Structural Equation Modeling: A Multidisciplinary Journal

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

A Corrected Goodness-of-Fit Index (CGFI) for Model Evaluation in Structural Equation Modeling

Kai Wang, Ying Xu, Chaolong Wang, Ming Tan & Pingyan Chen

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2019.1695213

A Corrected Goodness-of-Fit Index (CGFI) for Model Evaluation in Structural Equation Modeling

Kai Wang[1,2], Ying Xu[1,2], Chaolong Wang[3], Ming Tan[4], and Pingyan Chen[1,2]

[1]Southern Medical University; [2]National Clinical Research Center for Kidney Disease; [3]Huazhong University of Science and Technology Tongji Medical College; [4]Georgetown University Medical Center/Lombardi CCC

ABSTRACT

We propose a Corrected Goodness-of-Fit Index (CGFI) for model evaluation in Structural Equation Modeling (SEM). The CGFI is essentially a corrected index that takes into account model complexity and downward bias caused by small sample size. Using simulations based on pre-set SEM models, we compared the properties of CGFI, Goodness-of-Fit (GFI), and Adjusted Goodness-of-Fit Index (AGFI) under different settings of sample size, estimation method, magnitude of factor loadings, model complexity, and types and degrees of model misspecification. We find that the CGFI is more stable across different sample sizes and much more sensitive to detect model misspecification than the GFI and AGFI. We recommend a critical value of 0.90 for the proposed CGFI to evaluate the goodness of fit of SEM. Our proposed CGFI is easy to implement and can serve as a useful supplementary fit index to existing ones.

KEYWORDS

Structural equation modeling; Goodness-of-Fit Index; Monte Carlo simulation; sample size

Introduction

Structural Equation Modeling (SEM) involves model specification, parameter estimation and model evaluation based on how well the constructed model explains the pattern of the observed covariance matrix (Shevlin & Miles, 1998). As a vital part of SEM, various indexes have been proposed to assess the overall model fit, including the Root Mean Squared Error of Approximation (RMSEA, Steiger & Lind, 1980), the Standardized Root Mean squared Residual (SRMR, Joreskog & Sorbom, 1981), the Goodness-of-Fit Index (GFI, Joreskog & Sorbom, 1982, 1984), and so on (Bone, Sharma, & Shimp, 1989; Lei & Lomax, 2005; Saris, Satorra, & Sörbom, 1987; Sivo, Fan, Witta, & Willse, 2006; Yuan, Chan, Marcoulides, & Bentler, 2015). Ding, Velicer, and Harlow (1995) and Fan, Thompson, and Wang (1999) suggested that an ideal fit index should be: ① sensitive to all kinds of model misspecifications, such as the measurement model (factor loading[s]) misspecification, structural model (factor covariance[s]) misspecification, etc.; ② stable across different sample sizes, data distributions and estimation methods; ③ one that can penalize model complexity.

Among those indexes mentioned above, the GFI proposed by Joreskog and Sorbom (1982) is commonly used in SEM; however, Bollen, Stine, Bollen, and Stine (1993) have pointed out that the GFI tends to be greater than 1 when the fitted model is seriously misspecified. Gerbing and Anderson (1992) and Marsh, Balla, and McDonald (1988) found that the GFI is positively correlated with sample size, and has a moderate downward bias when the sample size is small, while Marsh and Balla (1994) concluded that it is reasonably stable with the increasing of sample size. Furthermore, Fan et al. (1999) and La Du and Tanaka (1989) demonstrated that the GFI is affected by the estimation method, especially in a misspecified model. To improve the performance of the GFI, Joreskog and Sorbom (1984) and Mulaik et al. (1989) proposed Adjusted Goodness-of-Fit Index (AGFI) and Parsimony unbiased Goodness-of-Fit Index (PGFI) respectively, by penalizing the degrees of freedom of model and free parameters to overcome the downward bias and weak stableness of the GFI. Unfortunately, the PGFI is not recommended in applied research because it is less sensitive to model misspecification than the GFI, and has relatively serious downward bias (Fan & Sivo, 2005; Fan et al., 1999; Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1998; Schreiber, 2008; Schreiber, Nora, Stage, Barlow, & King, 2006; Shevlin & Miles, 1998; Sivo et al., 2006). In the present study, we propose a Corrected Goodness-of-Fit Index (CGFI), which takes into account the model complexity and the downward bias caused by small sample size, to improve the evaluation of model fitness in comparison to the existing GFI and AGFI. We will show, based on both simulated and empirical data, that the proposed CGFI constitutes a stable and sensible alternative to the GFI and AGFI.

The paper proceeds as follows. We present the GFI and AGFI in Section “A brief introduction of the GFI and AGFI”. Then, we detailed the idea of the proposed CGFI and the design of our simulation studies in Sections “The Proposed

CONTACT Pingyan Chen chenpy99@126.com Department of Biostatistics, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515 China.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hsem.
Supplemental data for this article can be accessed on the publisher’s website.

Corrected GFI" and "Simulation studies", respectively. In Section "Simulation process" and "Assessment of fit indexes", we depict the concrete steps to perform simulation studies and explain how to interpret the simulation results. We comprehensively assess the properties of the CGFI, GFI, and AGFI based on these results and empirical examples in Sections "Results," "Discussion," and "Conclusion".

Methods

A brief introduction of the GFI and AGFI

Joreskog and Sorbom (1982) proposed the GFI to measure (1) the discrepancy between the sample covariance matrix (S) and the estimated covariance matrix $\hat{\Sigma}$, and (2) how much better the proposed model fits as compared to a null model. However, the GFI increases as more parameters added into a model. To solve this problem, Joreskog and Sorbom (1984) formulated an adjusted GFI (AGFI). The definitions and properties of the GFI and AGFI are shown in Table 1.

The proposed corrected GFI

Theoretically, the GFI is the proportion of variance accounted by the estimated population covariance (Joreskog & Sorbom, 1984). It is a weighted coefficient of determination, theoretically ranges from 0 (poor fit) to 1 (perfect fit), and is considered to be satisfactory when the value is greater than 0.90 (Tanaka & Huba, 1985). Some studies pointed out that the GFI has a positive relationship with sample size and has some bias under the true model (Shevlin & Miles, 1998, 1998). We confirmed the positive correlation between the GFI and sample size by simulation, as shown in Figure 1. The GFI is artificially biased downwards when the sample size is small. Therefore, to adjust for this downward bias, we propose to add a correction factor $\frac{1}{N}$ proportionated by the model complexity adjusting factor $\frac{k(k+1)}{p}$ to the GFI. Then the corrected GFI (CGFI) is defined as:

$$CGFI = GFI + \frac{k(k+1)}{p} \times \frac{1}{N} \tag{1}$$

where k is the number of observed variables, p is the number of free parameters and N is the sample size.

In SEM, the degrees of freedom and the number of free parameters (p) for the test model or the theoretical model (df_T) have a verified functional relationship as follows:

$$df_T = \frac{k(k+1)}{2} - p \tag{2}$$

Thus, an alternative form of CGFI is:

$$CGFI = GFI + \frac{2}{1 - \frac{2df_T}{k(k+1)}} \times \frac{1}{N} \tag{3}$$

Table 1. The definitions and properties of the GFI and AGFI.

Indexes	Algebraic definition	Range	Critical value
GFI (Goodness-of-Fit Index)	$1 - \frac{tr\left(\Sigma^{-1}S - I\right)^2}{tr\left(\Sigma^{-1}S\right)^2}$	(0.00, 1.00) Min value may < 0	0.90
AGFI (Adjusted Goodness-of-Fit Index)	$1 - \frac{k(k+1)}{2df_{test}}(1 - GFI)$		

S = the sample covariance matrix, Σ = the estimated covariance matrix, tr = trace of a matrix, I = identity matrix, df_{test}= the degrees of test model, k = the number of observed variables.

Simulation studies

We performed two Monte Carlo simulation studies under different settings to evaluate the properties of the CGFI, GFI, and AGFI. Data sets were generated from a standard multivariate normal distribution, and each simulation setting was repeated 1000 times.

