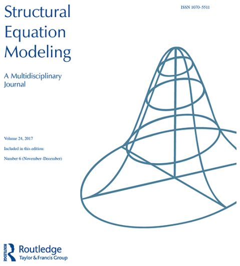
Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: http://www.tandfonline.com/loi/hsem20



Confidence Intervals of Fit Indexes by Inverting a Bootstrap Test

Chuchu Cheng & Hao Wu

To cite this article: Chuchu Cheng & Hao Wu (2017) Confidence Intervals of Fit Indexes by Inverting a Bootstrap Test, Structural Equation Modeling: A Multidisciplinary Journal, 24:6, 870-880, DOI: 10.1080/10705511.2017.1333432

To link to this article: https://doi.org/10.1080/10705511.2017.1333432

Structural Equation Modeling: A Multidisciplinary Journal, 24: 870–880, 2017
Copyright © Taylor & Francis Group, LLC
ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2017.1333432

Confidence Intervals of Fit Indexes by Inverting a Bootstrap Test

Chuchu Cheng and Hao Wu
Boston College

Fit indexes are an important tool in the evaluation of model fit in structural equation modeling (SEM). Currently, the newest confidence interval (CI) for fit indexes proposed by Zhang and Savalei (2016) is based on the quantiles of a bootstrap sampling distribution at a single level of misspecification. This method, despite a great improvement over naive and model-based bootstrap methods, still suffers from unsatisfactory coverage. In this work, we propose a new method of constructing bootstrap CIs for various fit indexes. This method directly inverts a bootstrap test and produces a CI that involves levels of misspecification that would not be rejected in a bootstrap test. Similar in rationale to a parametric CI of root mean square error of approximation (RMSEA) based on a noncentral $\chi2$ distribution and a profile-likelihood CI of model parameters, this approach is shown to have better performance than the approach of Zhang and Savalei (2016), with more accurate coverage and more efficient widths.

Keywords: bootstrap, fit indexes, likelihood-based CI, structural equation modeling

In structural equation modeling (SEM), fit indexes provide useful information about the quality of the model and have been widely used in model evaluation. In this article, we consider four fit indexes: root mean square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger, 1990; Steiger & Lind, 1980), comparative fit index (CFI; Bentler, 1990; McDonald & Marsh, 1990), goodness-of-fit index (GFI; Jöreskog & Sörbom, 1986), and standardized root mean square residual (SRMR; Bentler, 1995). With these four indexes, the method proposed in this article can be extended to other fit indexes that are functions of them, such as the adjusted goodness-of-fit index (AGFI, Jöreskog & Sörbom, 1986; Maiti & Mukherjee, 1990) and the Tucker–Lewis index (TLI) (Bentler & Bonett, 1980; Tucker & Lewis, 1973).

The confidence intervals (CIs) of fit indexes are very informative about the estimation of the model fit, especially for the test of close fit. Because the sample point estimate of a fit index is affected by sampling error, a CI would be helpful in quantifying the range of its plausible population values. With CIs of fit indexes, hypothesis testing about fit indexes can be conducted. For example, if the entire range of CI is outside the range of close fit, close fit is rejected; if inside the range, close fit can be established. Unfortunately, many fit indexes such as GFI, CFI, SRMR are not provided with a CI in most software because their analytic asymptotic sampling distributions are generally unknown (but see Ogasawara, 2001, for a CI based on higher order approximation). Although the CI of RMSEA is widely used and provided in most software, it is based on asymptotic theory with the normal assumption of data and might not be accurate under nonnormal distributions or with small sample sizes. In this context, nonparametric bootstrap (Beran & Srivastava, 1985; Efron, 1992) is a promising method for interval estimation of fit indexes (Bollen & Stine, 1992), because it only has mild requirements for sample size and does not make distribution assumptions.

Most recently, Zhang and Savalei (2016) proposed a bootstrap CI for fit indexes. Compared with naive bootstrap, an approach that uses the sampling distribution bootstrapped from the original data, Zhang and Savalei's method uses a sampling distribution bootstrapped from a transformed sample proposed by Yuan, Hayashi, and Yanagihara (2007). The quantiles of this bootstrap sampling distribution are used as

---

Correspondence should be addressed to Chuchu Cheng, Department of Psychology, Boston College, Chestnut Hill, MA 02467. E-mail: chuchu.cheng@bc.edu

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/HSEM.

the CI. This bootstrap method improves on naive bootstrap by adjusting an upward bias in the misspecification of the bootstrap samples. CIs constructed using this method have also been shown to perform better than the naive bootstrap CIs when the degrees of freedom (df) of the fitted model are large. However, the coverage rates are still not accurate enough to the desired level. The coverage of all four fit indexes still deviates significantly from the nominal level: The coverage of 95% CIs for RMSEA and CFI could be as small as 0.900 and as large as 0.985; for GFI and SRMR, there are conditions with zero coverage due to the bias in their sample estimates. In addition, only the overall coverage rates of CIs were discussed in their paper, but it is not known whether the CIs were balanced.

In this article, we propose a new method to construct CIs for fit indexes. Also based on the Yuan, Hayashi, and Yanagihara (2007) transformation, our method transforms the observed data to different levels of population misspecification and performs a bootstrap test at each level. Because a CI is the set of null values that will not be rejected in a test, our new method searches for levels of population misspecification that retain the null hypothesis in a bootstrap test. Then these levels are transformed to form the CI of different fit indexes. The new method, the bootstrap-test-based method, is predicted to perform better, given this relationship between test and CI.

This article is organized as follows. First, we review the fit indexes (RMSEA, CFI, GFI, and SRMR) considered in the study and different types of CIs, including normal theory CI, Wald-type CI, and likelihood-based CI. Second, we explain different nonparametric bootstrap approaches for fit indexes: the basic idea of bootstrap, a brief review of naive bootstrap and model-based bootstrap, Yuan, Hayashi, and Yanaginara transformation, and Zhang and Savalei's method. Then, we introduce our new method, explain the logic, and present an algorithm. Next, we report simulation studies that compare Zhang and Savalei's method and the new method. Finally, we conclude our findings and discuss future directions.

FIT INDEXES

Let p be the number of manifest variables, $\boldsymbol{\Sigma}^*$ be the $p \times p$ true population covariance matrix, $\boldsymbol{\theta}$ be the vector of model parameters, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ be the hypothesized covariance structure. The population maximum likelihood (ML) discrepancy function is defined by:

$$F(\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = -\ln\left|\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right| + \mathrm{tr}(\boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - p. \tag{1}$$

Let $\boldsymbol{\theta}_0$ be the minimizer of the population discrepancy function, $\boldsymbol{\Sigma}_0$ be the model-implied covariance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, and F_0 be the minimum of the function. Then we have $F_0 = F(\boldsymbol{\Sigma}^*, \boldsymbol{\Sigma}_0)$.

