariance matrix $\mathbf{S}_\mathbf{y}$ of $\mathbf{y}_i$ to estimate $\boldsymbol{\Pi}$, we have $\text{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Pi}}) = 22.873261$ and $\hat{\tau}_2 = [T_{ML} - \text{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Pi}})]/n = 0.1977215$. Now we can approximate $\boldsymbol{\Sigma}_0$ by $\mathbf{S}_a$ in (6). The $a = 0.757098$ corresponds to $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) = 0.1977214$, agreeing with $\hat{\tau}_2$ to the 6th decimal place.

Once τ or $\hat{\tau}$ is known, the search for $\mathbf{S}_a$, as illustrated above, can be performed using any standard SEM program (e.g., Bentler, 2006) that allows the printing of $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$ and $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ down to several decimals. The calculation of $\mathbf{S}_a$ in (6) can be done by any software that contains the functions of matrix addition and multiplication (e.g., SAS IML, Matlab, Splus). The estimation of τ based on (5) involves second derivatives of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, which is usually quite complicated and not any SEM program provides its calculation at present.

4. Conclusion and Discussion

A covariance structure model in practice is, at best, only an approximation to $\boldsymbol{\Sigma}_0$, there should be a great interest in studying the behavior of a statistic when $\tau > 0$ (see MacCallum, Browne & Sugawara, 1996). With unknown distributions of $\mathbf{x}$ in practice, the bootstrap remains a valuable tool for such studies. Because the behavior of any statistic T is closely related to $\boldsymbol{\Sigma}_0$, when $\boldsymbol{\Sigma}_0$ is unknown, finding a $\mathbf{S}_a$ that approximately equals $\boldsymbol{\Sigma}_0$ allows us to obtain more accurate evaluation of T_{ML}. The purpose of the paper is to provide the class of covariance matrices represented by $\mathbf{S}_a = a\mathbf{S} + (1-a)\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$. The nice properties of $\mathbf{S}_a$

make it an ideal choice for playing the role of the bootstrap population covariance matrix. Example 3 shows that it is straightforward to find a proper $\mathbf{S}_a$ once the τ or its estimate is given. Specific applications of $\mathbf{S}_a$ in bootstrap simulation were given in Yuan and Hayashi (2003).

When using $\mathbf{S}_a$ to approximate the unknown $\boldsymbol{\Sigma}_0$, the quality of the approximation depends on $\mathbf{S}$ as well as the goodness of the model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. When the sample size n is small or when the distribution of $\mathbf{x}$ is of heavy tails, $\mathbf{S}$ may not be a good estimate of $\boldsymbol{\Sigma}_0$ even it is unbiased, the $\hat{\tau}_1$ or $\hat{\tau}_2$ in the previous section can be a poor estimate of τ. It is possible that $\hat{\tau}_1$ or $\hat{\tau}_2$ may be negative even when $\tau > 0$, then we might have to estimate $\boldsymbol{\Sigma}_0$ by $\mathbf{S}_0 = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$. Although the form of $\mathbf{S}_a$ in (6) combines both theory and empirical data, it can still be quite different from the true $\boldsymbol{\Sigma}_0$. But such obtained $\mathbf{S}_a$, especially the one corresponding to $\hat{\tau}_2$, should be more close to $\boldsymbol{\Sigma}_0$ than those obtained previously in the literature where no effort was made in approximating the true $\boldsymbol{\Sigma}_0$. Of course, when $\boldsymbol{\Sigma}_0$ is known or a better estimator of $\boldsymbol{\Sigma}_0$ than that given in (6) is available, then such a covariance matrix should be used in the transformation (1) for more accurate bootstrap inference.

We have focused on using the normal theory based discrepancy function (2) to establish the properties of $\mathbf{S}_a$ in section 2. One may wonder if similar properties can be established for other types of discrepancy functions. In covariance structure analysis, a very general form of discrepancy function is (see Shapiro, 1985)