Simulation one

This simulation was to compare the properties of the CGFI, GFI, and AGFI under different estimation methods, complexities of model, and sample sizes (Table 2). Two kinds of

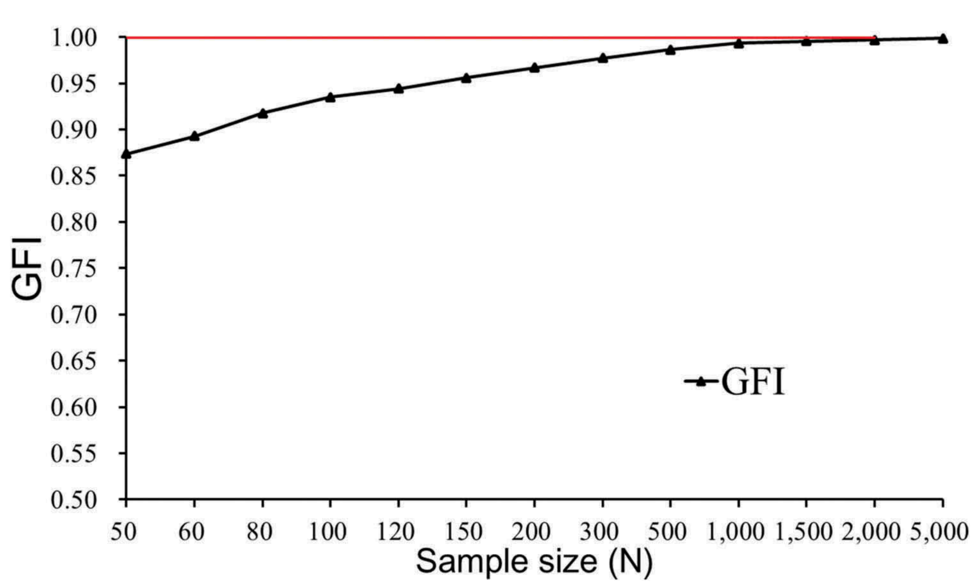
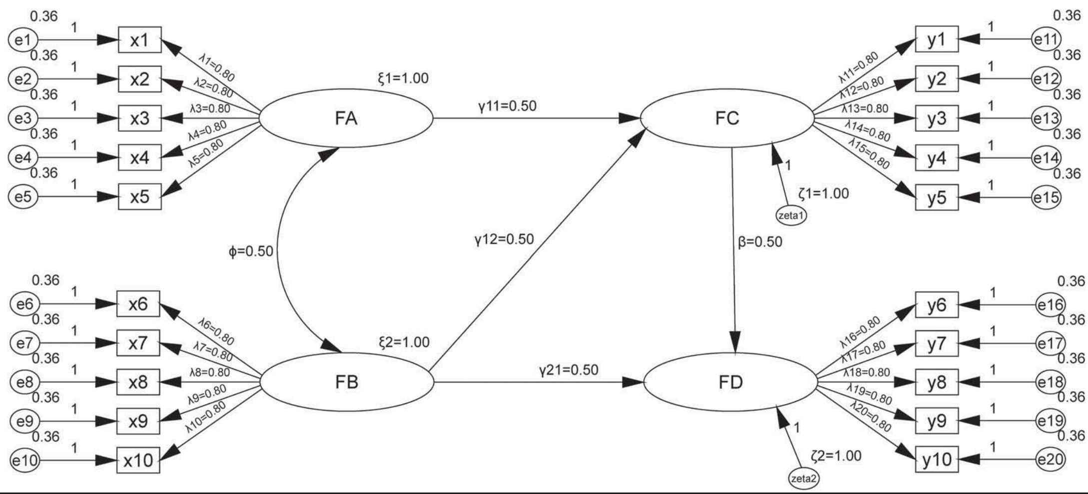


Figure 1. The trend of the GFI (Goodness-of-Fit) with the change of sample size in true model based on a pre-set SEM model.

Table 2. The basic information of the complexity of model and sample size settings.

Factors	Indicators[a]	Min N/k	Sample size (N)
2	3	5	30,40,50,80,100,120,150,200,300,500,1000,1500,2000,5000
2	4	5	40,50,60,80,100,120,150,200,300,500,1000,1500,2000,5000
2	5	5	50,60,80,100,120,150,200,300,500,1000,1500,2000,5000
3	3	5.6	50,60,80,100,120,150,200,300,500,1000,1500,2000,5000
3	4	5	60,80,100,120,150,200,300,500,1000,1500,2000,5000
3	5	5.3	80,100,120,150,200,300,500,1000,1500,2000,5000
4	3	5	60,80,100,120,150,200,300,500,1000,1500,2000,5000
4	4	5	80,100,120,150,200,300,500,1000,1500,2000,5000
4	5	5	100,120,150,200,300,500,1000,1500,2000,5000

[a]The numbers of indicators per factor, N = sample size, k = the number of indicators (or observed variables) in the model. The study only considered the balanced design.



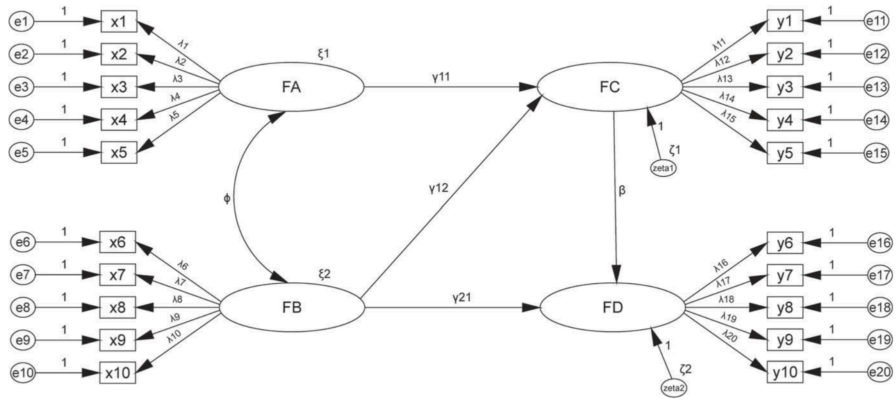
$$\Lambda_x = \begin{bmatrix} 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \end{bmatrix} \quad \theta_\delta = \begin{bmatrix} 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 \end{bmatrix}$$

$$\Lambda_y = \begin{bmatrix} 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.80 & 0.00 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \\ 0.00 & 0.80 \end{bmatrix} \quad \theta_\varepsilon = \begin{bmatrix} 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.36 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0.50 & 0.50 \\ 0.00 & 0.50 \end{bmatrix} \quad \phi = \begin{bmatrix} 1.00 & 0.50 \\ 0.50 & 1.00 \end{bmatrix} \quad \mathrm{B} = \begin{bmatrix} 0.00 & 0.00 \\ 0.50 & 0.00 \end{bmatrix} \quad \Psi = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix}$$

Figure 2. The SEM model with four factors and five indicators per factor, as well as the population parameters.

FA and FB are exogenous latent variables (factors); FC and FD are endogenous latent variables (factors); x1-x10 and y1-y10 are the x and y indicators, respectively; e1-e20 are error terms of the indicators; Λ_x (lambda-x) and Λ_y (lambda-y) are the parameter matrix of pattern coefficients for the x and y indicators, respectively; θ_δ (theta-delta) and θ_ε (theta-epsilon) are the covariance matrices of error terms of the x and y indicators, respectively; Γ is the parameter matrix for direct effects of exogenous factors on endogenous factors and B is the parameter matrix for direct effects of endogenous factors on each other; ϕ and Ψ are the covariance matrix for the disturbances of exogenous and endogenous factors, respectively.