Let $\mathbf{S}$ be the sample covariance matrix. The sample ML discrepancy function is given by:

$$F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = -\ln\left|\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}))\right| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})) - p. \tag{2}$$

Let n be the sample size, $\hat{\boldsymbol{\theta}}$ be the minimizer of the sample discrepancy function, and $\hat{\boldsymbol{\Sigma}}$ be the model-implied covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\theta})$. The minimum of F is given by $\hat{F} = F(\mathbf{S}, \hat{\boldsymbol{\Sigma}})$ and the test statistic is defined as $T = (n-1)\hat{F}$. T is asymptotically distributed as noncentral chi-square with noncentrality parameter $\lambda = (n-1)F_0$ under the convenient but unrealistic assumption of a sequence of local alternatives (Shapiro, 1983), also known as the Pitman drift assumption (McManus, 1991).

The population RMSEA and its sample estimate are defined as:

$$RMSEA_0 = \sqrt{\frac{F_0}{df}}, \tag{3}$$

$$RMSEA = \sqrt{\frac{\hat{F}}{df} - \frac{1}{n-1}} = \sqrt{\frac{\hat{\lambda}_n}{(n-1)df}}, \tag{4}$$

where $\hat{\lambda}_n$ is the noncentrality parameter estimated from the sample: $\hat{\lambda}_n = (T - df)$, and $\hat{\lambda}_n$ needs to be nonnegative.

The population and sample definitions of the ML-based CFI are given by:

$$CFI_0 = 1 - \frac{F_0}{F_{0,B}}, \tag{5}$$

$$CFI = 1 - \frac{\hat{F} - \frac{df}{n-1}}{\hat{F}_B - \frac{df_B}{n-1}} = 1 - \frac{\hat{\lambda}_n}{\hat{\lambda}_{n,B}}, \tag{6}$$

where $F_{0,B}$ and $\hat{F}_B$ are the population and sample ML discrepancy function values for the baseline model; $\hat{\lambda}_{n,B}$ is the estimated noncentrality parameter of the baseline model, $\hat{\lambda}_{n,B} = (n-1)\hat{F}_B - df_B$, and $\hat{\lambda}_{n,B}$, needs to be nonnegative.

The ML-based GFI has the following population definition and sample estimate:

$$GFI_0 = 1 - \frac{\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^* - \mathbf{I}\right)^2\right\}}{\mathrm{tr}\left\{\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}^*\right)^2\right\}}, \tag{7}$$

$$GFI = 1 - \frac{\mathrm{tr}\left\{\left(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S} - \mathbf{I}\right)^2\right\}}{\mathrm{tr}\left\{\left(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}\right)^2\right\}}. \tag{8}$$

Finally, the population SRMR and its sample estimate are defined as:

$$SRMR = \sqrt{\frac{2}{p(p+1)}\sum_{i=1}^{p}\sum_{j=1}^{i}\frac{(\sigma_{ij}^{*}-\sigma_{0,ij})^2}{\sigma_{ii}^{*}\sigma_{jj}^{*}}}, \tag{9}$$

$$SRMR_n = \sqrt{\frac{2}{p(p+1)}\sum_{i=1}^{p}\sum_{j=1}^{i}\frac{(s_{ij}-\hat{\sigma}_{ij})^2}{s_{ii}s_{jj}}}, \tag{10}$$

where $\sigma_{ij}^{*}, \sigma_{0,ij}, s_{ij}, \hat{\sigma}_{ij}$ are elements of $\mathbf{\Sigma}^{*}, \mathbf{\Sigma}_0, \mathbf{S}$, and $\hat{\mathbf{\Sigma}}$, respectively. For the models considered in our simulation studies, the correspondent diagonal elements of $\mathbf{S}$ and $\hat{\mathbf{\Sigma}}$ are identical. The preceding equation therefore computes the root mean square residual of the correlations.

DIFFERENT METHODS TO CONSTRUCT CI

Parametric Methods

In SEM, there are two widely used methods to construct parametric CIs for model parameters: likelihood-based CI and Wald-type CI. A Wald-type CI is constructed using the estimated standard error of the parameter estimate and quantiles from a normal distribution. To construct a Wald-type CI of parameter θ, one simply uses the formula: estimate ± critical value × standard error of estimate. The 95% Wald-type CI of θ is $\hat{\theta} \pm 1.96\hat{SE}_{\hat{\theta}}$.

A likelihood-based CI includes all parameter values that will not be rejected in a likelihood ratio test (LRT; Meeker & Escobar, 1995; Neale & Miller, 1997; Pek & Wu, 2015; Venzon & Moolgavkar, 1988; Wu & Neale, 2012). The likelihood-based CI of θ is constructed by inverting an LRT of the null hypothesis: $\theta = \theta_0$, for a given θ_0. In the context of SEM, if we denote the parameter of interest by θ, the other parameters by $\boldsymbol{\eta}$, then the discrepancy function can be written as $F(\mathbf{S}, \mathbf{\Sigma}(\theta, \boldsymbol{\eta}))$ and the LRT statistic as $G^2 = F(\mathbf{S}, \mathbf{\Sigma}(\theta_0, \tilde{\boldsymbol{\eta}})) - F(\mathbf{S}, \mathbf{\Sigma}(\hat{\theta}, \hat{\boldsymbol{\eta}}))$, where $\tilde{\boldsymbol{\eta}}$ is the maximum likelihood estimate (MLE) under the null hypothesis and $(\hat{\theta}, \hat{\boldsymbol{\eta}})$ is the MLE under the alternative hypothesis. G^2 follows an asymptotic χ^2 distribution with 1 df under the null hypothesis. A $100(1-\alpha)\%$ likelihood-based CI is constructed by searching for values of θ_0 such that the test statistic $G^2 \le \chi^2_{1,\alpha}$. The two values that make the LRT just significant ($G^2 = \chi^2_{1,\alpha}$) are the lower and upper limits of the CI. When parameter estimate $\hat{\theta}$ is not close to normally distributed, the profile likelihood-based CI has better performance than Wald-type CI in terms of coverage. They also enjoy the desired property of invariance to transformation (Cheung, 2009; Pek & Wu, 2015).

For fit indexes, the most widely used and reported is the normal theory CI of RMSEA. It is an inversion of the chi-square test of fit based on noncentral chi-square distribution (Browne & Cudeck, 1992; Steiger, 2016; see Wu & Browne, 2015a, 2015b, 2016 for a central chi-square based formulation). To be specific, the normal theory CI is constructed by searching for the values of F_0 that make the chi-square test just significant, that is, $T = \chi^2_{df,1-\alpha/2}((n-1)F_0)$ or $\chi^2_{df,\alpha/2}((n-1)F_0)$. Then these two values are used as lower and upper bounds of CI for F_0. The CI of RMSEA is obtained by transforming the values of F_0 back to RMSEA. The CI for the expected cross-validation index (ECVI; Browne & Cudeck, 1989), which is also a function of F_0, can also be obtained from the CI for F_0. ECVI evaluates the performance of a calibrated model to predict future samples. Its sample estimate is a scaled version of Akaike's information criterion (AIC; Akaike, 1998), a widely used model selection criterion. The model selection uncertainty in SEM due to the sampling variability of model selection criteria was discussed by Preacher and Merkle (2012). Similar to likelihood-based CI, a range of population values are searched based on the sampling distribution of T at those population values to produce a CI, but the standard error of $\hat{F}$ is not used. It has been widely reported that this normal theory CI has good performance for the range of misspecifications that are practically acceptable (Chun & Shapiro, 2009; Yuan, Hayashi, & Bentler, 2007).

