

UC Merced

UC Merced Electronic Theses and Dissertations

Title

The Robustness of the Delta Method and Bootstrap in Calculating Standardized Coefficient in Misspecified SEM Models

Permalink

<https://escholarship.org/uc/item/2xp9k61c>

Author

Zhang, Xiaoguang

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

The Robustness of the Delta Method and Bootstrap in Calculating Standardized Coefficient
in Misspecified SEM Models

A Thesis submitted in partial satisfaction of the requirements
for the degree of Master of Arts

in

Psychological Sciences

by

Xiaoguang Zhang

Committee in charge:

Professor Keke Lai, Chair
Professor Jack Vevea
Professor Sarah Depaoli

2017

Copyright (or ©)

Xiaoguang Zhang, 2017

All rights reserved

The Thesis of Xiaoguang Zhang is approved, and it is acceptable in
quality and form for publication on microfilm and electronically:

Sarah Depaoli

Jack Vevea

Keke Lai, Chair

University of California, Merced

2017

The Robustness of the Delta Method and Bootstrap in Calculating Standardized Coefficient
in Misspecified SEM Models

Xiaoguang Zhang

University of California, Merced

Abstract

In practice, models almost always have misfit. The misfit of a structural equation model (SEM) can affect the estimation of model parameters in terms of point estimates, standard errors (SEs), and confidence intervals (CIs). In this article, the performance of different methods are compared including: the delta method, the nonparametric bootstrap (NP-B) method, and the semi-parametric bootstrap (SP-B) method in estimating standardized model parameters under the influence of model misfit. Two methods are used to construct the CI of the bootstrap method, the standard percentile method and the bias-corrected and accelerated (BC_a) method. A simulation study is conducted using an SEM model with different amounts of model misfit and various sample sizes. The results show that if the model is correctly specified, all methods give correct point estimates, SE estimates, and CI coverage rates. However, as the amount of model misfit increases, the estimates based on NP-B remain accurate and consistent, whose based on the delta method and SP-B deteriorate

Keywords: misspecification, standard error, SEM, standardized parameters

The Robustness of the Delta Method and Bootstrap in Calculating Standardized Coefficient in Misspecified SEM Models

Structural equation modeling (SEM) is a commonly used statistical method in the social sciences. In an SEM study, researchers proposed several models to explain the relationships between variables based on substantive theories. After data are collected, researchers fit the candidate models to the data and compare the model fit. The model which is most consistent with both theories and the data is selected. The next step is to interpret the relationships between variables and such relationships are in the form of model coefficients. The model coefficients describes the strength and the direction of the relationship between variables, which can be used for developing and modifying theories.

There are two ways to estimate coefficients: point estimation and interval estimation. Point estimation seeks the most plausible single value for a population parameter. The advantage of that approach is that it gives us a straightforward view on the proposed model's coefficients. However, in practice, there is always uncertainty associated with the point estimates because the true population parameter may be larger or smaller than the sample value. Point estimation does not tell us how certain the estimation is unless we take standard error (SE) into consideration. Interval estimation overcomes these drawbacks by providing us with the certainty (via the interval's width) of our estimate. If the interval is narrow, we will have more confidence about our estimate. One common type of interval estimation is confidence interval (CI). According to the Publication Manual of the American Psychological Association (APA), researchers are strongly recommended to report the CI because it tells us the certainty of the estimation (APA, 2010, p.34).

Statistical methods for model evaluation, model selection and the estimation of coefficients have been mainly developed in the unstandardized context. However, if variables have different scales of measurement, their unstandardized coefficients are not comparable

and difficult to interpret. In order to solve this, researchers usually use standardized coefficients, which can be transformed from unstandardized coefficient to facilitate the interpretation. Most SEM software packages offer point estimates for standardized coefficients, and some can also calculate standard errors and confidence intervals. Although standardized coefficient provides us with an easy way to interpret models, they are less straightforward to estimate as compared to unstandardized ones. One naïve way is to fit the model to a correlation matrix to obtain standardized coefficients. But the usefulness of this method depends on whether or not the model is scale invariant. If the model is scale-invariant, the point estimation is unbiased, but the SE is not correct (see, e.g., Cudeck, 1989). If the model is scale-variant, neither the point estimation nor the SE is not accurate (e.g., Cudeck, 1989). A more reasonable approach is to estimate the unstandardized coefficients first and then transform them into standardized ones. In this case, standardized coefficients are a function of the unstandardized coefficients. Different types of coefficient have different transformation functions. There are two approaches to estimate standardized coefficients: the asymptotic approach and the empirical approach. The most common asymptotic approach is maximum likelihood (ML). The ML point estimate of standardized coefficient is transformed from the ML point estimate of unstandardized coefficient. One can then apply the delta method to construct the sampling distributions of the standardized coefficients so as to obtain SEs and CIs. The most common type of ML is normal theory ML and it relies on two crucial assumptions: normality and correct model specification. If these assumptions do not hold in practice, the estimation of standardized coefficient could be compromised. In fact, the two assumptions are usually violated in practice.

Another approach uses the bootstrap resampling technique to construct an empirical distribution of the estimates of the standardized population parameters. Then the standardized population parameters can be estimated from this empirical distribution without assuming

normality or correct model. One main difference between the asymptotic method and the bootstrap method is that the former estimates population parameters directly based on the sample, whereas the latter resamples from the sample. The bootstrap estimates population parameters are obtained based on the bootstrap resamples. The current study includes two bootstrap approaches: the non-parametric approach (NP-B) and the semi-parametric approach (SP-B). The difference between these two approaches is that the NP-B uses original sample space, whereas the SP-B uses a transformed. The purpose of the SP-B method is to compensate for the model misspecification (Yuan & Hayashi, 2006). Although in theory, the bootstrap methods should outperform the delta method when model is misspecified, no study has compared them in simulation. Therefore, the current study compares the performance of three methods: the asymptotic method, the NP-B method, and the SP-B method.

It is generally accepted that any model is only an approximation to reality in the SEM literature (e.g., MacCallum, 1986). The misspecification error can be caused by omitting variables and/or paths between variables. So the misspecification model is common in SEM because the relationship between variables in population level is unknown. The study of the impact of the model misspecification has great value because it can affect not only the overall model evaluation but also the parameter estimates (Intriligator, 1978, Yuan, Bentler, 2007). In particular, the use of chi-square as fit index in misspecified models can lead to accepting models with severe parameter bias (Kaplan, 1988). Moreover, the misspecification error can affect the ML estimates of the measurement and the structural parameters (see Judgem Griffiths, Hill, Lutkepohl, & Lee, 1985, pp. 857-859; Kaplan, 1988). Yuan and Hayashi (2006; see also Arminger & Schoenberg, 1989; Browne & Arminger, 1995; Shapiro, 1983) gave a consistent SE estimator for misspecified models using normal theory ML, and derived conditions for bootstrap's robustness to model misspecifications. Jennrich (2008) gave consistent SEs using infinitesimal jackknife (IJ), a special case of bootstrap (e.g., Efron &

Tibshirani, 1993, Chapter 21), without assuming correct model or normal data. But no simulation studies have been carry out to assess these treatments for model misspecifications. With respect to CIs, Bollen and Stine (1990), Ichikawa and Konishi (1995), and Lambert et al. (1991) investigated the utility of the bootstrap within a limited scope.

The consequences of violating the normality and correct model assumptions are not completely clear yet. The related literature has been primarily in the *unstandardized* context. When the population follows the multivariate normal distribution and model is correctly specified, the point estimates by the ML method are unbiased (Curran et al. 1996). For the bootstrap method, the accuracy of the estimated SEs increases with the sample size from $N=100$ to $N=1000$ (Nevitt & Hancock, 2001). Moreover, on the one hand, if the population is not normally distributed, the estimates by the ML method are unbiased (Anderson & Gerbing, 1984; Boomsma, 1986; Browne, 1982; Chou et al. 1991; Curran et al. 1996; Finch et al. 1997; Harlow, 1985; Hu et al. 1992), but the ML estimated SEs tend to be negatively biased (Finch et al. 1997). On the other hand, the SE estimates by the bootstrap method are larger than the ones obtained by the ML method (Boomsma, 1986). However, that result is not confirmed by Nevitt and Hancock (2001), who found that the bias of the SE estimates by the bootstrap method is actually smaller than the SE estimates by the ML method. Furthermore, if the model is misspecified and the population is normally distributed, the point estimates by the ML method are not substantially different from the population parameters (Curran et al. 1996). In addition, when the combining misspecified model and the nonnormality conditions, SE estimates by the ML method are not affected by misspecification.