$$\Lambda_x = \begin{bmatrix} \lambda_1 & 0.00 \\ \lambda_2 & 0.00 \\ \lambda_3 & 0.00 \\ \lambda_4 & 0.00 \\ \lambda_5 & 0.00 \\ 0.00 & \lambda_6 \\ 0.00 & \lambda_7 \\ 0.00 & \lambda_8 \\ 0.00 & \lambda_9 \\ 0.00 & \lambda_{10} \end{bmatrix} \quad \Lambda_y = \begin{bmatrix} \lambda_{11} & 0.00 \\ \lambda_{12} & 0.00 \\ \lambda_{13} & 0.00 \\ \lambda_{14} & 0.00 \\ \lambda_{15} & 0.00 \\ 0.00 & \lambda_{16} \\ 0.00 & \lambda_{17} \\ 0.00 & \lambda_{18} \\ 0.00 & \lambda_{19} \\ 0.00 & \lambda_{20} \end{bmatrix}$$

$$\theta_\delta = \begin{bmatrix} 1-{\lambda_1}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1-{\lambda_2}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1-{\lambda_3}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1-{\lambda_4}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_5}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_6}^2 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_7}^2 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_8}^2 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_9}^2 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{10}}^2 \end{bmatrix}$$

$$\theta_\varepsilon = \begin{bmatrix} 1-{\lambda_{11}}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1-{\lambda_{12}}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1-{\lambda_{13}}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1-{\lambda_{14}}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{15}}^2 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{16}}^2 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{17}}^2 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{18}}^2 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{19}}^2 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1-{\lambda_{20}}^2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0.50 & 0.50 \\ 0.50 & 0.50 \end{bmatrix} \quad \Phi = \begin{bmatrix} 1.00 & 0.50 \\ 0.50 & 1.00 \end{bmatrix} \quad \mathrm{B} = \begin{bmatrix} 0.00 & 0.00 \\ 0.50 & 0.00 \end{bmatrix} \quad \Psi = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix}$$

Figure 3. The figure of theoretical model.

FA and FB are exogenous latent variables (factors); FC and FD are endogenous latent variables (factors); x1-x10 and y1-y10 are the x and y indicators, respectively; e1-e20 are error terms of the indicators.

Table 3. The detailed information of model specification.

Model ID	Model specification	Detailed information
1	True model	All parameters were consistent with theoretical model
2	The slightly misspecification of measurement model	$\lambda_5 = 0$
3	The severely misspecification of measurement model	$\lambda_5 = 0$, $\lambda_7 = 0$, $\lambda_{12} = 0$
4	The slightly misspecification of structural model	$\phi_{12} = 0$, $\phi_{21} = 0$, $\gamma_{11} = 0$
5	The severely misspecification of structural model	$\phi_{12} = 0$, $\phi_{21} = 0$, $\gamma_{11} = 0$, $\gamma_{12} = 0$, $\gamma_{21} = 0$, $\beta = 0$

parameter estimation methods, Maximum Likelihood (ML) and Generalized Least Square (GLS), were chosen.

It is difficult to define the complexity of the SEM because it depends on the number of indicators, the number of factors, as well as the relationship among them (Fan et al., 1999). To illustrate the main points, we utilized a balanced design with 2–4 factors and each with 3–5 indicators, resulting in a total of nine theoretical models (Table 2). We set all the factor loadings to 0.80 and correlation coefficients between factors to 0.50. As an illustration, Figure 2 shows one theoretical model with four factors and five indicators under each factor, as well as corresponding population model parameters. The other eight theoretical models are listed in the supplementary material.

The small sample size is also hard to define clearly. For instance, the sample size of 100 may be relatively small for an SEM model with four factors and five indicators per factor, but large enough for an SEM model with two factors and three indicators per factor. Thompson (2000) demonstrated that the ratio of the sample size to the number of indicators (N/k) should be at least between 10:1 and 15:1 when determining the minimum sample size requirement based on the number of indicators. Therefore, we kept a minimum N/k to 5 or so to choose the minimum sample size for each theoretical model with different complexities. The maximum sample size was set to 5000, which is big enough for all theoretical models. Detailed parameter settings are shown in Table 2.

Simulation two

This simulation was to evaluate the performance of the CGFI, GFI, and AGFI taking into account sample size, magnitude of factor loadings, and types and degrees of model misspecification. The constructed theoretical model is shown in Figure 3, with four factors and each with five indicators. The design was a 2 ×2 ×11 ×7 (two types of model misspecification, two degrees of model misspecification, eleven levels of sample sizes, seven levels of factor loadings) factorial design.

Two common types of model misspecifications were considered in this simulation, measurement model and structural model misspecifications (Table 3). Although both underparameterized and overparameterized are known as incorrectly specified models, to be consistent with previous studies (Fan & Sivo, 2005; Fan et al., 1999; Hu & Bentler, 1998), our study focused on underparameterization.

We assigned two degrees of model misspecifications. For the measurement model, slight misspecification referred to the case when the factor loading of indicator X5 was set to zero, and severe misspecification was the case when the factor loadings of indicator X5, X7, and Y2 were set to zero (Figure 3).

For the structural model, slight misspecification referred to the case when the covariance of factor FA and FB and the regression coefficient of factor FA and FC were set to zero, and severe misspecification occurred when the covariance of factor FA and FB and all regression coefficients between factors were set to zero (Figure 3).

Table 4. Frequencies of improper solutions in simulation one.

	Estimation methods	
Sample size(N)	ML (%)	GLS (%)
30	9.70	15.40
40	1.75	3.85
50	0.30	1.78
60	0.06	1.56
80	0.00	0.18
100	0.00	0.00
120	0.00	0.00
150	0.00	0.00
200	0.00	0.00
300	0.00	0.00
500	0.00	0.00
1000	0.00	0.00
1500	0.00	0.00
2000	0.00	0.00
5000	0.00	0.00

ML: Maximum Likelihood; GLS: Generalized Least Square.

Table 5. η^2 of each study factor or interaction term for each fit index in simulation one.

	Fit index		
Study factors	GFI	AGFI	CGFI
Estimation Methods (EM)	<0.001	<0.001	<0.001
Model Complexity (MC)	0.140	0.056	0.202
Sample Size (SS)	0.750	0.841	0.211
EM*MC	<0.001	<0.001	0.001
EM*SS	<0.001	<0.001	0.002
MC*SS	0.110	0.043	0.149
EM*MC*SS	<0.001	<0.001	0.002

All the P-values of hypotheses test were<0.05.

We chose sample size of 150, 200, 300, 400, 500, 600, 800, 1000, 1500, 2000, and 5000, and used ML method for parameter estimation. In a measurement model, a factor loading of 0.40 or above is commonly used as a cutoff value for selecting an indicator. The corresponding indicator is regarded as reasonably stable when factor loading exceeds 0.80 (Velicer, Peacock, & Jackson, 1982). Therefore, factor loadings ranged from 0.30 to 0.90 with an increment of 0.10, and all indicators held the same loadings within each factor loading scenario. The correlation coefficients between factors equaled 0.50.

Simulation process

Once the model parameters are specified as above, the population covariance matrix (Σ) is obtained through Equation (4):

$$\Sigma = \begin{bmatrix} Cov_y & Cov_{yx} \\ Cov_{xy} & Cov_x \end{bmatrix} = \begin{bmatrix} \Lambda_y(I-B)^{-1}(\Gamma\phi\Gamma' + \Psi)(I-B')^{-1}\Lambda'_y + \theta_\varepsilon & \Lambda_y(I-B)^{-1}\Gamma\phi\Gamma\Lambda'_x \\ \Lambda_x\phi\Gamma'(I-B')^{-1}\Lambda'_y & \Lambda_x\phi\Lambda'_x + \theta_\delta \end{bmatrix} \tag{4}$$

where Λ_x (lambda-x) and Λ_y (lambda-y) are the parameter matrix of pattern coefficients for the x and y indicators, respectively; θ_δ (theta-delta) and θ_ε (theta-epsilon) are the covariance matrices of error terms of the x and y indicators, respectively; Γ is the parameter matrix for direct effects of exogenous factors on endogenous factors and B is the parameter matrix for direct effects of endogenous factors on each other; ϕ and Ψ are the covariance matrix for the disturbances of exogenous and endogenous factors, respectively; I is the identify matrix.

This matrix (Σ) is easily converted to a population correlation matrix (R) with Equation (5):

$$\Sigma = D_S * R * D_S \tag{5}$$

where R is the correlation matrix and D_S is a diagonal matrix containing the standard deviations of the variables (indicators) as its diagonal elements.