Nonparametric Bootstrap

Parametric methods are based on distributional assumptions and asymptotic theory. In reality, the distribution of the data does not necessarily follow this assumption and the sample size might not be large enough. In addition, the distributions of most fit indexes other than RMSEA are unknown. In these situations, nonparametric bootstrap can prove to be an ideal method.

In general, the nonparametric bootstrap method starts by drawing independent samples of size *n*, with replacement, from the original sample (Efron & Tibshirani, 1994). We call the original sample parent population. The additional samples generated with replacement are called bootstrap samples. The statistic of interest is further computed in each of these samples. The empirical distribution of the statistics calculated from these samples is used to approximate the shape of the true sampling distribution. This methodology has been applied to SEM for constructing CIs for model parameters (Bollen & Stine, 1992; Kim & Millsap, 2014; Nevitt & Hancock, 2001; Yung & Bentler, 1996).

Traditional bootstrap methods include naive bootstrap and model-based bootstrap. The naive bootstrap defines the parent population to be the original data set. It has been proved to perform well in obtaining CIs for SEM parameters (Enders, 2001; Nevitt & Hancock, 2001; Yuan & Hayashi, 2006; Yung & Bentler, 1996). It is not appropriate for fit indexes, however, because the misfit of the parent population is usually larger than the misfit of the true population. For example, even when the true population fits the model perfectly, this parent population as the current sample is always misspecified. As a result, fit indexes estimated from naive bootstrap always overestimate their true population values (Bollen & Stine, 1992).

The model-based bootstrap transforms the original data such that the hypothesized covariance structure holds exactly true in the transformed data (Beran & Srivastava, 1985; Bollen & Stine, 1992; Savalei & Yuan, 2009; Yung & Bentler, 1996) before they are used as a parent population for bootstrap. It transforms the original data $\mathbf{y}_i$ as follows:

$$\mathbf{z}_i = \hat{\boldsymbol{\Sigma}}^{1/2}\mathbf{S}^{-1/2}\mathbf{y}_i. \tag{11}$$

This method is not appropriate for fit indexes as well, for it removes misfit entirely from the parent population.

Yuan, Hayashi, and Yanagihara transformation.

Yuan, Hayashi, & Yanagihara (2007) proposed a method that transforms the original data to parent populations with different levels of misfit, and produces covariance matrices of the parent population that fall between $\mathbf{S}$ (covariance of original data set) and $\hat{\boldsymbol{\Sigma}}$ (fitted covariance matrix). In particular, they defined a series of target covariance matrices $\mathbf{S}_a$ for the parent population that have the form:

$$\mathbf{S}_a = a\mathbf{S} + (1-a)\hat{\boldsymbol{\Sigma}}, \tag{12}$$

where a is a constant between 0 and 1.

They have shown that the model-implied covariance matrix produced by fitting $\mathbf{S}_a$ is $\hat{\boldsymbol{\Sigma}}$, and that the level of misspecification monotonically increases with a in the range of $(0, 1)$. This method has been implemented in the R package lavaan (Rosseel, 2012) by Merkle for his work (Preacher & Merkle, 2012).

Yuan, Hayashi, and Yanagihara (2007) provided an iterative algorithm for solving a for any given level of misspecification between 0 and $\hat{F}$. Once a is found, $\mathbf{S}_a$ is computed, and the original data $\boldsymbol{y}_i$ are rotated to the transformed data $\boldsymbol{x}_i$ as follows:

$$\mathbf{x}_i = \mathbf{S}_a^{1/2}\mathbf{S}^{-1/2}\boldsymbol{y}_i \tag{13}$$

and can be used for bootstrap.

Zhang and Savalei's method. Zhang and Savalei (2016) proposed the newest method of bootstrap CI for fit indexes. In this method, a covariance matrix $\mathbf{S}_a$ is first constructed using Yuan, Hayashi, and Yanagihara (2007) transformation such that its population discrepancy function value $F(\mathbf{S}_a, \hat{\boldsymbol{\Sigma}})$ equals the bias-adjusted estimate from the sample $\hat{F} - \frac{df}{n-1}$:

$$\hat{F}_a = F(\mathbf{S}_a, \hat{\boldsymbol{\Sigma}}) = \hat{F} - \frac{df}{n-1} \tag{14}$$

Second, the original data set $\boldsymbol{y}_i$ is transformed using Equation 13. The resultant data are used as a parent population with a level of misspecification of $F(\mathbf{S}_a, \hat{\boldsymbol{\Sigma}})$. With this parent population, the bootstrap sampling distribution of a fit index is obtained. Next, a CI is constructed with percentiles $100(1-\alpha/2)$ and $100\alpha/2$ of the bootstrap sampling distribution.

Because Zhang and Savalei's method uses the quantiles of a single bootstrap sampling distribution, it is expected to have similar limitations of a Wald-type CI and might not perform well when the sampling distribution of the fit index has a bias or changes quickly.

Bootstrap-test-based CI. In this article, we propose a new method that is comparable to a likelihood-based CI or a normal theory CI. It searches for a range of levels of misfit that will not be rejected in a bootstrap test based on the test statistic T. It is only different from the normal theory CI in that the new method uses bootstrap sampling distributions of the test statistic instead of the noncentral χ^2 sampling distribution under normal theory. Because of the advantages of likelihood-based CIs over Wald-type CIs, we expect that the new method has a better performance in estimating CIs of fit indexes. The details of the new method are explained next.

METHOD

We propose a new method to calculate the CI of fit indexes. In this new method, instead of using quantiles of one single bootstrap sampling distribution, we search for the range of misspecification such that the bootstrap test retains the null hypothesis. The logic of this method is explained next. First, the sample covariance matrix $\mathbf{S}$ and fitted covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the data are obtained. Then for a given constant a, a covariance matrix $\mathbf{S}_a$ can be constructed using Equation 12. A bootstrap test can be performed using the bootstrap sampling distributions of T constructed based on the transformed data set defined by Equation 13. When the value of a changes, the corresponding bootstrap sampling distribution and its quantiles also change. We search for the values of a such that this bootstrap test retains the null hypothesis. The detailed algorithm is presented next.

Algorithm

1. Obtain test statistic T and $\hat{\boldsymbol{\Sigma}}$. Given the original data set, the covariance matrix $\mathbf{S}$ is calculated and fitted to the model to obtain test statistic T and $\hat{\boldsymbol{\Sigma}}$.