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]' \mathbf{W} [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]$$

for a proper weight matrix $\mathbf{W}$. It can be verified directly that, when $\mathbf{W}$ is only a function of $\boldsymbol{\theta}$ or a constant matrix, $F(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ will enjoy the same properties of $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ as in Theorems 1 and 2. Special cases are the normal theory based likelihood discrepancy function and the least squares discrepancy function. However, when $\mathbf{W}$ involves $\mathbf{x}_i^{(a)}$ the properties in Theorems 1 and 2 may no longer apply to $F(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. A special case for this is when $\mathbf{W} = (\mathbf{S}_{\mathbf{y}}^{(a)})^{-1}$, where $\mathbf{S}_{\mathbf{y}}^{(a)}$ is the sample covariance matrix of $\mathbf{y}_i^{(a)} = \text{vech}[(\mathbf{x}_i^{(a)} - \bar{\mathbf{x}}_a)(\mathbf{x}_i^{(a)} - \bar{\mathbf{x}}_a)']$ with $\bar{\mathbf{x}}_a = \sum_{i=1}^n \mathbf{x}_i^{(a)}/n$. Comparing several commonly used discrepancy functions for power analysis, Yuan and Hayashi (2003) showed that, with a proper downweighting of heavy tails, the bootstrap based on the normal theory discrepancy function in (2) leads to the most

powerful test for covariance structure analysis.

Because a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ represents prior information about the population covariance matrix, we propose to use the $\mathbf{S}_a$ in (6) to estimate the population covariance matrix. When no such information is available, the form $\mathbf{S}_{ab} = a\mathbf{S} + b\mathbf{I}$ has been shown to enjoy some nice properties with respect to several loss functions (Efron & Morris, 1976; Haff, 1980). It is also interesting to study properties of $\mathbf{S}_a$ with respect to some loss functions and to find a proper prior distribution to formally justify its Bayes nature.

Standard errors in CSA are also affected by model misspecification (Yuan & Hayashi, in press). Because different $\mathbf{S}_a$'s correspond to the same $\hat{\boldsymbol{\theta}}$, choosing a bootstrap population covariance matrix in the form of $\mathbf{S}_a$ in (6) will facilitate the study of standard error changes with varying degree of model misspecification.

We have mainly discussed the candidacy of $\mathbf{S}_a$ in playing the role of the bootstrap population covariance matrix. It can also be used as the population covariance matrix in Monte Carlo studies. Since the behavior of a statistic is closely related to $\boldsymbol{\Sigma}_0$ (Yuan et al., 2005), a $\boldsymbol{\Sigma}_0$ that reflects the reality will enhance the validity of the conclusion of a Monte Carlo study. Most previous Monte Carlo studies constructed $\boldsymbol{\Sigma}_0$ according to certain structures while the models are created by omitting or fixing a subset of parameters in $\boldsymbol{\theta}$. Such created model misspecifications might be interesting but may not be realistic. The $\mathbf{S}_a$ in (6) should better reflect the fact that "All models are wrong but some are useful." (Box, 1979).

Appendix

Proof of Theorem 1: It is obvious that $\hat{\boldsymbol{\theta}}$ minimizes $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ when $a = 0$ or 1. For a general $a \in [0, 1]$ we can rewrite $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ as

$$\begin{aligned} F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta})) &= a\mathrm{tr}[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] + (1-a)\mathrm{tr}[\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] - \ln|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})| - p \\ &\quad + \ln[|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})|/|a\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) + (1-a)\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})|] \\ &= a\mathrm{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] + F_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\boldsymbol{\theta})) \\ &\quad + \ln[|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})|/|a\mathbf{S} + (1-a)\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})|]. \end{aligned} \tag{A1}$$

Since $\hat{\boldsymbol{\theta}}$ minimizes $F_{ML}(\mathbf{S}_a, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ when $a = 1$, we have

$$\begin{aligned}\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})] + \ln[|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})|/|\mathbf{S}|] &\leq \text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] \\ &\quad + F_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\boldsymbol{\theta})) + \ln[|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})|/|\mathbf{S}|]\end{aligned}$$

for any $\boldsymbol{\theta}$, which further leads to

$$\begin{aligned}a\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})] &\leq a\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] + aF_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\boldsymbol{\theta})) \\ &\leq a\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] + F_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\boldsymbol{\theta})).\end{aligned}$$

Thus,

$$a\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})] + F_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})) \leq a\text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] + F_{ML}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \boldsymbol{\Sigma}(\boldsymbol{\theta})). \tag{A2}$$

Notice that the last term in (A1) does not involve $\boldsymbol{\theta}$; the minimization of (A1) only involves the previous two terms. The theorem is a result of (A2).