Then, we can use the matrix decomposition procedure by Kaiser and Dickman (1962) to generate multivariable distribution data with the population covariance matrix and resultant correlation matrix. Given a desired population correlation matrix R, the matrix decomposition procedure is as follows:

$$\hat{Z}_{(K\times N)} = F_{(K\times K)} \times \hat{X}_{(K\times N)}$$

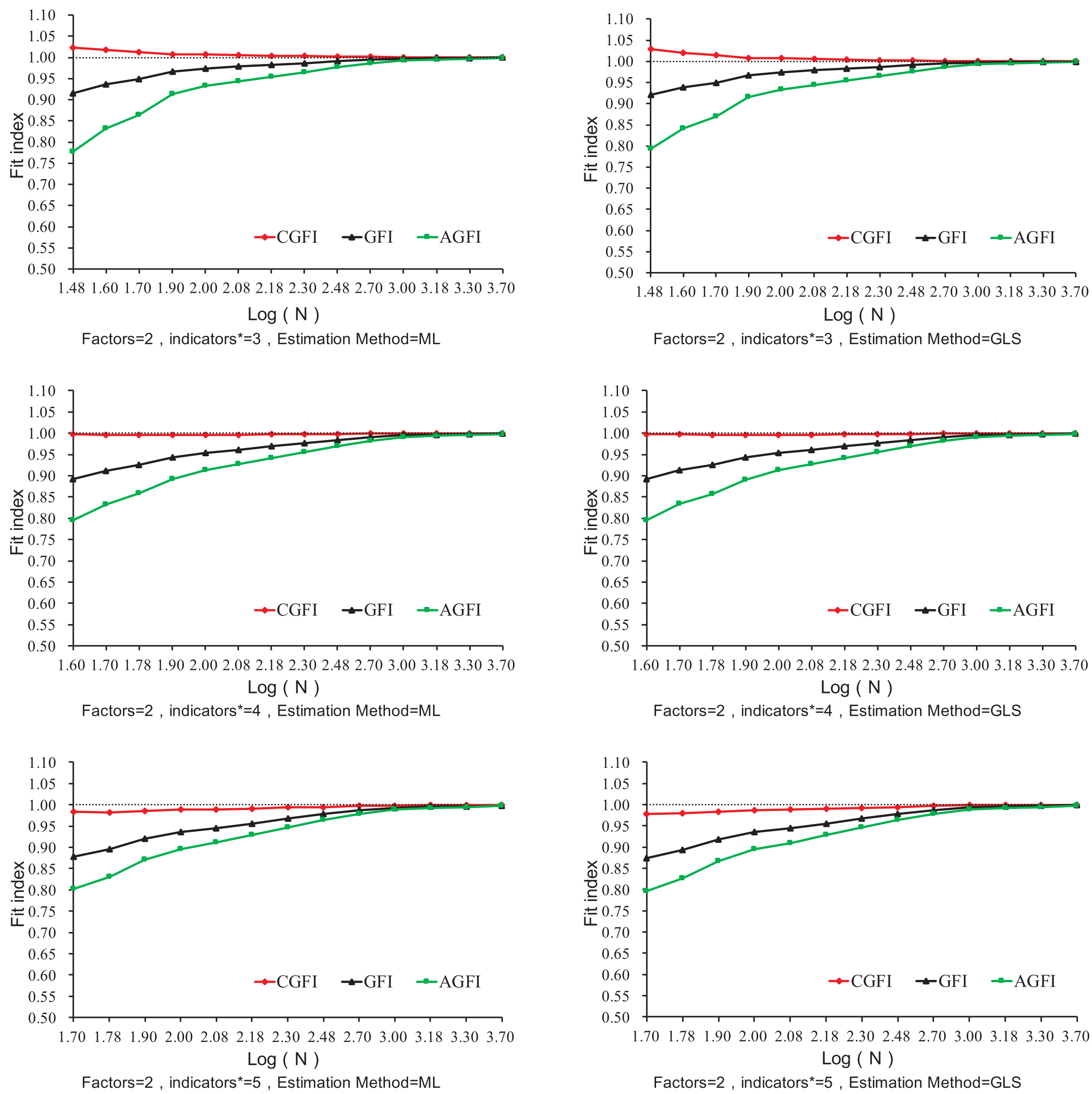


Figure 4. The trends of three fit indexes with the change of sample size under two factors, different numbers of indicators and estimation methods.

Where: $\hat{X}$ is a sample data matrix, with N observations and K uncorrelated variables each sampled from N (0, 1); F is a factor pattern matrix generated by applying principal component factorization to the given population correlation matrix R; and $\hat{Z}$ is the synthetic $K \times N$ sample data matrix, as if sampled from a population with the given population correlation matrix R (Fan & Fan, 2005).

Our simulations followed the steps below:

(1) Specify the theoretical model parameters and compute the $K \times K$ population correlation matrix R by using the SAS PROC IML;
(2) Conduct a factor analysis using principal component as the factor extraction method, completed by the SAS PROC FACTOR. Keep K factors and obtain the $K \times K$ matrix of factor pattern F;
(3) Generate K uncorrelated random normal variables, each with N observations from $N(0,1)$. Use the pseudorandom number generator with a fixed random seed number. Then transpose the above $N \times K$ sample data matrix to $K \times N$ matrix $\hat{X}$;
(4) Multiply the factor pattern matrix F with sample data matrix $\hat{X}$ and obtain the $K \times N$ sample data matrix $\hat{Z}$;
(5) Fit the test model with sample data matrix $\hat{Z}$ under different simulation scenarios, using the SAS PROC CALIS. All results were saved for later analysis.

The simulations were performed in SAS (v. 9.4; SAS Institute Inc, Cary, NC). Simulation programming was implemented through a combination of the SAS MACRO, the SAS PROC IML, and the SAS PROC CALIS for SEM model fitting (Fan et al., 1999). A two-tailed 0.05 significance level is used.

Assessment of fit indexes

Nonconvergence and improper solutions can occur in simulation studies of SEM. Nonconvergence means that the SEM estimation fails to converge on a solution for a sample after reaching the maximum number of iterations permitted,