2. For a given constant a, construct S_a. Following Yuan, Hayashi, and Yanagihara (2007) transformation, a covariance matrix $\mathbf{S}_a$ can be constructed using $\mathbf{S}$ and $\hat{\boldsymbol{\Sigma}}$ through Equation 12, where a is a positive constant. Note that the value of a will be determined later.
3. Transform the data. The original data are transformed using Equation 13 and the resultant data are used as the parent population for bootstrap. The covariance matrix in this parent population is $\mathbf{S}_a$.
4. Take bootstrap samples. B bootstrap samples are generated with replacement independently from these transformed data. Then we fit the model to each bootstrap sample and obtain test statistics $\hat{T}_b$ ($b = 1, 2, 3, ..., B$). These test statistics from the bootstrap samples form a bootstrap sampling distribution.
5. Perform a bootstrap test. For a given bootstrap sampling distribution, we compare the observed test statistic T from the original data set with percentiles $100(1 - \alpha/2)$ and $100\alpha/2$ of the bootstrap sampling distribution. If the observed T is within this region, we fail to reject the null hypothesis; if T is outside this region, we reject the null hypothesis.
6. Construct CIs for a. For the lower limit of this CI, search for a such that percentile $100(1 - \alpha/2)$ of the bootstrap sampling distribution equals to the observed statistic T, which makes the bootstrap test just significant. For the upper limit of a, we compare T to percentile $100\alpha/2$. These two a values are the upper and lower limits of a CI for a.
7. Construct CIs for fit indexes. Once the upper and lower limits of CI for a are found, we can construct two covariance matrices $\mathbf{S}_{a,l}$ and $\mathbf{S}_{a,u}$ using those two values. Then the fit indexes RMSEA, CFI, GFI, and SRMR are calculated as the upper and lower limits of the CIs based on $\mathbf{S}_{a,l}$ and $\mathbf{S}_{a,u}$, using Equations 3, 5, 7, and 9. We can transform the CI for a to form CIs for different fit indexes because it is proved in the Appendix that the fit indexes monotonically change with a under mild conditions.

Algorithm Efficiency

The following details are designed to improve the efficiency of the algorithm.

In the bootstrap step, we generated a single set of bootstrap samples from the original data set and then transformed it under different levels of a. If we draw bootstrap samples independently for each value of a, it would cost more time and the bootstrap sampling error would affect the convergence of the search for a.

By only bootstrapping a single set of samples and using different a values to further transform them, the algorithm is more efficient and stable.

To search for the desired values of a more efficiently, proper search bounds and starting values are needed. To find these values, the normal theory CI (F_l, F_u) and the bias-adjusted discrepancy value $\hat{F}_0 = \hat{F} - \frac{df}{n-1}$ are obtained. These discrepancy function values are then transformed to values of a (a_l, a_u,a_0) using the Newton–Raphson algorithm provided by Yuan, Hayashi, and Yanagihara (2007), which is now available in lavaan. The values of a corresponding to the normal theory CI can serve as starting values of the search. The value a_0 used in Zhang and Savalei's (2016) method can be used as a bound for the search.

To perform the bootstrap, we modified the function bootstrapLavaan() of lavaan package (Rosseel, 2012) of R (R Core Team, 2016). Additional functions to find the quantiles of the bootstrap distribution and to search for the a parameter have been written. See Supplementary Materials[1] for details.

SIMULATION

Design

A simulation study was conducted to evaluate empirically the performance of the new method. We used a confirmatory factor analysis (CFA) model with two factors and three indicators for each factor. Four representative models of misspecification conditions were examined in this study: true model (TM) conditions, correlated residuals (CR) conditions, cross-loading (CL) conditions, and wrong model (WM) conditions. These model conditions were selected to represent a range of population values of fit indexes meaningful in practice and were also similar in construction to the simulation design of Zhang and Savalei (2016). Details of these conditions are described in the next paragraph. The sample size was set to have three levels of 100, 200, and 500. In addition to the new method, our simulation study also included Zhang and Savalei's method for comparison.

In the TM conditions, the true model contained a correlation of .3 between two factors (the between-factor correlation is the same for CR and CL conditions) and the loadings for all indicators are .5 or .7; the fitted model is the same CFA model with free parameters. In the CR conditions, the true model had loadings of .7 for all indicators and contained one pair of correlated residuals, with correlation of .2, .3 or .4. This correlation is either within factor (CR-WF) or between factors (CR-CF), where WF means that the correlation is between items loaded on the same factor and CF means the correlation is between items loaded on different factors. In the CL conditions, the true model had loadings of .7 for all indicators and contained one cross-loading of .4. In both CR and CL conditions, the fitted model does not have correlated residual or cross-loading. In the WM conditions,

---

[1] Supplementary materials can be found at https://github.com/chuchucheng/Bootstrap-Test-Based-CI

the true model had a between-factor correlation of .5 or .7, and the loadings for all indicators are .5; the fitted model only contained one factor. We chose these true values in the four conditions such that the RMSEA and CFI population values are in a practically meaningful range.

The fit indexes used in this simulation study are RMSEA, CFI, GFI, and SRMR. A 90% CI was constructed for each fit index under different conditions. The distribution of the sample data is multivariate normal. For each condition, we simulated 1,000 data sets, and for each data set, 1,000 bootstrap samples were used to obtain the bootstrap sampling distributions.

For each fit index, the coverage rate and mean widths were calculated. The coverage rate is the proportion of bootstrapped CIs that contained the true population value of the fit index, which captures the accuracy. In addition, to evaluate whether the CIs are balanced or not, we also separately calculate the coverage of the lower and upper limits of these 90% CIs as 95% lower and upper confidence bounds (CBs). As explained in the method, the CIs for fit indexes were all transformed from the CI for a, so they had the same coverage rate. The mean width is the average width of the bootstrap CIs, which captures efficiency of this method.

Results

Detailed results of eleven conditions are summarized in this section. Among them, three conditions correspond to correlated residuals cross-factors (CR-CF), three conditions are for correlated residuals within factor (CR-WF), one condition corresponds to cross-loading (CL), two conditions correspond to the true model (TM), and two conditions are for the wrong model (WM).

Table 1 and Figure 1 show coverage rates of the 90% CIs of fit indexes under the bootstrap-test-based method and Zhang and Savalei's (2016) method for all conditions. For the bootstrap-test-based method, because we obtained the CIs for fit indexes by transforming the CIs for a, the coverage rates of CIs for all fit indexes are the same.

Compared with Zhang and Savalei's (2016) method, the coverage rate of the bootstrap-test-based method is closer to the nominal level of .95 on each CB, and the coverage of the CI is closer to the nominal level of .90. Specifically, for the bootstrap-test-based method, the overall coverage rate has a mean of .904, lower bound coverage rate has a mean of .952, and the upper bound coverage also has a mean of .952. For Zhang and Savalei's method, the mean overall coverage rates of RMSEA, CFI, GFI, and SRMR are .935, .940, .390, and .385, respectively; the mean lower bound coverage rates are .970, .965, .994, and .410, respectively; the mean upper bound coverage rates are .965, .975, .397, and .993, respectively.

The deviations of CI coverage rates from the nominal level are smaller in the bootstrap-test-based method. As shown in Table 1, the deviation of overall coverage rate from 0.90 under the bootstrap-test-based method ranges from .002 to .031 with a mean difference of .011. In Zhang and Savalei's (2016) method, the deviation of overall coverage rate ranges from .011 to .123 with a mean difference of .060 for RMSEA, from .007 to .087 with a mean of .057 for CFI, from .003 to .900 with a mean of .510 for GFI, and from .013 to .900 with a mean of .518 for SRMR.

We also note that the coverage rates of CIs under the bootstrap-test-based method are generally more balanced than those of Zhang and Savalei's (2016) method. The lower bound coverage rates and upper bound coverage rates are both closer to .95 in the bootstrap-test-based method; however, in Zhang and Savalei's method, for many conditions, the coverage rate of one bound is close to 1, but the coverage rate of the other bound is below .90 and even could be close to 0.