Proof of Theorem 2: Let the eigenvalues of $\mathbf{S}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})$ be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$, then we can rewrite $g(a)$ as

$$g(a) = a\sum_{i=1}^{p}\lambda_i + (1-a)p - \sum_{i=1}^{p}\ln[(1-a) + a\lambda_i] - p.$$

The derivative of $g(a)$ is

$$\begin{aligned}\dot{g}(a) &= \sum_{i=1}^{p}\lambda_i - p - \sum_{i=1}^{p}\frac{(\lambda_i - 1)}{(1-a) + a\lambda_i} \\ &= a\sum_{i=1}^{p}\frac{(\lambda_i - 1)^2}{(1-a) + a\lambda_i} \geq 0\end{aligned}$$

for $a \in [0, 1]$. The equality sign holds only when $\lambda_1 = \lambda_2 = \cdots = \lambda_p = 1$, which happens only when $\mathbf{S} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$. So $g(a)$ is a strictly increasing function of a.

References

Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Annals of Statistics, 18*, 1453–1463.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48*, 493-517.

Bentler, P. M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Dijkstra, T. K. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42). Amsterdam: North-Holland.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, directions. *Annual Review of Psychology, 47*, 563–592.

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34*, 181–197.

Beran, R. (1986). Simulated power functions. *Annals of Statistics, 14*, 151–173.

Beran, R., & Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics, 13*, 95–115.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605–634.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. In C. C. Clogg (Ed.), *Sociological methodology 1990* (pp. 115–140). Oxford: Basil Blackwell.

Bollen, K. A., & Stine, R. (1993). Bootstrapping goodness of fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Boomsma, A. (1986). On the use of bootstrap and jackknife in covariance structure analysis. *Compstat 1986*, 205–210.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). New York: Academic Press

Browne, M. W. (1982). Covariance structure analysis. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge: Cambridge University Press.

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

Browne, M. W., & Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology, 41*, 193–208.

Chatterjee, S. (1984). Variance estimation in factor analysis: An application of the bootstrap. *British Journal of Mathematical and Statistical Psychology, 37*, 252–262.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44*, 347–357.

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika, 57*, 357–369.

Curran, P. S., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26.

Efron, B., & Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Annals of Statistics, 4*, 22–32.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Enders, C.K. (2002). Applying the Bollen-Stine bootstrap for goodness-of-fit measures to structural equation models with missing data. *Multivariate Behavioral Research, 37*, 359–377.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling, 7*, 356–410.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics, 8*, 586–597.

Hall, P., & Titterington, D. M. (1989). The effect of simulation order on level accuracy and power of Monte Carlo tests. *Journal of the Royal Statistical Society. Series B, 51*, 459–467.

Hayashi, K., & Sen, P. K. (2002). Bias-corrected estimator of factor loadings in Bayesian factor analysis. *Educational and Psychological Measurement, 62*, 944–959.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution*. University of Chicago: Supplementary Educational Monographs, No. 48.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351–362.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor

analysis. *Psychometrika, 34*, 183–202.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

Kano, Y., Berkane, M., & Bentler, P. M. (1990). Covariance structure analysis with heterogeneous kurtosis parameters. *Biometrika, 77*, 575–585.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, 88*, 365–411.

Lee, S.-Y., & Jennrich, R. I. (1984). The analysis of structural equation models by means of derivative free nonlinear least squares. *Psychometrika, 49*, 521–528.

MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201–226.

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131–151.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Newbury Park, CA: Sage.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83–90.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal, 17*, 33–81.

Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to GLS estimators. *South African Statistical Journal, 19*, 73–81.

Shapiro, A., & Browne, M. W. (1987). Analysis of covariance structures under elliptical distributions. *Journal of the American Statistical Association, 82*, 1092–1097.

Steiger, J. H., & Lind, J. M. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society. Iowa City,

IA.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*, 253–264.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40*, 115–148.

Yuan, K.-H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, *92*, 767–774.

Yuan, K.-H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, *51*, 289–309.

Yuan, K.-H., & Bentler, P. M. (1999a). F-tests for mean and covariance structure analysis. *Journal of Educational and Behavioral Statistics*, *24*, 225–243.

Yuan, K.-H., & Bentler, P. M. (1999b). On normal theory and associated test statistics in covariance structure analysis under two classes of nonnormal distributions. *Statistica Sinica*, *9*, 831–853.

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology, 56*, 93–110.

Yuan, K.-H., & Hayashi, K. (in press). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology.*

Yuan, K.-H., Hayashi, K., & Bentler, P. M. (2005). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. Under review.

Yuan, K.-H., & Marshall, L. L. (2004). A new measure of misfit for covariance structure models. *Behaviormetrika, 31*, 67–90.

Yung, Y. F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Techniques and issues* (pp. 195–226). Hillsdale, New Jersey: Lawrence Erlbaum.