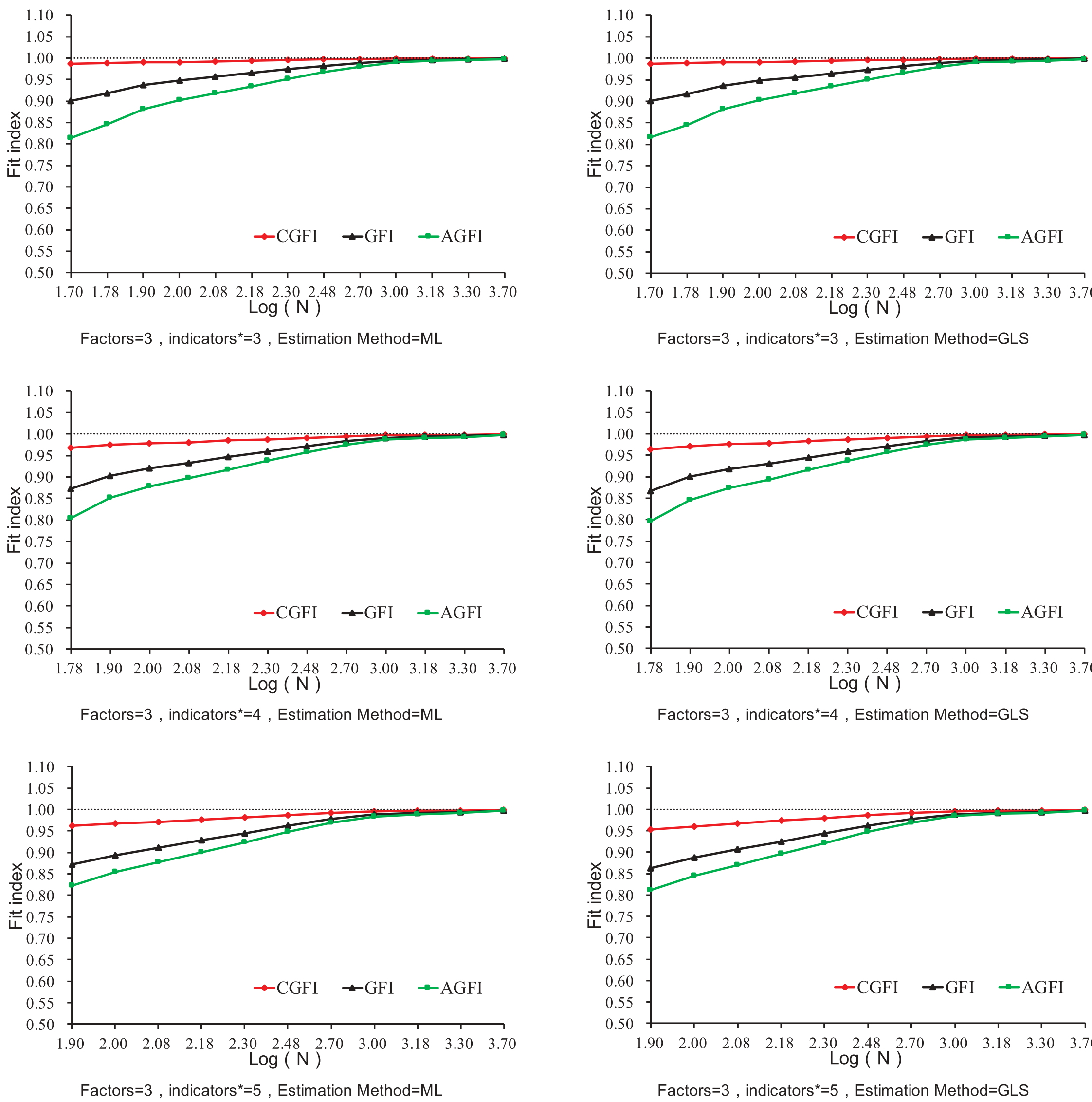


Figure 5. The trends of three fit indexes with the change of sample size under three factors, different numbers of indicators and estimation methods.

whereas the problem of an improper solution happens when some statistically impossible values, such as negative variances (also called "Heywood cases", Kolenikov & Bollen, 2012; Rindskof, 1984), are obtained from the estimation. In this study, the Levenberg-Marquardt optimization procedure was used (v. 9.4; SAS Institute Inc, Cary, NC), and the maximum number of iterations was set at 1,000,000. The nonconvergence issue did not appear in our study. We calculated the frequency of improper solutions for each simulation. Subsequent analyses were based on the entire data set. A series of analyses of variance (ANOVA) were performed for each fit index obtained. The η^2, indicating the proportion of variance in each fit index accounted for by each factor or interaction term, is equivalent to R^2 and is used to evaluate the degree of the effect for each factor or interaction term on each fit index (Hays, 1988). The η^2 is calculated by dividing the type 3 sum of squares for a given factor or interaction term by the corrected total sum of squares, and $\eta^2 \leq 0.01$ is considered a negligible effect even though the *P*-value may be significant (Hu & Bentler, 1998). Besides, how the CGFI, GFI, and AGFI change with various sample sizes or factor loadings was depicted to assess the properties of these indexes.

Results

Simulation one

Improper solutions

Collapsing model complexity, we presented the frequencies of improper solutions under different sample sizes and estimation methods in Table 4. Apparently, the ML estimation method had fewer improper solutions than GLS estimation method, and the improper solutions occurred when the sample size is relatively small (<100).

The proportion of variance (η^2)

We performed a nested analysis of variance. The factors considered were the estimation method, model complexity, and sample size as shown in Table 5. The CGFI, GFI, and

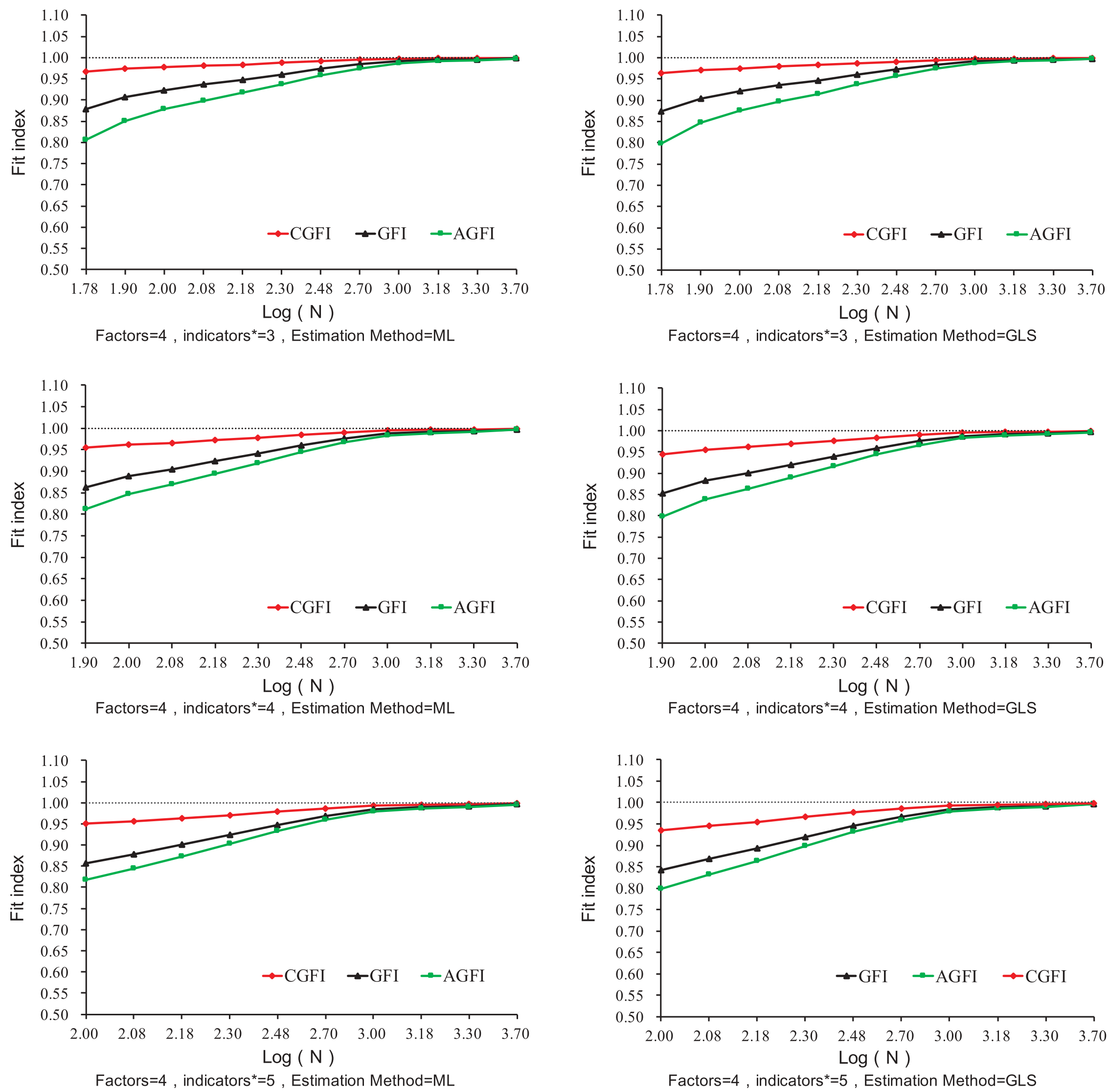


Figure 6. The trends of three fit indexes with the change of sample size under four factors, different numbers of indicators and estimation methods.

Table 6. Frequencies of improper solutions in simulation two.

	Model ID				
Factor loadings	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
0.30	10.08	12.00	15.10	10.33	8.33
0.40	1.06	1.50	2.49	0.98	0.71
0.50	0.03	0.03	0.18	0.04	0.03
0.60	0.00	0.00	0.00	0.00	0.00
0.70	0.00	0.00	0.00	0.00	0.00
0.80	0.00	0.00	0.00	0.00	0.00
0.90	0.00	0.00	0.00	0.00	0.00

Table 7. η^2 of each study factor or interaction item on each fit index in simulation two.