In addition, as shown in Table 2 and Figure 2, the mean width of the CIs for RMSEA and CFI under the bootstrap-test-based method is generally smaller than or approximately the same as Zhang and Savalei's (2016) method. For RMSEA, the mean widths of CIs under the bootstrap-test-based method are narrower in 21 conditions, wider in 10 conditions, and the same in 2 conditions; for CFI, the mean widths of the CIs under the bootstrap-test-based method are narrower in 30 conditions and wider in 3 conditions. We did not notice great difference of mean widths between the two methods for GFI and SRMR, and the mean widths under Zhang and Savalei's method are better than the bootstrap-test-based method for some conditions. However, the coverage rates for GFI and SRMR are generally not satisfactory under Zhang and Savalei's method, so the advantage of mean widths is not meaningful in this situation.

In summary, our simulation study shows that the CIs of the bootstrap-test-based method are substantially better than those for the Zhang and Savalei (2016) method in general, with more accurate and balanced coverage rates. In addition, for conditions where CIs from both methods have acceptable coverage, the bootstrap-test-based method gives generally shorter widths. This indicates that the bootstrap-test-based method is more powerful.

SUMMARY AND CONCLUSION

In this article, we propose a new method for constructing CIs for fit indexes using the bootstrap method. Our approach directly inverts a bootstrap test by searching the levels of misspecification (as measured by the a parameter of Yuan, Hayashi, Yanigahara transformation) that make the bootstrap test not significant. Specifically, for a given value of a, a covariance matrix $\mathbf{S}_a$ is obtained and the observed data are transformed so that their misfit as a parent population is $\mathbf{S}_a$. Using this parent population, a bootstrap test is performed with test statistic T. By comparing the observed T and the quantiles of the bootstrap sampling distribution, the null hypothesis is rejected or retained. The CI of a includes all values that are not rejected in this bootstrap test. Because the fit indexes are

TABLE 1
Coverage Rates of 90% Confidence Intervals and 95% Confidence Bounds by Bootstrap-Test-Based Method and Zhang and Savalei's Method

		PFI				*BTB Method*			*Zhang RMSEA*			*Zhang CFI*			*Zhang GFI*			*Zhang SRMR*		
Condition	*N*	*RMSEA*	*CFI*	*GFI*	*SRMR*	*LB*	*UB*	*CI*	*LB*	*UB*	*CI*	*LB*	*UB*	*CI*	*LB*	*UB*	*CI*	*LB*	*UB*	*CI*
CR-CF	100	.023	.984	.992	.020	.946	.970	.916	.974	1	.974	1	.975	.975	1	.044	.044	0	1	0
cr=.2	200					.953	.961	.914	.976	1	.976	1	.976	.976	1	.618	.618	.371	1	.371
	500					.947	.951	.898	.967	.905	.872	.906	.965	.871	1	.813	.813	.672	1	.672
CR-CF	100	.052	.964	.983	.030	.940	.958	.898	.975	1	.975	.999	.978	.977	1	.596	.596	.426	1	.426
cr=.3	200					.956	.950	.906	.973	.903	.876	.900	.971	.871	1	.734	.734	.609	1	.609
	500					.952	.955	.907	.965	.924	.889	.927	.966	.893	.981	.858	.839	.787	1	.787
CR-CF	100	.095	.936	.970	.040	.945	.953	.898	.969	.917	.886	.912	.967	.879	1	.728	.728	.596	1	.596
cr=.4	200					.954	.948	.902	.970	.911	.881	.903	.971	.874	.989	.846	.835	.755	1	.755
	500					.944	.942	.886	.962	.916	.878	.912	.959	.871	.971	.869	.840	.816	.997	.813
CR-WF	100	.001	.999	1	.008	.958	.962	.920	.982	1	.982	1	.978	.978	1	0	0	0	1	0
cr=.2	200					.951	.962	.913	.981	1	.981	1	.981	.981	1	0	0	0	1	0
	500					.936	.949	.885	.976	1	.976	1	.977	.977	1	0	0	0	1	0
CR-WF	100	.003	.998	.999	.012	.964	.967	.931	.985	1	.985	1	.985	.985	1	0	0	0	1	0
cr=.3	200					.961	.952	.913	.981	1	.981	1	.981	.981	1	0	0	0	1	0
	500					.949	.961	.910	.976	1	.976	1	.977	.977	1	0	0	.533	1	.533
CR-WF	100	.004	.998	.999	.015	.958	.967	.925	.979	1	.979	1	.980	.980	1	0	0	0	1	0
cr=.4	200					.940	.958	.898	.968	1	.968	1	.970	.970	1	0	0	.001	1	.001
	500					.952	.959	.911	.982	1	.982	1	.979	.979	1	0	0	.944	1	.944
CL	100	.106	.944	.964	.061	.939	.945	.884	.961	.903	.864	.937	.939	.876	1	.794	.794	.901	.978	.879
cl=.4	200					.950	.942	.892	.971	.913	.884	.936	.957	.893	.971	.873	.844	.936	.939	.875
	500					.944	.941	.885	.956	.923	.879	.945	.937	.882	.958	.909	.867	.939	.940	.879
TM	100	0	1	1	0	.956	.967	.923	.777	1	.777	1	.981	.981	1	0	0	0	1	0
l=.7	200					.948	.954	.902	.978	1	.978	1	.979	.979	1	0	0	0	1	0
	500					.958	.950	.908	.981	1	.981	1	.982	.982	1	0	0	0	1	0
TM	100	0	1	1	0	.958	.951	.909	.985	1	.985	1	.985	.985	1	0	0	0	1	0
l=.5	200					.960	.942	.902	.988	1	.988	1	.987	.987	1	0	0	0	1	0
	500					.948	.950	.898	.981	1	.981	1	.981	.981	1	0	0	0	1	0
WM	100	.071	.824	.973	.052	.972	.936	.908	.987	.981	.968	.922	.998	.920	1	.799	.799	.785	1	.785
fc=.5	200					.953	.924	.877	.987	.867	.854	.856	.994	.850	.977	.890	.867	.881	.974	.855
	500					.954	.943	.897	.978	.895	.873	.900	.984	.884	.951	.946	.897	.933	.954	.887
WM	100	.029	.938	.989	.031	.951	.960	.911	.984	1	.984	1	.984	.984	1	.295	.295	.217	1	.217
fc=.7	200					.944	.951	.895	.975	1	.975	.982	.975	.957	1	.641	.641	.608	1	.608
	500					.961	.949	.910	.975	.899	.874	.902	.974	.876	.992	.840	.832	.806	.994	.211
Mean Coverage						.952	.952	.904	.970	.965	.935	.965	.975	.940	.994	.397	.390	.410	.993	.385
Max Deviation from Nominal						.022	.026	.031	.173	.083	.123	.094	.048	.087	.050	.950	.900	.950	.050	.900
Min Deviation from Nominal						.000	.000	.002	.006	.026	.011	.005	.007	.007	.001	.004	.003	.006	.004	.013
Mean Deviation from Nominal						.006	.008	.011	.030	.047	.060	.045	.026	.057	.044	.553	.510	.540	.044	.518

Note. BTB method = bootstrap-test-based method. Since the CIs for fit indices by BTB method were all transformed from the CI for *a*, they had the same coverage rate.