No study has examined different methods to estimate *standardized* coefficient in the context of SEM. That is the purpose of the current paper. Specifically, this paper compares the performance of the asymptotic approach and the empirical approach in estimating

standardized coefficients under different model-correctness conditions. There are three misfit conditions with varying amounts of model misfit, and one correct model condition. For typical SEM simulation studies, sample size ranges from 200 to 1000. Sample size less than 200 can result inaccurate estimates and sample size larger than 1000 requires extensive amount of time to complete. Therefore, I select three sample size conditions, 300, 500, and 800. In each sample size condition, I use the ML method, the non-parametric bootstrap (NP-B), and the semi-parametric bootstrap (SP-B), which will be introduced in the following section, to estimate standardized model parameter. There are three methods to estimate the SEs of the standardized coefficients: the delta method, the NP-B method, and the SP-B method. For the estimation of the CI coverage rates, there are five methods: the delta method, the NP-B percentile method, the NP-B bias-corrected and accelerated (BC_a) method, the SP-B percentile method, and the SP-B BC_a method.

The structure of the current study is as follows: I introduce different methods to estimate the standardized coefficients with respect to point estimates, SEs, and the CI coverage rates. The design of the simulation is covered in the next section, followed by a description of the evaluation criteria. I present the simulation results and give recommendations on the choice of method. Finally, I present a discussion of the results

Standardized Model Parameter Estimation

Maximum Likelihood Estimation

Let Σ denote the variance covariance matrix of vector \mathbf{x} with p variables. Let $\Sigma(\cdot)$ denote the structure of interest. By using a fit function, the goal is to minimize the discrepancy between Σ and $\Sigma(\cdot)$, which is $F(\Sigma, \Sigma(\cdot))$. In this paper, I assume the discrepancy measure is the normal-theory ML fit function (i.e., $F_{ML}(\Sigma, \Sigma(\cdot)) = \ln|\Sigma(\cdot)| + \text{tr}(\Sigma\Sigma(\cdot)^{-1}) - \ln|\Sigma| - p$). Let θ denote the argument that minimizes the fit function $F(\Sigma, \Sigma(\cdot))$.

So θ is considered to be the unstandardized population parameter, no matter $F(\Sigma, \Sigma(\cdot)) = 0$ (correct model) or $F(\Sigma, \Sigma(\cdot)) > 0$ (misspecified model). Let $\hat{\theta}$ denote the ML estimate of θ . Then standard asymptotics lead to

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(\mathbf{0}, \Omega), \quad (1)$$

where \xrightarrow{D} means converge in distribution and Ω is the asymptotic variance covariance matrix of $\sqrt{n}\hat{\theta}$.

A standardized population parameter is defined as

$$\theta_{ij}^s = \theta_{ij}(\sigma_{jj}/\sigma_{ii}), \quad (2)$$

where the superscript s represents the standardized form, θ_{ij} is the unstandardized population parameter, i is the index of the dependent variable, j is the index of the independent variable, and σ is the population model-implied standard deviation of the i and j variable in the unstandardized context (e.g., Bollen, 1989, p.349). σ is not parameter in the model but function of model parameters. Based on Equation 2, the standardized coefficient can be considered a function of its unstandardized coefficient.

Delta Method

Based on the relationship between the standardized and unstandardized coefficients, one can use the delta method to estimate the asymptotic distribution of the standardized coefficients from Equation 1,

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \xrightarrow{D} N(0, \text{var}(\hat{\theta})[g'(\theta)]^2), \quad (3)$$

where g is the function transforming the unstandardized coefficients to the standardized coefficients (Equation 2), $g(\theta)$ is the value of g evaluated at the population coefficient θ and the first derivative $g'(\theta)$ exists and is no zero. Therefore, the SE estimate at a given sample size n is $\hat{\sigma}(\hat{\theta})g'(\hat{\theta})$, obtained by substituting $\hat{\theta}$ for θ in calculating $\sigma(\hat{\theta})$ and $g'(\theta)$, where

$\sigma(\hat{\theta}) = \sqrt{\text{var}(\hat{\theta})}$. One can use the SE to obtain the confidence limits of a $(1 - \alpha)\%$ CI as $[g(\hat{\theta}) - z_{1-\alpha} \cdot \hat{\sigma}(\hat{\theta})g'(\hat{\theta}), g(\hat{\theta}) + z_{1-\alpha} \cdot \hat{\sigma}(\hat{\theta})g'(\hat{\theta})]$.

Bootstrap Estimation of Standardized parameters

The bootstrap method draws multiple samples with replacement from the existing sample. The population parameters are estimated from each bootstrap sample. Then one can use these estimates to construct a distribution of the estimated population parameters to obtain the point estimates, the SE estimates, and the CI coverage rates of the population parameters.

Given a sample $\mathbf{x}_i (i=1, 2, \dots, n)$, let us construct a bootstrap sample space, $\mathbf{x}_i^* = \mathbf{S}_a^{1/2} \mathbf{S}^{-1/2} \mathbf{x}_i$, for $i = 1, 2, \dots, n$, where, \mathbf{x}_i^* is the bootstrap sample space, \mathbf{S} is the covariance matrix of the sample space, and \mathbf{S}_a depends on the method of bootstrap. If $\mathbf{S}_a = \mathbf{S}$, it is the nonparametric bootstrap (NP-B). That is, the bootstrap sample space is equal to the original sample space. If \mathbf{S}_a is the model-implied covariance matrix, it is the semi-parametric bootstrap (SP-B). It can be proved that the covariance matrix of the SP-B sample space is equal to the model-implied covariance matrix. The usefulness of SP-B is proposed by Yuan and Hayashi (2006) in order to compensate for the inaccuracy of the model-implied covariance matrix when the model is incorrectly specified.

The procedure of bootstrap method to estimate the population parameter is as follow. Let \mathbf{S}^* denote the covariance matrix of the bootstrap sample and $\hat{\boldsymbol{\theta}}^*$ denote the argument, minimizing the ML fitting function, $F(\mathbf{S}^*, \boldsymbol{\Sigma}(\cdot))$. $\hat{\boldsymbol{\theta}}^*$ is considered as the estimates of unstandardized population parameters. From the Equation 2 one can obtain the estimates of the standardized population parameters, $\hat{\boldsymbol{\theta}}^{*s}$. A large number (B) of bootstrap samples are drawn from the bootstrap sample space, obtaining $\hat{\boldsymbol{\theta}}_b^{*s}$, where $b= 1, 2, \dots, B$. The point estimates of the standardized population parameters are the means of the corresponding

bootstrap samples, $\hat{\theta}^* = (\sum_{b=1}^B \hat{\theta}_b^{*s}) / B$. The SE estimates of the standardized parameter are the standard deviation of the empirical distribution, $\widehat{se}^* = sd(\hat{\theta}_b^{*s})$. The current study uses two methods to construct CIs: the percentile method and the bias-corrected and accelerated method (BC_a). In the percentile method, the lower $((\alpha/2) 100\%)$ and upper $((1 - (\alpha/2)) 100\%)$ boundaries of $(1 - \alpha) 100\%$ CI are the percentile of the estimates of the standardized population parameter from bootstrap samples $(\hat{\theta}_1^{*s}, \hat{\theta}_2^{*s}, \dots, \hat{\theta}_B^{*s})$. In the BC_a method, extra steps are required. One needs to calculate bias-corrected component, \hat{z}_0 , and acceleration component, $\hat{\alpha}$ in order to construct the BC_a CIs (Efron, Tibshirani, 1993, pp.185-186). Efron and Tibshirani (1993) proposed the BC_a method in order to compensate for the discrepancy in median values between the original sample and the bootstrap samples.