	Fit index		
Study factors	GFI	AGFI	CGFI
	Measurement model misspecification		
Factor Loadings (FL)	0.178	0.179	0.245
Sample Size (SS)	0.279	0.285	0.025
Model Misspecification (MM)	0.365	0.359	0.484
FL*SS	0.001	0.001	0.002
FL*MM	0.162	0.162	0.223
SS*MM	0.003	0.004	0.007
FL*SS*MM	0.001	0.001	0.002
	Structural model misspecification		
Factor Loadings (FL)	0.064	0.064	0.108
Sample Size (SS)	0.406	0.420	0.043
Model Misspecification (MM)	0.436	0.421	0.684
FL*SS	0.001	0.001	0.001
FL*MM	0.071	0.071	0.120
SS*MM	0.004	0.005	0.012
FL*SS*MM	0.001	0.001	0.001

All the *P*-values of hypotheses test were<.05.

AGF were not sensitive to estimation method (η^2< 0.001) and were slightly sensitive to model complexity (η^2s ranged from 0.056 to 0.202). The sample size accounted for a very large proportion of variance (η^2s = 0.750 and 0.841) in the GFI and AGFI. However, the proportion was much less in the CGFI (η^2 = 0.211), suggesting that the CGFI is much less sensitive to sample size. Besides, relatively small interaction effect existed between model complexity and sample size (η^2s ≤0.149), indicating that the influence of model complexity on these indexes is not uniform for different sample sizes.

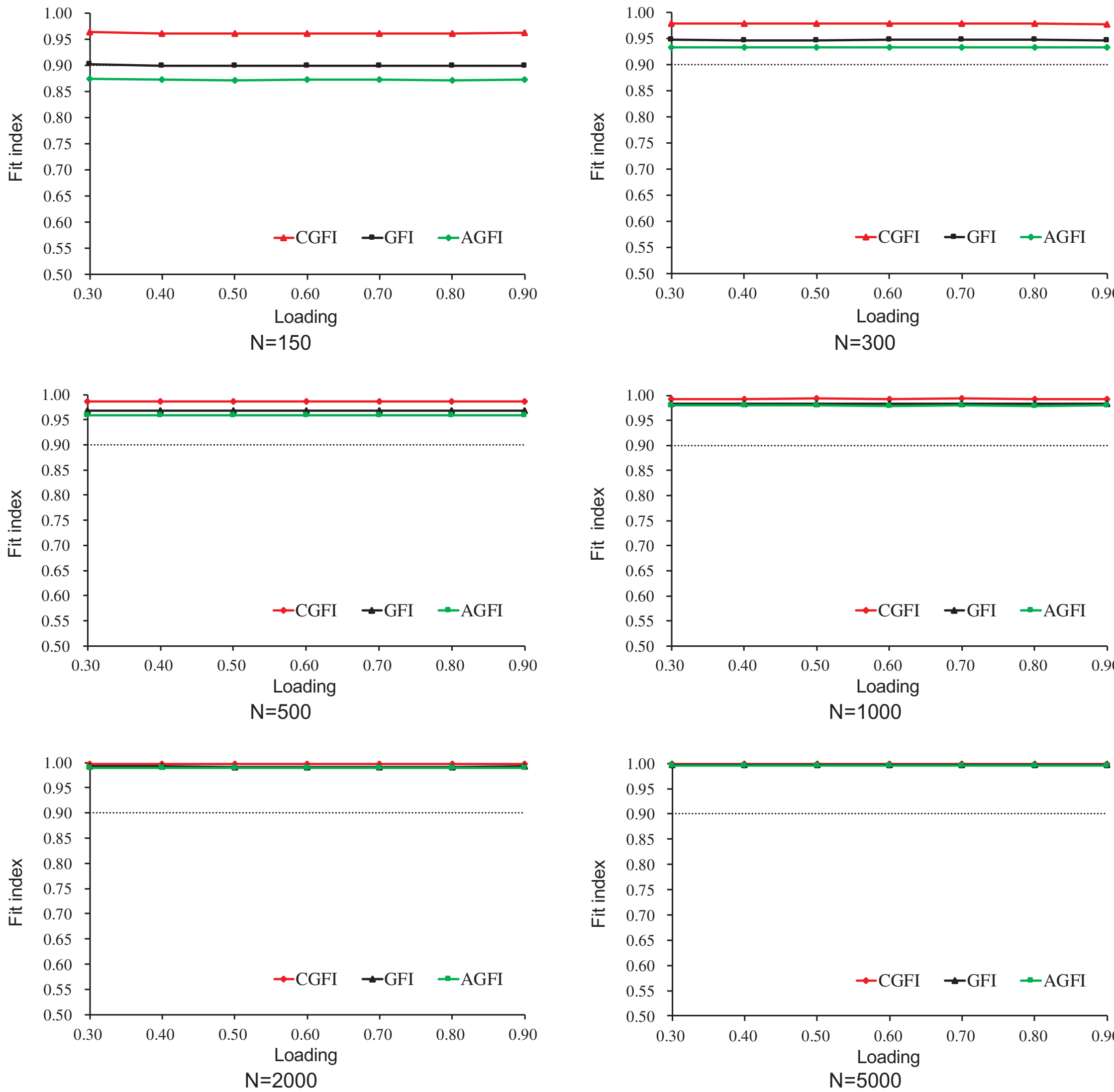


Figure 7. The trends of three fit indexes with the change of factor loadings under the true model with different sample sizes.

The Trends with the Change of Sample Size. The GFI and AGFI were affected by sample size (Figures 4–6). Especially, when the sample size is less than 150 there was a moderately large downward bias. In addition to a small bias when the sample size is less than 150, the CGFI was insensitive to sample size. As model complexity increasing, the values of the CGFI, GFI, and AGFI were slightly reduced, which indicated that these indexes have some degree of penalty effect on model complexity. Among them, the CGFI presented a stronger penalty effect than other indexes. This penalty effect diminished with the increase of sample size. Two estimation methods performed similarly on all the three indexes.

Simulation two

Improper solutions

Collapsing sample sizes, we illustrate the frequencies of improper solutions under different factor loadings and model specifications in Table 6. The improper solutions mainly occurred in relatively lower factor loadings (≤0.50), and measurement model misspecification (Model ID equals 2 and 3) might result in a higher proportion of improper solutions than structural model misspecifications (Model ID equals 4 and 5).

The proportion of variance (η^2)

The factorial analysis of variance was performed among factor loading, sample size, types and degrees of model misspecification as shown in Table 7.

For measurement model misspecifications, the CGFI, GFI, and AGFI showed slight sensitivity to factor loadings with η^2 ranging from 0.178 to 0.245. The GFI and AGFI were sensitive to sample size (η^2s = 0.279 and 0.285), while the CGFI was not (η^2=0.025). Besides, the CGFI appeared much more sensitive to model misspecification than the GFI and AGFI, with η^2 equaling to 0.484, 0.365 and 0.359, respectively. There was a relatively small interaction effect among factor loadings and model misspecification (η^2s ≤0.223) for these indexes.

For structural model misspecifications, the CGFI, GFI, and AGFI were not sensitive to factor loadings with η^2s less than 0.108. The GFI and AGFI were moderately sensitive to sample