PFI=Population fit index value. LB=lower bound; UB=upper bound; CI=confidence interval. CR-CF=correlated residual cross factors; CR-WF=correlated residual within factors. CL=cross loading. TM=true model. WM=wrong model. cr=correlated residual. cl=cross loading. l=factor loading. fc=factor correlation.

Factor loadings are .7 for all CF, WF, CL conditions. For TM, factor loading is .5 or .7. For WM, factor loading is .5.

generally monotonically related to *a*, their CIs can be directly calculated from the CI of *a*.

The new method, the bootstrap-test-based method, is theoretically predicted to perform better than Zhang and Savalei's (2016) method, which uses the quantiles of a single bootstrap sampling distribution to construct CI and is comparable to a Wald-type CI in parametric statistics. The reason is that CIs based on inverting the distribution of the test statistic has been shown to have better performance than Wald-type CIs (Pek & Wu, 2015). Especially, a Wald-type CI is based on the distribution of sample estimates of fit indexes that might overestimate their population value, leading to biased results. To be specific, Zhang and Savalei's method has poor performance when constructing CIs for GFI and SRMR because their sample estimates do not involve bias adjustment. Our method searches for levels of population misfit directly and does not suffer from this problem, so it is also predicted to provide particularly more accurate bootstrap CIs for GFI and SRMR.

From the simulation results, the CIs for all four fit indexes under the bootstrap-test-based method are more accurate and efficient than Zhang and Savalei's (2016) method: Coverage rates of both CIs and CBs (the upper

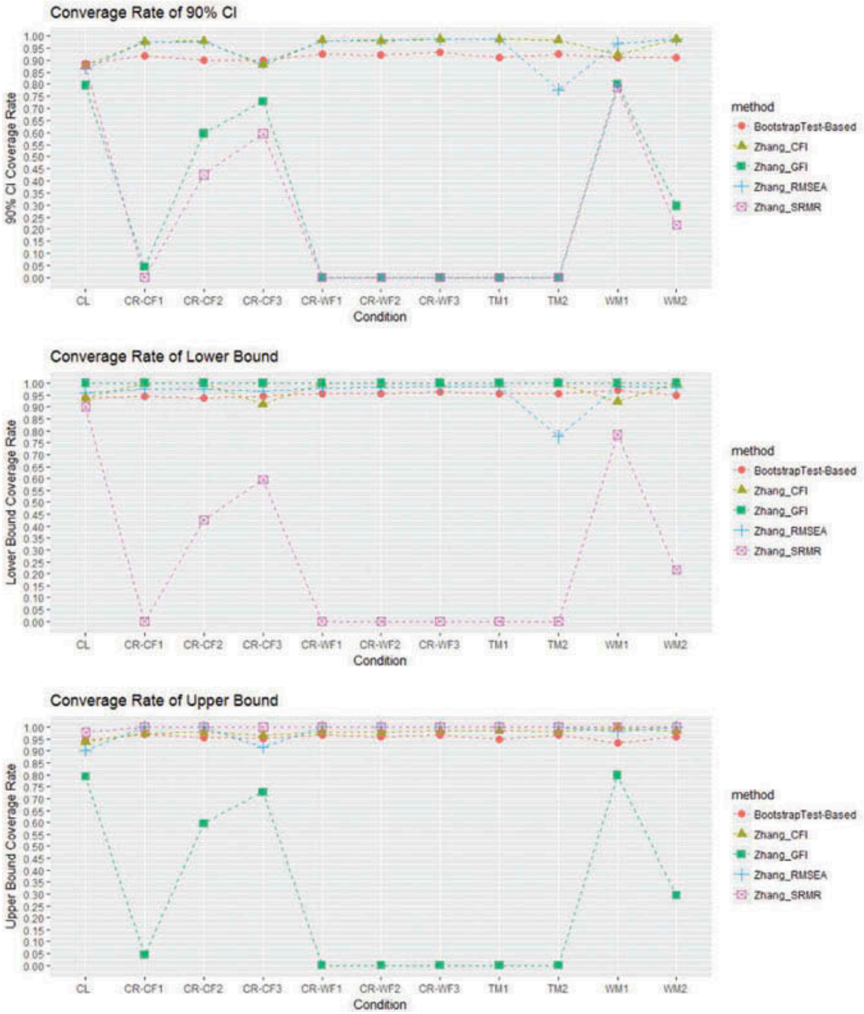


FIGURE 1 Coverage Rates of 90% Cis and 95% CBs by BTB Method and Zhang and Savalei's Method. *Note*. The sample size is N = 100. CB = Confidence Bounds, upper and lower bounds of CI. BTB method = bootstrap test based method. The desired coverage rate of CI is 90%; the desired coverage rate of CB is 95%.

and lower limits of CI) constructed by the bootstrap test-based method are closer to their nominal levels; the CIs are also narrower when the coverage rates were satisfactory for both methods. In particular, our method does not suffer the zero or unity coverages frequently present in Zhang and Savalei's method. These results confirmed our predictions.

Similar to Zhang and Savalei's (2016) study, we only considered normally distributed data in this article, because the definitions of fit indexes could still be controversial (Brosseau-Liard, Savalei, & Li, 2012; Yuan, 2005; Yuan & Marshall, 2004). However, the bootstrap-test-based method can be generalized to nonnormal data if the same definitions of fit indexes are used. For other definitions of fit indexes under conditions of nonnormal distribution, one first needs to establish that the fit indexes increase monotonically with the value of a. This line of research will be considered in the future.

The bootstrap-test-based method, when applied to CFI, could be restrictive: It only works for covariance structures that satisfy $\mathrm{diag}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}(\mathbf{S})$. Although a large class of models satisfy this condition (see the comments at the end of the Appendix), structural equation models with a non-saturated factor correlation matrix do not. This restriction excludes most models for longitudinal data, such as latent growth curve models. Constructing bootstrap CIs for CFI in more general conditions could be investigated in the future.

This method only applies to data without missing values. With missing data, the bootstrap procedures could be more complicated (Savalei & Yuan, 2009), even for the test of exact fit. How to properly transform a data set with missing values to obtain different levels of misspecification will be the pivotal problem in the extension of this method to data with missing values. This is beyond the scope of this article, but could be an interesting topic for future research.