Design of Simulation Study

I compare the performance of these different methods in a SEM model (Figure 1). There are two exogenous latent variables (F_1, F_2) and three endogenous latent variables (F_3, F_4, F_5). Each latent variable has four manifest variables with *standardized* loadings of 0.6, 0.7, 0.8, and 0.9. The error variances of corresponding loads are .64, .51, .36, and .19. All of the measurement errors and disturbances are independent. For the structural coefficients, because they are in the standardized context, I choose values that imply a somewhat medium magnitude in effects. The relationships among the latent variables in the standardized metric are: $cov(F_1, F_2) = 0.5$; $F_3 = 0.3F_1 + 0.7F_2 + D_3$; $F_4 = 0.6F_3 + 0.3F_2 + D_4$; $F_5 = 0.4F_4 + 0.5F_1 + D_5$. In order to obtain model-implied covariance matrix, I assign the standard deviations to 20 observable variables to be [1, 2, ..., 9, 1, 2, ..., 9, 1, 2]. Then I use the MASS package (Venables & Ripley, 2002) in R (R Core Team, 2016) to generate random samples from a multivariate normal distribution using the model-implied covariance matrix as the population covariance matrix; then I use the lavaan package (Rosseel, 2012) in R to fit the model to random samples.

There are four model correctness conditions: one correct model condition, and three misspecification conditions with varying amount of model misfit. In the slightly misfit condition (misfit 1), I remove b_{F1F5} , yielding comparative fit index (CFI) = .98, root mean square error of approximation (RMSEA) = .037, and standardized root mean residual (SRMR) = .052. In the moderate misfit condition (misfit 2), we remove b_{F1F5} and b_{F2F3} , with CFI = .932, RMSEA = .068, and SRMR = .099. In the severe misfit condition (misfit 3), we remove b_{F1F5} , b_{F2F3} , and b_{F1F3} , yielding CFI = .889, RMSEA = .087, and SRMR = .231. Note that the model in misfit 3 condition does not resemble common models in the SEM literatures because $F1$ only correlates with $F2$ and $F3$ and does not have regression effects on the endogenous latent variables. I have tried various other misspecifications but none of them can lead to the desired amount of misfit. Although the model in Misfit Condition 3 is not straightforward to interpret substantively, for the purpose of evaluating model estimation methods it is still statistically viable. There are three sample-size conditions, 300, 500, and 800. Crossing these sample sizes with model-correctness conditions above results in 12 simulation conditions. In each condition, I draw 1500 samples from model implied covariance matrix. Within each sample, there are 1500 bootstrap samples.

Evaluation Criteria

I use different metrics to evaluate the performance of the methods including point estimates, SE estimates, the coverage rate of CIs, and the width of the CIs. I will also examine the convergence rate of NP-B and SP-B.

I consider both absolute bias and relative bias to evaluate point estimates. Absolute bias is calculated by $mean(\hat{\theta}_i) - \theta$, where $\hat{\theta}_i$ is the estimate of θ by a certain method and i is the index for replications. Relative bias is to capture the magnitude relative to the true value of population parameters. It is calculated by $[mean(\hat{\theta}_i) - \theta] / \theta$. For estimation of the SEs, because the true SE of $\hat{\theta}$ is unknown, I approximate it by the standard deviation, $sd(\hat{\theta}_i)$, of

the empirical distribution of $\hat{\theta}_i$. The estimates of the SE for a certain method are calculated by the mean of the SEs of the replications, $mean(\widehat{se}_i)$. Therefore, the bias of the SE estimates is $mean(\widehat{se}_i) - sd(\hat{\theta}_i)$ and the relative bias of the SE estimates is $[mean(\widehat{se}_i) - sd(\hat{\theta}_i)] / sd(\hat{\theta}_i)$. To assess the coverage rate of the CIs, let $\hat{\theta}_{Li}$ and $\hat{\theta}_{Ui}$ denote the lower $(\alpha/2) 100\%$ and upper $((1 - (\alpha/2)) 100\%)$ bounds of the $(1 - \alpha) 100\%$ CI in each replication. For the delta method, the boundaries of the CI can be calculated as $\hat{\theta}_{Li} = \hat{\theta}_i - z_{\alpha/2} \widehat{se}_i$ and $\hat{\theta}_{Ui} = \hat{\theta}_i + z_{\alpha/2} \widehat{se}_i$, where $z_{\alpha/2}$ is the z score of the intended CI coverage. For the bootstrap method, $\hat{\theta}_{Li}$ and $\hat{\theta}_{Ui}$ is the $(\alpha/2) 100\%$ and $(1 - (\alpha/2)) 100\%$ percentile of the empirical distribution. I assess the coverage rates of the CIs by counting the number of times $[\hat{\theta}_{Li}, \hat{\theta}_{Ui}]$ containing θ . This empirical coverage rate is compared with the nominal confidence level: 50%, 60%, 70%, 80%, 90% and 95%. Additionally, I compare the average width of the CIs in each nominal level. If the CI coverage rates are correct, the narrow the CI width is, the more accurate the estimates are. Finally, because some bootstrap samples may not converge, I also calculate the convergence rates of the bootstrap samples by dividing the number of the successfully converged bootstrap samples by the total number of the bootstrap samples.

Results

I evaluate the performance of different methods in the following sequence: the point estimation, the SE estimation, the coverage of CI rates, the width of CIs, and convergence rate. The performance of different methods is evaluated on the estimation of four structural coefficients (b_{F1F2} , b_{F2F3} , b_{F3F4} , and b_{F4F5}) and two factor loadings (b_{F4V1} and b_{F4V2}). Table 1 shows the results of the point estimation. Overall, the discrepancy between the estimated model parameters and the population parameters is small for all methods, with none of the relative bias exceeding 5%. The results are the same for all sample conditions and all model-correctness conditions.

Table 2 shows the results of the SE estimation. The performance of different methods depends on the model-correctness conditions. That is, when model is correctly specified, the estimated SEs of different methods are close to the empirical standard deviations. However, as the model becomes incorrectly specified, the bias and the relative bias increase for some of the parameters in the both delta method and the SP-B method. The estimation of SE by NP-B method remains accurate. This tendency is the same across all sample conditions. Moreover, model correctness does not affect the estimation of all parameters. For example, the estimation of the coefficient between F1 and F2 is less affected than the estimation of the coefficient between F2 and F4 when the model is incorrectly specified. Finally, the model correctness has less influence on the estimation of factor loadings (b_{F4V1} , b_{F4V2}) than on the estimation of structural coefficients.

Figures 2 and 3 show the CI coverage rates for the coefficients, b_{F1F2} and b_{F2F4} , for a sample size of 300. The results are similar for larger sample conditions (i.e., 500, 800). When the model is correctly specified, the CI coverages of all five methods are correct. This result persists if the model is slightly misspecified (misfit 1). However, as the severity of model misfit increases, the differences in CI coverages between different methods emerge. On the one hand, the model misfit does not affect the CI coverages by the NP-B percentile and the NP-B BC_a method. That is, the CI coverages by the NP-B percentile and the NP-B BC_a method are correct in misfit 3 and 4 conditions. On the other hand, model misfit affects the CI coverage rate by the delta method, the SP-B percentile method, and, the SP-B BC_a method. Specifically, in the misfit 2 condition, the delta method, SP-B percentile, and SP-B BC_a method give incorrect CI coverage rate for the coefficients b_{F1F2} and b_{F2F4} . In the misfit 3 condition, the CI coverage of the coefficient b_{F1F2} is correct but the CI coverage of the coefficient b_{F2F4} is incorrect. Figure 4 shows the CI coverage rates for of the factor loading, b_{F4V1} , with sample size of 300. The results are similar for larger sample conditions (i.e., 500,

800). The CI coverages rates of all methods are correct in correctly specified model, the misfit 1, and the misfit 2 conditions. In the misfit 3 condition, the delta method, SP-B percentile method, and the SP-B BC_a method give incorrect CI coverage rates.

Table 3 shows the CI width for b_{F4V2} estimated by different methods. The narrower the CI width is, the more accurate the estimation is. I compare the CI widths only when the CI coverage is accurate. Therefore, because the accuracy of CI coverage rates by the delta method, SP-B percentile method, and SP-B BC_a method are affected by model misfit, their CI widths are not discussed in conditions with model misspecified. For correct-model condition, the CI widths estimated by all methods are similar. That result persists in the misfit 1 condition. In the misfit 2 and misfit 3 conditions, the CI widths estimated by the NP-B percentile method and the NP-B BC_a method are not very different from each other, and the CI widths for other parameters are similar. That is, if methods have correct CI coverage rates, the CI widths are almost the same.