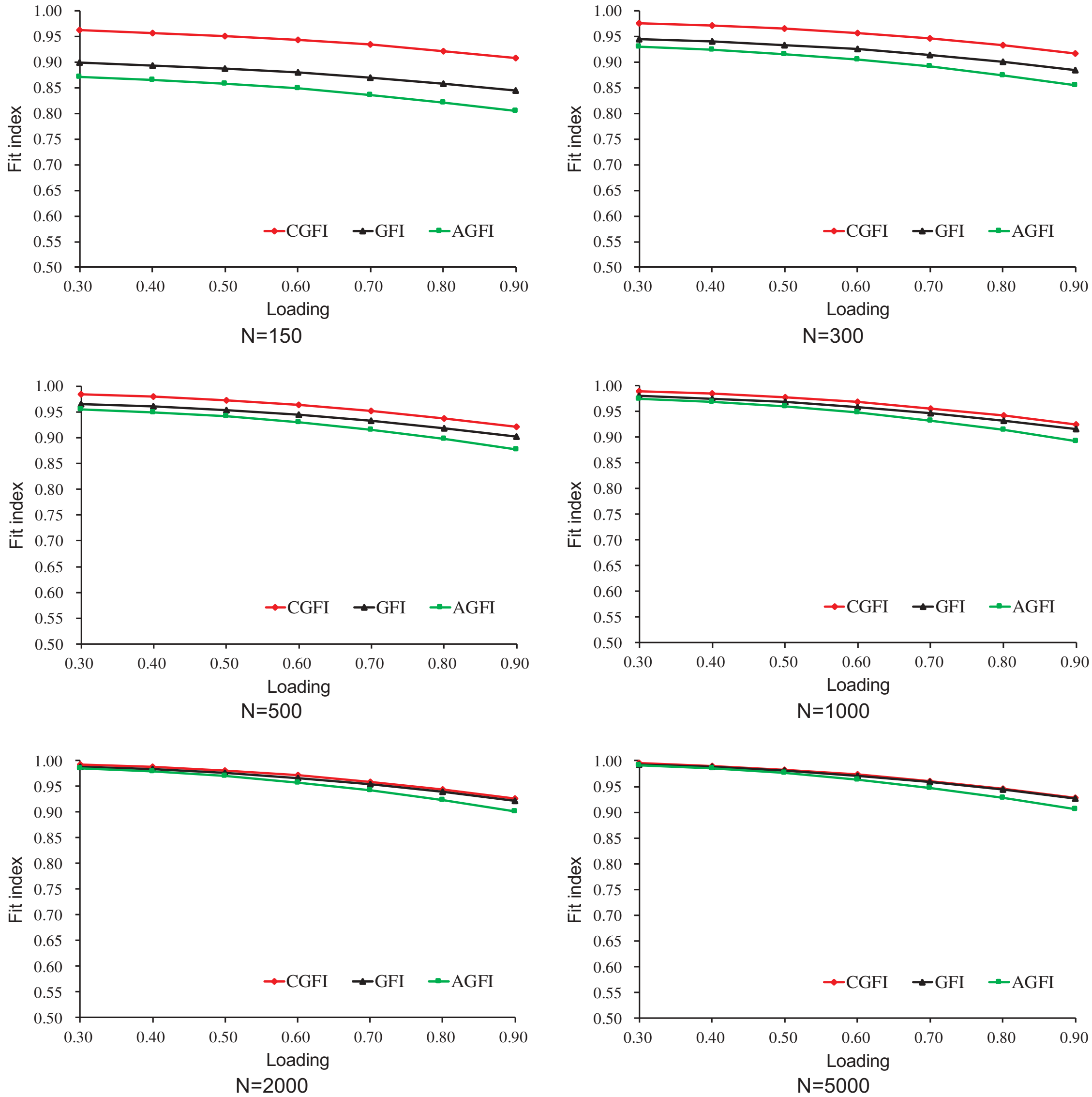


Figure 8. The trends of three fit indexes with the change of factor loadings under the slight misspecification of measurement model with different sample sizes.

size with η^2 equaling 0.406 and 0.420, respectively. As expected, the CGFI was insensitive to sample size (η^2 = 0.043) and very sensitive to model misspecification (η^2 = 0.684) than the other two indexes.

The trends with the change of factor loadings

The Trends with the Change of Factor Loadings under the True Model. The CGFI, GFI, and AGFI were not sensitive to factor loadings overall (Figure 7). When the sample size was relatively small (≤300), the CGFI had slightly downward bias while the AGFI and GFI revealed a much more obviously downward bias. This downward bias disappeared quickly as the sample size increases.

The Trends with the Change of Factor Loadings under Measurement Model Misspecifications. Theoretically, the larger the factor loadings are, the bigger the effect on fit indexes by the misspecification of factor loadings and corresponding to the lower index values, when fixing the model misspecification degree. The results were consistent with the above assumption, which indicated good sensitivity of these fit indexes to misspecification (Figures 8–9). The CGFI was not sensitive to sample size, while the GFI and AGFI were much more sensitive to small sample size (≤500), especially for the AGFI.

The Trends with the Change of Factor Loadings under Structural Model Misspecifications. Figures 10–11 show the CGFI, GFI, and AGFI are less sensitive to factor loadings under the conditions of structural model misspecifications compared to the conditions of measurement model misspecifications. Three fit indexes were sensitive to factor loadings and presented much lower values with increasing misspecification degrees. The CGFI was insensitive to sample size, while the GFI and AGFI were moderately sensitive to small sample size (≤500).

Real examples

We applied the CGFI, GFI, and AGFI to some real examples with different sample sizes and model fitting degrees, which

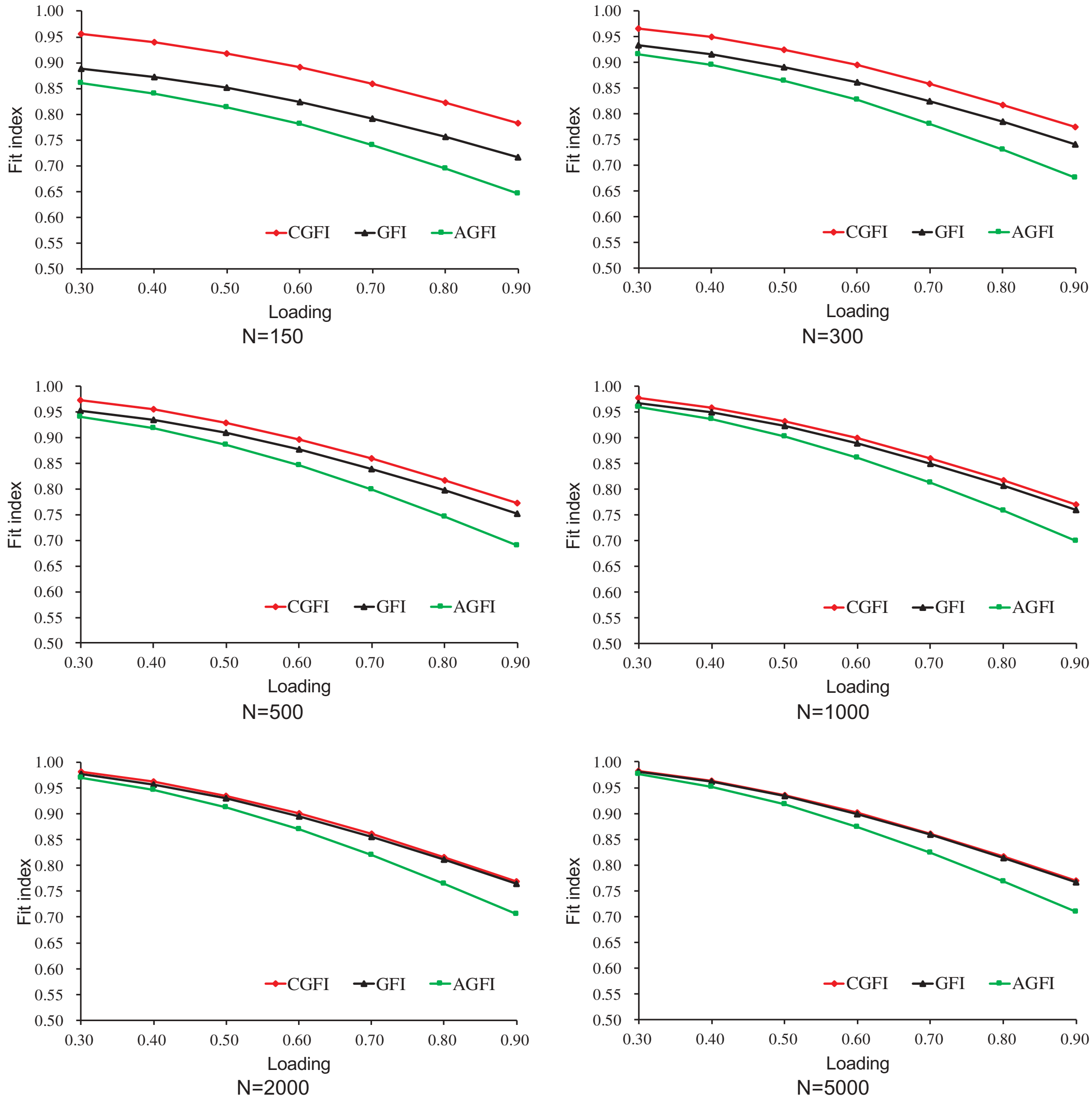


Figure 9. The trends of three fit indexes with the change of factor loadings under the severe misspecification of measurement model with different sample sizes.

covered medical, sociology, and economics, etc., to compare the properties of the three fit indexes in practice. The real examples and related model information are depicted in Table 8.