TABLE 2
Mean Widths of 90% Bootstrap-Test-Based Method and Zhang and Savalei's Method Confidence Intervals

		BTB Method				*Zhang and Savalei's Method*			
Condition	*N*	*RMSEA*	*CFI*	*GFI*	*SRMR*	*RMSEA*	*CFI*	*GFI*	*SRMR*
CR-CF,cr=.2	100	0.150	0.093	0.045	0.054	0.178	0.150	0.050	0.048
	200	0.085	0.056	0.027	0.040	0.094	0.066	0.029	0.034
	500	0.049	0.034	0.016	0.026	0.049	0.034	0.015	0.022
CR-CF,cr=.3	100	0.180	0.111	0.053	0.056	0.199	0.134	0.055	0.049
	200	0.120	0.076	0.036	0.039	0.120	0.081	0.035	0.035
	500	0.071	0.047	0.022	0.022	0.070	0.048	0.021	0.023
CR-CF,cr=.4	100	0.236	0.136	0.066	0.057	0.234	0.151	0.062	0.051
	200	0.154	0.093	0.045	0.036	0.150	0.098	0.042	0.036
	500	0.093	0.057	0.027	0.020	0.092	0.061	0.027	0.023
CR-WF,cr=.2	100	0.110	0.066	0.034	0.050	0.159	0.103	0.046	0.047
	200	0.055	0.035	0.018	0.035	0.077	0.051	0.024	0.033
	500	0.024	0.015	0.008	0.023	0.031	0.021	0.010	0.021
CR-WF,cr=.3	100	0.110	0.062	0.034	0.050	0.159	0.097	0.046	0.048
	200	0.056	0.034	0.018	0.036	0.077	0.048	0.024	0.033
	500	0.026	0.016	0.008	0.025	0.032	0.020	0.010	0.021
CR-WF,cr=.4	100	0.155	0.061	0.035	0.051	0.162	0.093	0.046	0.049
	200	0.060	0.034	0.019	0.038	0.080	0.047	0.025	0.034
	500	0.027	0.016	0.009	0.027	0.033	0.019	0.011	0.022
CL,cl=.4	100	0.248	0.113	0.072	0.067	0.243	0.128	0.066	0.055
	200	0.166	0.080	0.051	0.046	0.160	0.086	0.047	0.043
	500	0.101	0.049	0.032	0.028	0.099	0.053	0.031	0.029
TM,l=.7	100	0.116	0.199	0.033	0.057	0.207	0.112	0.056	0.062
	200	0.053	0.118	0.017	0.040	0.076	0.055	0.024	0.032
	500	0.021	0.053	0.007	0.026	0.030	0.022	0.010	0.020
TM,l=.5	100	0.109	0.199	0.033	0.057	0.151	0.420	0.044	0.049
	200	0.053	0.118	0.017	0.040	0.073	0.199	0.023	0.034
	500	0.021	0.053	0.007	0.026	0.030	0.081	0.010	0.022
WM,fc=.5	100	0.210	0.328	0.065	0.076	0.207	0.491	0.059	0.052
	200	0.143	0.256	0.048	0.055	0.132	0.295	0.042	0.040
	500	0.087	0.169	0.031	0.033	0.082	0.188	0.028	0.029
WM,fc=.7	100	0.162	0.240	0.050	0.064	0.184	0.394	0.053	0.047
	200	0.100	0.173	0.032	0.049	0.103	0.215	0.032	0.034
	500	0.058	0.109	0.020	0.032	0.056	0.117	0.019	0.024

Note. BTB method = bootstrap-test-based method. PFI=Population fit index value. CF=correlated residual cross factors. WF=correlated residual within factors. CL=cross loading. TM=true model. WM=wrong model. cr=correlated residual. cl=cross loading. l=factor loading. fc=factor correlation.

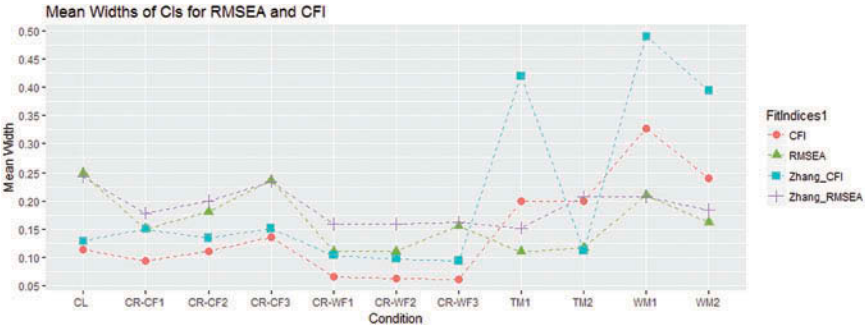


FIGURE 2 Mean Widths of 90% BTB Method and Zhang and Savalei's Method Cis for RMSEA and CFI. *Note*. The sample size is N = lOO. BTB method = bootstrap test based method. The Cis for GFI and SRMR are not included, because the CI coverage rates of them under Zhang and Savalei's method are generally not satisfactory.

ACKNOWLEDGMENTS

We are grateful to Dr. Edgar C. Merkle for sharing the R code he wrote for the lavaan package and to Dr. Michael D. Hunter for his tips on R programming. We would also like to thank the two anonymous reviewers for their suggestions.

FUNDING

Work on this article by both authors during Summer 2016 was supported by a Research Expense Grant from Boston College. Work on this article by Hao Wu during Autumn 2015 was supported by a Faculty Fellowship from Boston College.

ORCID

Chuchu Cheng http://orcid.org/0000-0002-1578-7497
Hao Wu http://orcid.org/0000-0001-6471-1774

REFERENCES

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa, *Selected papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer.

Anderson, T. (2003). *An introduction to multivariate statistical analysis*. New York, NY: Wiley.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238.

Bentler, P. M. (1995). *EQS structural equations program manual*. Chicago, IL: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588.

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

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, *21*, 205–229.

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, *47*, 904–930.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445–455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*, 230–258.

Cheung, M. W.-L. (2009). Constructing approximate confidence intervals for parameters with structural equation models. *Structural Equation Modeling*, *16*, 267–294.

Chun, S. Y., & Shapiro, A. (2009). Normal versus noncentral chi-square asymptotics of misspecified models. *Multivariate Behavioral Research*, *44*, 803–827.

Efron, B. (1992). Bootstrap methods: Another look at the jackknife. In S. Kotz & N. L. Johnson, *Breakthroughs in statistics* (pp. 569–593). Springer.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC.

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*, 352.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Chicago, IL: Scientific Software.

Kim, H., & Millsap, R. (2014). Using the Bollen–Stine bootstrapping method for evaluating approximate fit indices. *Multivariate Behavioral Research*, *49*, 581–596.

Lawley, D., & Maxwell, A. (1963). *Factor analysis as a statistical method*. Oxford, UK: Butterworths.

Maiti, S. S., & Mukherjee, B. N. (1990). A note on distributional properties of the Jöreskog–Sörbom fit indices. *Psychometrika*, *55*, 721–726.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247.

McManus, D. A. (1991). Who invented local power analysis? *Econometric Theory*, *7*, 265–268.

Meeker, W. Q., & Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, *49*, 48–53.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, *27*, 113–120.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, *8*, 353–377.

Ogasawara, H. (2001). Approximations to the distributions of fit indexes for misspecified structural equation models. *Structural Equation Modeling*, *8*, 556–574.

Pek, J., & Wu, H. (2015). Profile likelihood-based confidence intervals and regions for structural equation models. *Psychometrika*, *80*, 1123–1145.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*, 1–14.

R Core Team. (2016). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Savalei, V., & Yuan, K.-H. (2009). On the model-based bootstrap with missing data: Obtaining a p-value for a test of exact fit. *Multivariate Behavioral Research*, *44*, 741–763.

Shapiro, A. (1983). Asymptotic distribution theory in the analysis of covariance structures. *South African Statistical Journal*, *17*(1), 33–81.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling: 23*, 777–781.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. In *Annual Meeting of the Psychometric Society* (Vol. 758, pp. 424–453). Iowa City, IA.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Venzon, D., & Moolgavkar, S. H. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, *37*, 87–94.