Finally, the convergence rates are assessed in the NP-B method and the SP-B method. All the bootstrap samples converged in both the NP-B method and the SP-B method. An interesting question applied researchers often encounter is NP-B tends to suffer from nonconvergence if misfit is large, and the nonconvergence problem may undermine of usefulness of NP-B. Based on the present simulation study, nonconvergence does not seem to be a drawback NP-B.

In summary, the NP-B method demonstrate superior performance over other methods in terms of the SE estimation, the CI coverage rates, and the convergence rates. Although all methods give accurate point estimates in all sample-size conditions and model-correctness conditions, some of the SE estimates are biased by the delta method and the SP-B method if model is not correctly specified (misfit 2 and misfit 3). In contrast, the SE estimates by the NP-B method is accurate the consistent. The similar results are obtained from the estimation

of the CI coverage rates. In particular, for some of the structural coefficients, the delta method and the SP-B method underestimate the CI coverage rates. But for some factor loadings, they overestimate the CI coverage rates. In terms of the width of the CI coverage, if methods have the correct CI coverage rates, the widths of their CI coverages are the similar.

Discussion

In general, the results from the simulation study suggest that if the model is correctly specified, all methods perform equally well. But if the model has misfit, NP-B outperform the delta method and the SP-B with respect to the SE estimation and the CI coverage rates. This result is consistent across all sample size conditions. This information is useful when using the bootstrap method because the bootstrap is criticized if sample size is small. For example, Ichikawa and Konishi (1995) found that if the sample size is 150, the bootstrap overestimated the empirical SEs. The current study can be considered as an extension of previous study that when sample size is above 300, NP-B estimated SE was correct.

Model misfit plays an important role in the estimation of the SEs. When we use the asymptotic method in software to estimate SEs, there is an assumption that the population is normally distributed and the model is correctly specified. However it is generally accepted that the model is only an approximation to the real world and that no model is absolutely correct. If the model is misfit, then, estimation of the SEs by asymptotic method (e.g., the delta method) could be incorrect. On the other hand, the bootstrap method for estimating the SEs takes both the distribution of the data and the finite sample size into consideration (Bollen & Stine, 1990; Boomsma, 1986; Chatterjee, 1984; Ichikawa & Konishi, 1995; Yung & Benlter, 1996). Therefore, the bootstrap should outperform the delta method if the model is misspecified. The current study confirms that expectation. If the model is correctly specified, every method estimates the SEs correctly. However, as the amount of model misfit increases, the NP-B method gives us better estimation of the SEs overall than the delta method and the

SP-B. Specifically, the SEs estimated by the NP-B are close to the empirical standard deviation for all coefficients. However, the SEs estimated by the delta method and the SP-B method are correct only for some coefficients. For the factor loadings, the estimated SEs by the delta method and the SP-B method are less affected by the model misfit. This result is not consistent with previous studies. Ichikawa and Konishi (1995) showed that the bootstrap method performs differently for different coefficients. In the current study, only the SP-B method shows a similar pattern and the NP-B method accurately estimates SEs for all model parameters.

Which bootstrap method is better at forming confidence intervals, the percentile method or the BC_a method? Bollen and Stine (1990) recommended the use of the percentile method in covariance structure analysis due to its ease of implementation. While Efron and Tibshirani (1993) argued that because of the good mathematical properties (transformation respecting and second-order correctness), the BC_a method is better. If the sample size is infinitely large, the two methods become equivalent. But for finite sample sizes, the BC_a method is believed to approach its asymptotic properties more quickly than the percentile CI method. The results of the current study suggest that the performance of the two methods are similar. The reason might be that because the samples of the current study are drawn from a normal population and because the smallest sample size condition (300) is large enough, the bootstrap samples are representative to the population. Therefore, the BC_a method does not need to adjust any bias and skewness in the bootstrap samples. Finally, the two methods perform similar if model is misspecified.

I conducted this simulation study to compare the performance of the delta method, the NP-B method and the SP-B method in different sample-size conditions and different model-misfit conditions. Based on the results of the simulation, I recommended the use of the NP-B method to estimate model parameters if the model is likely to be misspecified. It performs

best if sample size is above 300. One limitations of the study is that I did not address the issue of the distributional form of the population. This issue is potentially as important as model misfit because normality is one of the assumptions for the asymptotic approach in estimating model parameters. The violation of such an assumption, which is commonly seen in practice, could lead to inaccurate estimation. For example, Ichikawa and Konishi (1995) examined this issue and found that the SEs estimated by the bootstrap method is less biased than asymptotic method if the population is not normally distributed. Therefore, future research should address this issue, while taking model misfit into consideration.

Reference

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solution, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173.
- Bollen, K. A. (1989). Structural equations with latent variables. New York, NY: Wiley.
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 115–140). Oxford, UK: Basil Blackwell.
- Boomsma, A. (1986). On the use of bootstrap and jackknife in covariance structure analysis. In F. de Antoni, N. Lauro, & A. Rizzi (Eds.), *Compstat 86 Proceedings in Computational Statistics* (pp. 205-210). Heidelberg: Physica-Verlag.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge, England: Cambridge University Press.
- Chatterjee, S. (1984). Variance estimation in factor analysis: An application of the bootstrap. *British Journal of Mathematical and Statistical Psychology*, 37, 252–262.
doi:10.1111/bmsp.1984.37.issue-2
- Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A monte carlo study. *British Journal of Mathematical and Statistical Psychology*, 44, 347-357.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317–327. doi:10.1037/0033-2909.105.2.317

- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Method, 1*, 16-29.
- Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. New York, NY: Chapman & Hall.
- Finch J. F. (1997). Effect of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 4*, 87-107.
- Harlow, L. L. (1985). Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A monte carlo study. Unpublished doctoral dissertation, University of California, Los Angeles.
- Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.
- Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika, 60*, 77-93. doi:10.1007/BF02294430
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T. C. (1985). *The Theory and Practice of Econometrics*. New York: John Wiley and Sons.
- Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research, 23*, 69-86.
- 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.
- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*, 107-120.

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. doi:10.1348/000711005X85896

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, NJ: Erlbaum.

Table 1

Point estimates of the standardized population parameters

Poplation			<i>n</i> =300			<i>n</i> =500			<i>n</i> =800		
Correct	Value	Bias	ML	NP-B	SP-B	ML	NP-B	SP-B	ML	NP-B	SP-B
<i>b</i> _{F1F2}	.500	Absolute	.001	.002	.002	.002	.002	.002	.001	.001	-
		Relative	.002	.004	.004	.004	.004	.004	.002	.002	-
<i>b</i> _{F2F4}	.300	Absolute	.003	.003	.005	.007	.007	.001	.002	.002	.003
		Relative	.010	.010	.017	.023	.023	.003	.007	.007	.010
<i>b</i> _{F3F4}	.600	Absolute	.002	.002	.005	.007	.007	.002	.001	.001	.004
		Relative	.003	.003	.008	.012	.012	.003	.002	.002	.007
<i>b</i> _{F4F5}	.400	Absolute	-	-	-	-	-	-	-	-	.001
		Relative	-	-	-	-	-	-	-	-	.003
<i>b</i> _{F4V1}	.600	Absolute	.003	.003	-	.001	.001	-	-	-	.001
		Relative	.005	.005	-	.002	.002	-	-	-	.002
<i>b</i> _{F4V2}	.700	Absolute	.002	.002	.001	.001	.001	.002	-	-	.001
		Relative	.003	.003	.001	.001	.001	.003	-	-	.001
Misfit 1											
<i>b</i> _{F1F2}	.503	Absolute	.002	.003	.001	-	-	.001	-	-	.002
		Relative	.004	.006	.002	-	-	.002	-	-	.004
<i>b</i> _{F2F4}	.249	Absolute	.006	.011	-	.003	.006	.006	.001	.001	.004
		Relative	.024	.044	-	.012	.024	.024	.004	.004	.016
<i>b</i> _{F3F4}	.665	Absolute	.005	.010	.001	.003	.006	.006	.001	.001	.004
		Relative	.008	.015	.002	.005	.009	.009	.002	.002	.006
<i>b</i> _{F4F5}	.701	Absolute	.002	.002	.001	.001	-	.001	-	-	.001
		Relative	.003	.003	.001	.001	-	.001	-	-	.001
<i>b</i> _{F4V1}	.598	Absolute	.001	.002	-	-	-	-	-	-	.001
		Relative	.002	.003	-	-	-	-	-	-	.002
<i>b</i> _{F4V2}	.698	Absolute	.002	.003	.001	-	.001	.001	-	.001	-
		Relative	.003	.004	.001	-	.001	.001	-	.001	-

Note . Bold values are relative bias exceeding 5%; ML=maximum likelihood; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap.

Table 1 continue

Point estimates of the standardized population parameters

	Poplation	Bias	n=300			n=500			n=800		
Misfit 2	Value		ML	NP-B	SP-B	ML	NP-B	SP-B	ML	NP-B	SP-B
b_{F1F2}	.585	Absolute	.001	-	.003	-	-	.001	-	.001	.002
		Relative	.002	-	.005	-	-	.002	-	.002	.003
b_{F2F4}	.407	Absolute	-	.001	-	-	-	.002	.002	.001	.001
		Relative	-	.002	-	-	-	.005	.005	.002	.002
b_{F3F4}	.617	Absolute	.002	.003	.002	.001	.002	.004	.001	.002	.003
		Relative	.003	.005	.003	.002	.003	.006	.002	.003	.005
b_{F4F5}	.680	Absolute	.001	.001	-	.001	.001	.001	.001	.001	.001
		Relative	.001	.001	-	.001	.001	.001	.001	.001	.001
b_{F4V1}	.573	Absolute	.001	.001	-	-	-	.002	-	-	.002
		Relative	.002	.002	-	-	-	.003	-	-	.003
b_{F4V2}	.673	Absolute	.002	.002	.001	.001	.001	.001	-	-	-
		Relative	.003	.003	.001	.001	.001	.001	-	-	-
Misfit 3											
b_{F1F2}	.510	Absolute	.004	.005	.001	-	-	.001	.001	.001	-
		Relative	.008	.010	.002	-	-	.002	.002	.002	-
b_{F2F4}	.479	Absolute	.001	.001	.001	.001	.001	.001	.002	.002	-
		Relative	.002	.002	.002	.002	.002	.002	.004	.004	-
b_{F3F4}	.675	Absolute	.004	.007	.004	.001	.003	.001	.001	.002	.001
		Relative	.006	.010	.006	.001	.004	.001	.001	.003	.001
b_{F4F5}	.628	Absolute	.001	-	.001	.001	.002	.003	-	-	.001
		Relative	.002	-	.002	.002	.003	.005	-	-	.002
b_{F4V1}	.529	Absolute	.001	.002	.001	.002	.002	.002	-	-	.001
		Relative	.002	.004	.002	.004	.004	.004	-	-	.002
b_{F4V2}	.631	Absolute	.002	.003	.001	.001	.002	-	-	-	-
		Relative	.003	.005	.002	.002	.003	-	-	-	-

Note . Bold values are relative bias exceeding 5%; ML=maximum likelihood; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap.

Table 2

Standard errors of the standardized population parameters

Correct	Bias	Empirical		<i>n</i> =300		Empirical		<i>n</i> =500		Empirical		<i>n</i> =800	
		SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B
<i>b</i> _{F1F2}	Absolute	.052	.001	-	.001	.039	.001	.001	.001	.032	.001	.001	.001
	Relative		.019	-	.019		.026	.026	.026		.031	.031	.031
<i>b</i> _{F2F4}	Absolute	.098	-	.006	.003	.078	.003	.001	.001	.061	.002	-	.001
	Relative		-	.061	.031		.038	.013	.013		.033	-	.016
<i>b</i> _{F3F4}	Absolute	.096	.001	.005	.003	.076	.003	.001	.001	.060	.002	.001	.002
	Relative		.010	.052	.031		.039	.013	.013		.033	.017	.033
<i>b</i> _{F4F5}	Absolute	.056	.002	.001	.001	.043	.001	-	.001	.033	-	.001	-
	Relative		.036	.018	.018		.023	-	.023		-	.030	-
<i>b</i> _{F4V1}	Absolute	.041	.001	.001	.001	.030	.001	.001	.001	.025	.001	.001	.001
	Relative		.024	.024	.024		.033	.033	.033		.040	.040	.040
<i>b</i> _{F4V2}	Absolute	.032	.001	.001	.001	.026	.001	.001	.001	.020	-	-	-
	Relative		.031	.031	.031		.038	.038	.038		-	-	-
Misfit 1													
<i>b</i> _{F1F2}	Absolute	.052	.001	-	.001	.038	.001	.002	.001	.032	.001	-	.001
	Relative		.019	-	.019		.026	.053	.026		.031	-	.031
<i>b</i> _{F2F4}	Absolute	.106	.009	.007	.006	.080	.005	.002	.004	.063	.004	-	.003
	Relative		.085	.066	.057		.063	.025	.050		.063	-	.048
<i>b</i> _{F3F4}	Absolute	.103	.010	.007	.007	.078	.006	.001	.005	.061	.005	-	.004
	Relative		.097	.068	.068		.077	.013	.064		.082	-	.066
<i>b</i> _{F4F5}	Absolute	.039	.002	-	.057	.029	.001	.001	.044	.023	.001	-	.034
	Relative		.051	-	1.462		.034	.034	1.517		.043	-	1.478
<i>b</i> _{F4V1}	Absolute	.041	.001	.001	.001	.031	-	-	-	.024	-	-	-
	Relative		.024	.024	.024		-	-	-		-	-	-
<i>b</i> _{F4V2}	Absolute	.032	.001	.001	.001	.026	.001	.001	.001	.020	-	-	-
	Relative		.031	.031	.031		.038	.038	.038		-	-	-

Note . Bold values are relative bias exceeding 5%; Delta=the delta method; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap. 1: Empirical standard deviation

Table 2 continue

Standard errors of the standardized population parameters

Misfit 2	Bias	Empirical	<i>n</i> =300			Empirical	<i>n</i> =500			Empirical	<i>n</i> =800		
		SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B
<i>b</i> _{F1F2}	Absolute	.058	.012	-	.012	.044	.008	-	.008	.035	.007	.001	.007
	Relative		.207	-	.207		.182	-	.182		.200	.029	.200
<i>b</i> _{F2F4}	Absolute	.068	.023	.001	.022	.053	.018	-	.018	.041	.013	.001	.013
	Relative		.338	.015	.324		.340	-	.340		.317	.024	.317
<i>b</i> _{F3F4}	Absolute	.063	.021	-	.021	.049	.017	.001	.016	.038	.013	-	.012
	Relative		.333	-	.333		.347	.020	.327		.342	-	.316
<i>b</i> _{F4F5}	Absolute	.040	.001	.001	.001	.032	.002	.001	.002	.024	-	.001	-
	Relative		.025	.025	.025		.063	.031	.063		-	.042	-
<i>b</i> _{F4V1}	Absolute	.041	.001	-	.001	.031	.002	-	.001	.024	.002	.001	.002
	Relative		.024	-	.024		.065	-	.032		.083	.042	.083
<i>b</i> _{F4V2}	Absolute	.034	.001	-	.001	.026	.001	.001	.001	.021	-	-	-
	Relative		.029	-	.029		.038	.038	.038		-	-	-
Misfit 3													
<i>b</i> _{F1F2}	Absolute	.052	.002	-	.002	.041	.002	.001	.002	.031	-	.001	-
	Relative		.038	-	.038		.049	.024	.049		-	.032	-
<i>b</i> _{F2F4}	Absolute	.075	.032	.002	.031	.060	.026	-	.026	.047	.020	-	.020
	Relative		.427	.027	.413		.433	-	.433		.426	-	.426
<i>b</i> _{F3F4}	Absolute	.065	.027	-	.027	.050	.021	-	.021	.039	.016	.001	.016
	Relative		.415	-	.415		.420	-	.420		.410	.026	.410
<i>b</i> _{F4F5}	Absolute	.039	.005	-	.005	.030	.004	-	.004	.024	.003	-	.003
	Relative		.128	-	.128		.133	-	.133		.125	-	.125
<i>b</i> _{F4V1}	Absolute	.040	.005	.001	.005	.031	.004	.001	.004	.024	.004	-	.004
	Relative		.125	.025	.125		.129	.032	.129		.167	-	.167
<i>b</i> _{F4V2}	Absolute	.034	.005	-	.005	.026	.004	.001	.004	.021	.003	-	.003
	Relative		.147	-	.147		.154	.038	.154		.143	-	.143