Wu, H., & Browne, M. W. (2015a). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*(3), 571–600.

Wu, H., & Browne, M. W. (2015b). Random model discrepancy: Interpretations and technicalities (A Rejoinder). *Psychometrika*, *80*(3), 619–624.

Wu, H., & Browne, M. W. (2016). Erratum to: Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *81*(4), 1168.

Wu, H., & Neale, M. C. (2012). Adjusted confidence intervals for a bounded parameter. *Behavior Genetics*, *42*, 886–898.

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

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, *59*, 397–417.

Yuan, K.-H., Hayashi, K., & Bentler, P. M. (2007). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. *Journal of Multivariate Analysis*, *98*, 1262–1282.

Yuan, K.-H., Hayashi, K., & Yanagihara, H. (2007). A class of population covariance matrices in the bootstrap approach to covariance structure analysis. *Multivariate Behavioral Research*, *42*, 261–281.

Yuan, K.-H., & Marshall, L. L. (2004). A new measure of misfit for covariance structure models. *Behaviormetrika*, *31*(1), 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: Issues and techniques* (pp. 195–226). Mahwah, NJ: Lawrence Erlbaum.

Zhang, X., & Savalei, V. (2016). Bootstrapping confidence intervals for fit indexes in structural equation modeling. *Structural Equation Modeling*, *23*, 392–408.

APPENDIX

In this appendix we establish the fact that the fit indexes used in this article monotonically increase with the value of a under mild conditions.

Proposition 1

If $\mathbf{S} \neq \hat{\boldsymbol{\Sigma}}$, *then in the parent population with covariance matrix* $\mathbf{S}_a$ *given by Equation 12, RMSEA strictly increases as a function of* a *for its entire range; CFI strictly decreases as a function of* a *if the following relationships are satisfied:* $\mathrm{diag}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}(\mathbf{S})$, $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) = p$ *and* $\hat{\boldsymbol{\Sigma}}$ *is not a diagonal matrix; GFI strictly decreases as a function of* a *for covariance structures whose ML estimate satisfies* $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) = p$*; SRMR strictly increases as a function of* a.

Proof. The discrepancy function value of the parent population defined by $\mathbf{S}_a$ is

$$F_a = F(\mathbf{S}_a, \hat{\boldsymbol{\Sigma}}) = -\ln\left|\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a\right| + \mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) - p. \qquad (\mathrm{A}.1)$$

Let the eigenvalues of $\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}$ be $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_p$. Yuan, Hayashi, and Yanagihara (2007) proved that for $0 \leq a \leq 1$, F_a is a strictly increasing function of a. The derivative of F_a as provided in Yuan et al.'s proof is

$$F_a' = a\sum_{i=1}^{p}\frac{(\lambda_i - 1)^2}{(1-a) + a\lambda_i}.$$

It can be easily observed that this is a positive function on the full range of a for which F_a exists, which is $0 < a < \frac{1}{1-\lambda_1}$. Because RMSEA is strictly increasing with F in the population, it is also strictly increasing with a.

The equation we used for CFI is

$$CFI_a = 1 - \frac{F_a}{F_{B,a}}, \qquad (\mathrm{A}.2)$$

where F_a is the discrepancy function value associated with $\mathbf{S}_a$ and $F_{B,a}$ is the baseline discrepancy function value. Because of the relationship $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) = p$, the discrepancy function value is simplified as $F_a = -\ln\left|\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a\right|$ and the baseline discrepancy function value is simplified as $F_{B,a} = -\ln\left|\hat{\mathbf{D}}^{-1}\mathbf{S}_a\right|$.

If the model satisfies the relationship $\mathrm{diag}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}(\mathbf{S})$, the null model estimate $\hat{\mathbf{D}}$ must satisfy $\hat{\mathbf{D}} = \mathrm{diag}(\mathbf{S}_a) = \mathrm{diag}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}(\mathbf{S})$ and does not depend on a. The baseline discrepancy function value can be written as $F_{B,a} = -\ln\left|\hat{\mathbf{D}}^{-1}\mathbf{S}_a\right| = -\ln\left|\hat{\mathbf{D}}^{-1}\hat{\boldsymbol{\Sigma}}\right| - F_a$, and as a result,

$$CFI_a = 1 - \frac{F_a}{F_{B,a}} = 1 - \frac{F_a}{\ln\left|\hat{\mathbf{D}}^{-1}\hat{\boldsymbol{\Sigma}}\right| + F_a}.$$

Because $\hat{\boldsymbol{\Sigma}}$ is not a diagonal matrix, we have $\ln\left|\hat{\mathbf{D}}^{-1}\hat{\boldsymbol{\Sigma}}\right| < 0$, so CFI strictly decreases with F_a and therefore with the value of a.

Using the relationship $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) = p$, we can express GFI as

$$GFI_a = 1 - \frac{\mathrm{tr}\left\{(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a)^2 - 2(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) + \mathbf{I}\right\}}{\mathrm{tr}\left\{(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a)^2\right\}} = \frac{p}{\mathrm{tr}\left\{(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a)^2\right\}}.$$

Now we only need to prove $\mathrm{tr}\left\{(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a)^2\right\}$ is strictly increasing with a. Express $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a)^2 = \sum_i^p (a\lambda_i + 1 - a)^2 = a^2\sum_i^p (\lambda_i - 1)^2 + 1$ and we see it is indeed the case.

For SRMR, we express

$$\sum_{i=1}^{p}\sum_{j=1}^{i}\frac{\left(s_{ij}^{(a)} - \hat{\sigma}_{ij}\right)^2}{s_{ii}^{(a)}s_{jj}^{(a)}} = \sum_{i=1}^{p}\sum_{j=1}^{i}\frac{(as_{ij} - a\hat{\sigma}_{ij})^2}{[a(s_{ii} - \hat{\sigma}_{ii}) + \hat{\sigma}_{ii}][a(s_{jj} - \hat{\sigma}_{jj}) + \hat{\sigma}_{jj}]}$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{i}\frac{(s_{ij} - \hat{\sigma}_{ij})^2}{[(s_{ii} - \hat{\sigma}_{ii}) + \hat{\sigma}_{ii}/a][(s_{jj} - \hat{\sigma}_{jj}) + \hat{\sigma}_{jj}/a]}$$

and see it is strictly increasing with a. Now we have prove the proposition.

Although there are some restrictions in our proposition, these restrictions are generally satisfied by a large class of models. In fact, all exploratory factor analysis (EFA) models (Anderson, 2003, section 14.3.1) and CFA models where all fixed loadings are zero and factor correlations and unique variances are free parameters (Lawley & Maxwell, 1963, Exercise 7.3) satisfy the relationship $\mathrm{diag}(\hat{\boldsymbol{\Sigma}}) = \mathrm{diag}(\mathbf{S})$. As long as the covariance structure is closed under scalar multiplications (for all C and $\boldsymbol{\theta}$, there exists a $\boldsymbol{\theta}_c$ that satisfies $c\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_c)$), it satisfies $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_a) = p$. This second condition is met by most models in practice.