Note . Bold values are relative bias exceeding 5%; Delta=the delta method; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap. 1: Empirical standard deviation

Table 1

Point estimates of the standardized population parameters

Poplation			<i>n</i> =300			<i>n</i> =500			<i>n</i> =800		
Correct	Value	Bias	ML	NP-B	SP-B	ML	NP-B	SP-B	ML	NP-B	SP-B
<i>b</i> _{F1F2}	.500	Absolute	.001	.002	.002	.002	.002	.002	.001	.001	-
		Relative	.002	.004	.004	.004	.004	.004	.002	.002	-
<i>b</i> _{F2F4}	.300	Absolute	.003	.003	.005	.007	.007	.001	.002	.002	.003
		Relative	.010	.010	.017	.023	.023	.003	.007	.007	.010
<i>b</i> _{F3F4}	.600	Absolute	.002	.002	.005	.007	.007	.002	.001	.001	.004
		Relative	.003	.003	.008	.012	.012	.003	.002	.002	.007
<i>b</i> _{F4F5}	.400	Absolute	-	-	-	-	-	-	-	-	.001
		Relative	-	-	-	-	-	-	-	-	.003
<i>b</i> _{F4V1}	.600	Absolute	.003	.003	-	.001	.001	-	-	-	.001
		Relative	.005	.005	-	.002	.002	-	-	-	.002
<i>b</i> _{F4V2}	.700	Absolute	.002	.002	.001	.001	.001	.002	-	-	.001
		Relative	.003	.003	.001	.001	.001	.003	-	-	.001
Misfit 1											
<i>b</i> _{F1F2}	.503	Absolute	.002	.003	.001	-	-	.001	-	-	.002
		Relative	.004	.006	.002	-	-	.002	-	-	.004
<i>b</i> _{F2F4}	.249	Absolute	.006	.011	-	.003	.006	.006	.001	.001	.004
		Relative	.024	.044	-	.012	.024	.024	.004	.004	.016
<i>b</i> _{F3F4}	.665	Absolute	.005	.010	.001	.003	.006	.006	.001	.001	.004
		Relative	.008	.015	.002	.005	.009	.009	.002	.002	.006
<i>b</i> _{F4F5}	.701	Absolute	.002	.002	.001	.001	-	.001	-	-	.001
		Relative	.003	.003	.001	.001	-	.001	-	-	.001
<i>b</i> _{F4V1}	.598	Absolute	.001	.002	-	-	-	-	-	-	.001
		Relative	.002	.003	-	-	-	-	-	-	.002
<i>b</i> _{F4V2}	.698	Absolute	.002	.003	.001	-	.001	.001	-	.001	-
		Relative	.003	.004	.001	-	.001	.001	-	.001	-

Note . Bold values are relative bias exceeding 5%; ML=maximum likelihood; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap.

Table 1 continue

Point estimates of the standardized population parameters

	Poplation	Bias	n=300			n=500			n=800		
Misfit 2	Value		ML	NP-B	SP-B	ML	NP-B	SP-B	ML	NP-B	SP-B
b_{F1F2}	.585	Absolute	.001	-	.003	-	-	.001	-	.001	.002
		Relative	.002	-	.005	-	-	.002	-	.002	.003
b_{F2F4}	.407	Absolute	-	.001	-	-	-	.002	.002	.001	.001
		Relative	-	.002	-	-	-	.005	.005	.002	.002
b_{F3F4}	.617	Absolute	.002	.003	.002	.001	.002	.004	.001	.002	.003
		Relative	.003	.005	.003	.002	.003	.006	.002	.003	.005
b_{F4F5}	.680	Absolute	.001	.001	-	.001	.001	.001	.001	.001	.001
		Relative	.001	.001	-	.001	.001	.001	.001	.001	.001
b_{F4V1}	.573	Absolute	.001	.001	-	-	-	.002	-	-	.002
		Relative	.002	.002	-	-	-	.003	-	-	.003
b_{F4V2}	.673	Absolute	.002	.002	.001	.001	.001	.001	-	-	-
		Relative	.003	.003	.001	.001	.001	.001	-	-	-
Misfit 3											
b_{F1F2}	.510	Absolute	.004	.005	.001	-	-	.001	.001	.001	-
		Relative	.008	.010	.002	-	-	.002	.002	.002	-
b_{F2F4}	.479	Absolute	.001	.001	.001	.001	.001	.001	.002	.002	-
		Relative	.002	.002	.002	.002	.002	.002	.004	.004	-
b_{F3F4}	.675	Absolute	.004	.007	.004	.001	.003	.001	.001	.002	.001
		Relative	.006	.010	.006	.001	.004	.001	.001	.003	.001
b_{F4F5}	.628	Absolute	.001	-	.001	.001	.002	.003	-	-	.001
		Relative	.002	-	.002	.002	.003	.005	-	-	.002
b_{F4V1}	.529	Absolute	.001	.002	.001	.002	.002	.002	-	-	.001
		Relative	.002	.004	.002	.004	.004	.004	-	-	.002
b_{F4V2}	.631	Absolute	.002	.003	.001	.001	.002	-	-	-	-
		Relative	.003	.005	.002	.002	.003	-	-	-	-

Note . Bold values are relative bias exceeding 5%; ML=maximum likelihood; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap.

Table 2

Standard errors of the standardized population parameters

Correct	Bias	Empirical		<i>n</i> =300		Empirical		<i>n</i> =500		Empirical		<i>n</i> =800	
		SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B
<i>b_{F1F2}</i>	Absolute	.052	.001	-	.001	.039	.001	.001	.001	.032	.001	.001	.001
	Relative		.019	-	.019		.026	.026	.026		.031	.031	.031
<i>b_{F2F4}</i>	Absolute	.098	-	.006	.003	.078	.003	.001	.001	.061	.002	-	.001
	Relative		-	.061	.031		.038	.013	.013		.033	-	.016
<i>b_{F3F4}</i>	Absolute	.096	.001	.005	.003	.076	.003	.001	.001	.060	.002	.001	.002
	Relative		.010	.052	.031		.039	.013	.013		.033	.017	.033
<i>b_{F4F5}</i>	Absolute	.056	.002	.001	.001	.043	.001	-	.001	.033	-	.001	-
	Relative		.036	.018	.018		.023	-	.023		-	.030	-
<i>b_{F4V1}</i>	Absolute	.041	.001	.001	.001	.030	.001	.001	.001	.025	.001	.001	.001
	Relative		.024	.024	.024		.033	.033	.033		.040	.040	.040
<i>b_{F4V2}</i>	Absolute	.032	.001	.001	.001	.026	.001	.001	.001	.020	-	-	-
	Relative		.031	.031	.031		.038	.038	.038		-	-	-
Misfit 1													
<i>b_{F1F2}</i>	Absolute	.052	.001	-	.001	.038	.001	.002	.001	.032	.001	-	.001
	Relative		.019	-	.019		.026	.053	.026		.031	-	.031
<i>b_{F2F4}</i>	Absolute	.106	.009	.007	.006	.080	.005	.002	.004	.063	.004	-	.003
	Relative		.085	.066	.057		.063	.025	.050		.063	-	.048
<i>b_{F3F4}</i>	Absolute	.103	.010	.007	.007	.078	.006	.001	.005	.061	.005	-	.004
	Relative		.097	.068	.068		.077	.013	.064		.082	-	.066
<i>b_{F4F5}</i>	Absolute	.039	.002	-	.057	.029	.001	.001	.044	.023	.001	-	.034
	Relative		.051	-	1.462		.034	.034	1.517		.043	-	1.478
<i>b_{F4V1}</i>	Absolute	.041	.001	.001	.001	.031	-	-	-	.024	-	-	-
	Relative		.024	.024	.024		-	-	-		-	-	-
<i>b_{F4V2}</i>	Absolute	.032	.001	.001	.001	.026	.001	.001	.001	.020	-	-	-
	Relative		.031	.031	.031		.038	.038	.038		-	-	-

Note . Bold values are relative bias exceeding 5%; Delta=the delta method; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap. 1: Empirical standard deviation

Table 2 Continue

Standard errors of the standardized population parameters

Misfit 2	Bias	Empirical	<i>n</i> =300			Empirical	<i>n</i> =500			Empirical	<i>n</i> =800		
		SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B	SD ¹	Delta	NP-B	SP-B
<i>b</i> _{F1F2}	Absolute	.058	.012	-	.012	.044	.008	-	.008	.035	.007	.001	.007
	Relative		.207	-	.207		.182	-	.182		.200	.029	.200
<i>b</i> _{F2F4}	Absolute	.068	.023	.001	.022	.053	.018	-	.018	.041	.013	.001	.013
	Relative		.338	.015	.324		.340	-	.340		.317	.024	.317
<i>b</i> _{F3F4}	Absolute	.063	.021	-	.021	.049	.017	.001	.016	.038	.013	-	.012
	Relative		.333	-	.333		.347	.020	.327		.342	-	.316
<i>b</i> _{F4F5}	Absolute	.040	.001	.001	.001	.032	.002	.001	.002	.024	-	.001	-
	Relative		.025	.025	.025		.063	.031	.063		-	.042	-
<i>b</i> _{F4V1}	Absolute	.041	.001	-	.001	.031	.002	-	.001	.024	.002	.001	.002
	Relative		.024	-	.024		.065	-	.032		.083	.042	.083
<i>b</i> _{F4V2}	Absolute	.034	.001	-	.001	.026	.001	.001	.001	.021	-	-	-
	Relative		.029	-	.029		.038	.038	.038		-	-	-
Misfit 3													
<i>b</i> _{F1F2}	Absolute	.052	.002	-	.002	.041	.002	.001	.002	.031	-	.001	-
	Relative		.038	-	.038		.049	.024	.049		-	.032	-
<i>b</i> _{F2F4}	Absolute	.075	.032	.002	.031	.060	.026	-	.026	.047	.020	-	.020
	Relative		.427	.027	.413		.433	-	.433		.426	-	.426
<i>b</i> _{F3F4}	Absolute	.065	.027	-	.027	.050	.021	-	.021	.039	.016	.001	.016
	Relative		.415	-	.415		.420	-	.420		.410	.026	.410
<i>b</i> _{F4F5}	Absolute	.039	.005	-	.005	.030	.004	-	.004	.024	.003	-	.003
	Relative		.128	-	.128		.133	-	.133		.125	-	.125
<i>b</i> _{F4V1}	Absolute	.040	.005	.001	.005	.031	.004	.001	.004	.024	.004	-	.004
	Relative		.125	.025	.125		.129	.032	.129		.167	-	.167
<i>b</i> _{F4V2}	Absolute	.034	.005	-	.005	.026	.004	.001	.004	.021	.003	-	.003
	Relative		.147	-	.147		.154	.038	.154		.143	-	.143

Note . Bold values are relative bias exceeding 5%; Delta=the delta method; NP-B=nonparametric bootstrap; SP-B=semi-parametric bootstrap. 1: Empirical standard deviation

Table 3

Width of the empirical CI coverage with a nominal level of 95%

	<i>n</i> = 300					<i>n</i> = 500					<i>n</i> = 800				
	Delta	NP-B percentile	NP-B BCa percentile	SP-B percentile	SP-B Bca percentile	Delta	NP-B percentile	NP-B BCa percentile	SP-B percentile	SP-B Bca percentile	Delta	NP-B percentile	NP-B BCa percentile	SP-B percentile	SP-B Bca percentile
Correct															
<i>b_{F1F2}</i>	.199	.201	.202	.199	.199	.155	.156	.156	.154	.155	.122	.122	.122	.122	.122
<i>b_{F2F4}</i>	.384	.409	.409	.398	.398	.294	.303	.303	.301	.301	.233	.237	.237	.235	.235
<i>b_{F3F4}</i>	.374	.398	.398	.387	.387	.285	.295	.295	.292	.292	.226	.230	.230	.228	.228
<i>b_{F4F5}</i>	.213	.217	.217	.213	.214	.165	.167	.167	.166	.166	.131	.132	.132	.131	.131
<i>b_{F4V1}</i>	.157	.157	.157	.156	.156	.121	.121	.121	.121	.121	.096	.095	.096	.095	.096
<i>b_{F4V2}</i>	.128	.129	.129	.127	.128	.099	.099	.099	.098	.099	.078	.078	.078	.078	.078
Misfit 1															
<i>b_{F1F2}</i>	.200	.203	.203	.198	.199	.154	-	-	.154	.154	.122	.123	.123	.122	.122
<i>b_{F2F4}</i>	-	-	-	-	-	-	.322	.321	-	-	-	.248	.248	.233	.233
<i>b_{F3F4}</i>	-	-	-	-	-	-	.311	.311	-	-	-	.240	.240	-	-
<i>b_{F4F5}</i>	-	.150	.151	-	-	.111	.115	.116	-	-	.088	.091	.091	-	-
<i>b_{F4V1}</i>	.156	.157	.157	.155	.156	.121	.121	.121	.121	.121	.096	.096	.096	.095	.095
<i>b_{F4V2}</i>	.128	.129	.130	.128	.128	.099	.099	.099	.099	.099	.078	.078	.079	.078	.078
Misfit 2															
<i>b_{F1F2}</i>	-	.227	.227	-	-	-	.172	.172	-	-	-	.135	.135	-	-
<i>b_{F2F4}</i>	-	.269	.269	-	-	-	.206	.206	-	-	-	.162	.162	-	-
<i>b_{F3F4}</i>	-	.248	.248	-	-	-	.189	.189	-	-	-	.150	.150	-	-
<i>b_{F4F5}</i>	.152	.158	.159	.153	.153	-	.123	.123	-	-	-	.097	.097	.093	.093
<i>b_{F4V1}</i>	.165	.158	.159	.164	.164	-	.123	.123	.127	.127	-	.097	.097	-	-
<i>b_{F4V2}</i>	.136	.133	.134	.136	.136	.107	.104	.104	.106	.106	.084	.082	.082	.084	.084
Misfit 3															
<i>b_{F1F2}</i>	.198	.203	.203	.197	.197	.153	.156	.156	.152	.152	.121	.123	.123	.121	.121
<i>b_{F2F4}</i>	-	.301	.301	-	-	-	.234	.234	-	-	-	.185	.185	-	-
<i>b_{F3F4}</i>	-	.254	.255	-	-	-	.196	.196	-	-	-	.155	.155	-	-
<i>b_{F4F5}</i>	-	.154	.155	-	-	-	.118	.119	-	-	-	.094	.094	-	-
<i>b_{F4V1}</i>	-	.153	.154	-	-	-	.119	.119	-	-	-	.094	.094	-	-
<i>b_{F4V2}</i>	-	.134	.135	-	-	-	.104	.104	-	-	-	.082	.082	-	-

Figure 1

Structural part of the SEM model

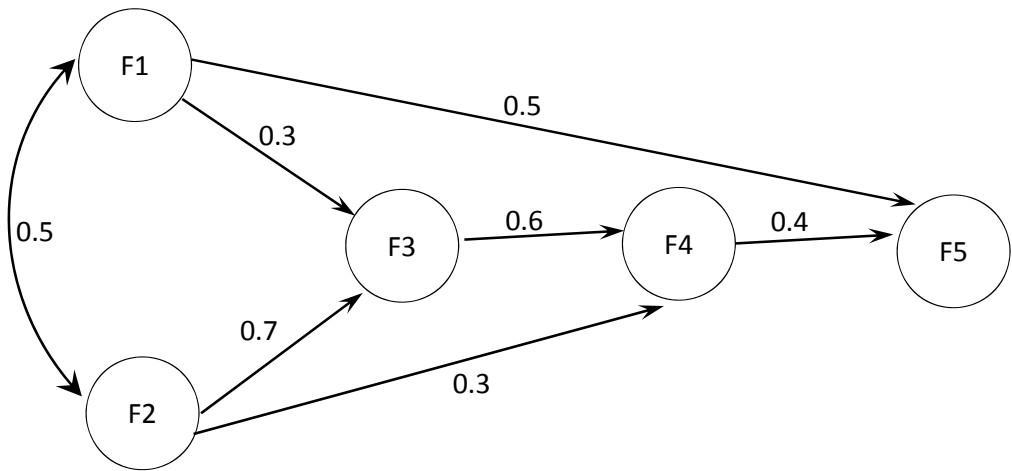


Figure 2

Empirical CI coverage rates for b_{F1F2} with sample size of 300

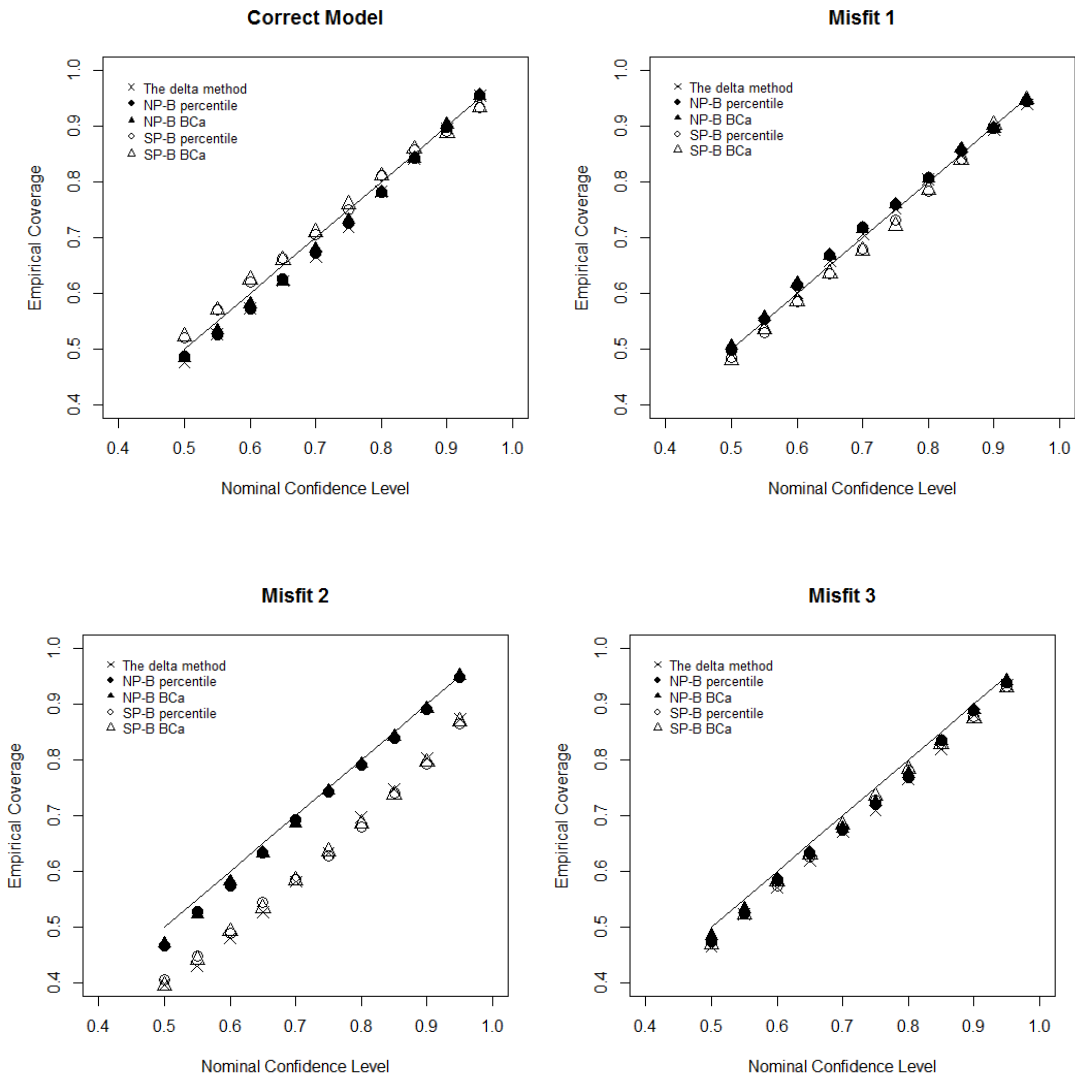


Figure 3
Empirical CI coverage rates for b_{F2F4} with sample size of 300

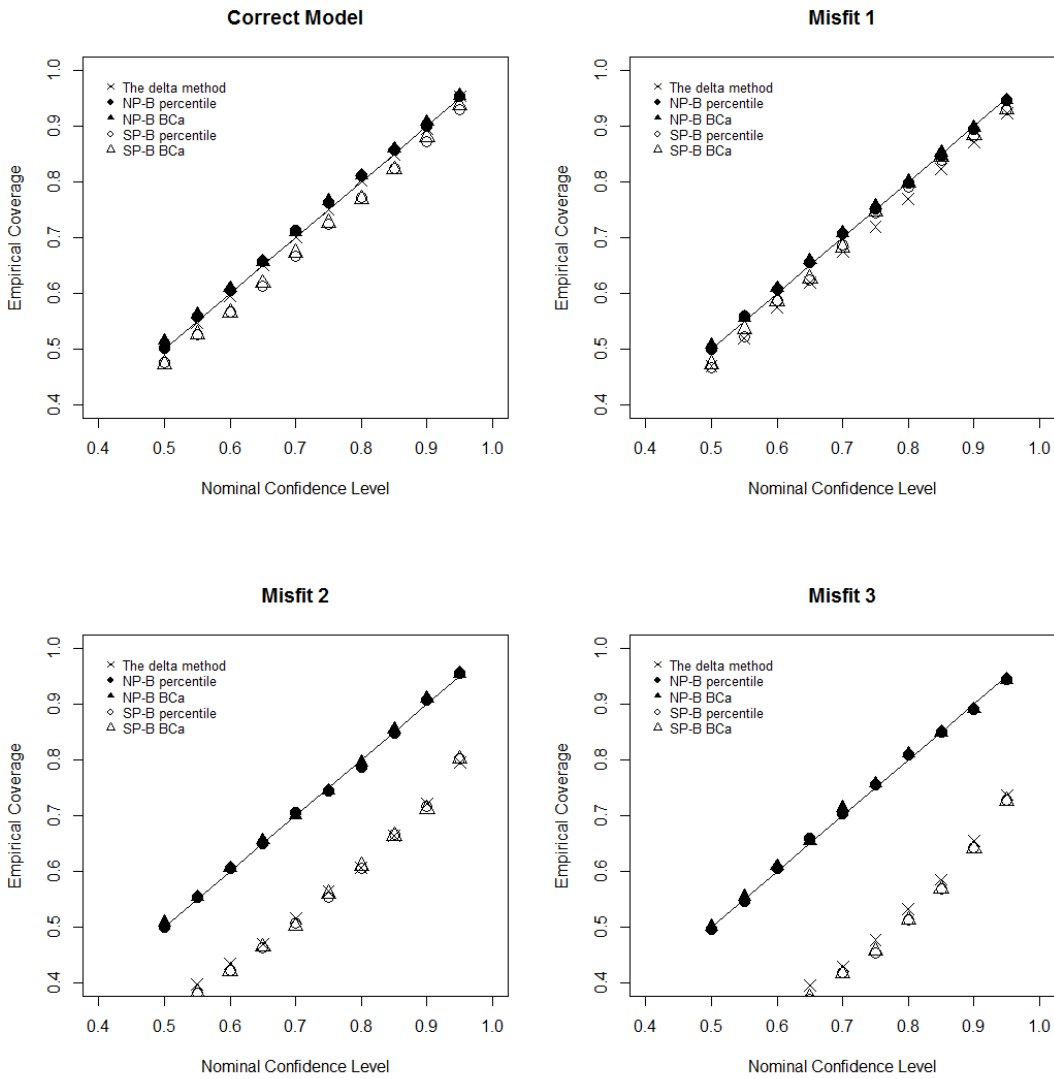


Figure 4

Empirical CI coverage rates for b_{F4VI} with sample size of 300