First three examples had a small sample size ($N \le 148$), and it was obvious that the values of the GFI and AGFI were lower than the value of the CGFI, especially the AGFI (Goubert, Crombez, & Van Damme, 2004; Wu, Wang, & Lin, 2007; Yoon & Uysal, 2005). While the fitting results reflected by the CGFI were consistent with the overall evaluation results of the models, which indicated the models fitted well. With the increase of sample size, Kim (2008) and Gibbs, Giever, and Higgins (2003)'s results revealed that the value of the AGFI was lower than the CGFI's, while the gap between the GFI and CGFI's was very small. In the example of Cook, Brawer, and Vowles (2006), the CGFI could also identify a bad model fit just like the GFI, etc. For the last three examples with large enough sample sizes, the three fit indexes had almost the same value and behaved in a manner consistent with the overall evaluation, which confirmed the findings of the above simulation studies (Kelly, 2011; Ko & Stewart, 2002; Papi, 2010).

Discussion

We formulated a corrected fit index–CGFI based on the GFI. The main idea was to correct the downward bias caused by small sample size by rescaling it with $\frac{1}{N}$ and penalized the complexity of model with a scale of $\frac{2}{1-\frac{2df_T}{k(k+1)}}\frac{1}{N}$. We then performed comprehensive Monte Carlo simulations to compare the properties of the CGFI, GFI, and AGFI, across a wide range of parameter settings including sample size, estimation method, factor loadings, model complexity, and types and degrees of model misspecification. The results revealed that the proposed CGFI was not sensitive to sample size, and much more sensitive to model misspecifications than the GFI and AGFI, and thus more suitable for assessing model fitting. For ease of use, the critical value of the CGFI was

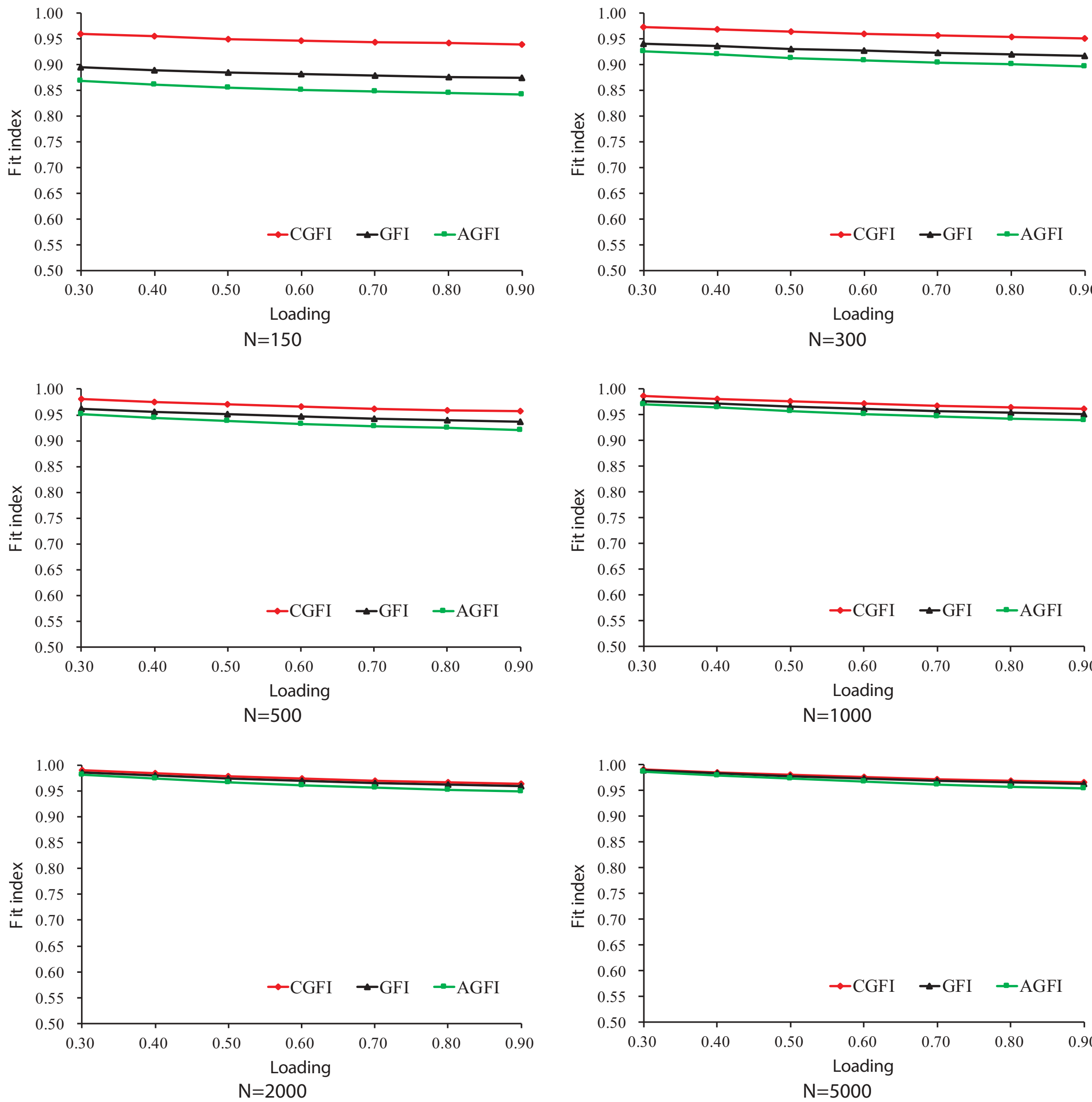


Figure 10. The trends of three fit indexes with the change of factor loadings under the slight misspecification of structural model with different sample sizes.

maintained at a fixed level (0.90) empirically, as was consistent with other fit indexes including the GFI, CFI, NNFI and so on (Fan et al., 1999; Hu & Bentler, 1998; Shevlin & Miles, 1998). However, more elaborate choices of the cutoff values of the CGFI under different conditions similar to Yuan, Yang, and Jiang (2017) may potentially be derived.

The impact of model complexity, a known factor that affects the extrapolation of study findings (Kenny & McCoach, 2003; Shevlin & Miles, 1998), has been examined. In applied research, the number of the factor for most SEM falls within the range of 2–5. Besides, theoretical studies have pointed out that when the SEM is constructed with 2–6 factors and each with 2–6 indicators, the research results obtained under such design have better extrapolation (Anderson & Gerbing, 1984; Ding et al., 1995; Kenny & McCoach, 2003). In addition, Anderson and Gerbing (1984) found that the lower bound of the number of indicators per factor is 2. Thus, to evaluate the effect of model complexity on fit indexes properly, we included 2–4 factors and 3–5 indicators per factor in Simulation 1. And in Simulation 2, we constructed the theoretical model with four factors and each with five indicators, which is more relevant to the findings from theory and application (Ding et al., 1995), making our findings more general.

Although a true model is relatively easy to specify in a simulation study, model misspecification is difficult to set for at least two reasons: (a) model misspecification can take a myriad of forms, and (b) the degree of model misspecification is not easily quantified. Some scholars proposed that the degree of misspecification could be accurately quantified by using the power of a specified model. However, this method highly depends on sample size, which varies among different simulation conditions and could not be effectively handled in practice, and it is not scientific to use it as a quantitative indicator (Fan & Sivo, 2005; Saris et al., 1987). Therefore, in Simulation 2, we specified two degrees of model misspecification, slight and severe misspecifications (Fan & Sivo, 2005, 2007; Hu & Bentler, 1998; Kenny & McCoach, 2003), and we considered both types of model misspecifications in measurement and structural model by adjusting the misspecification

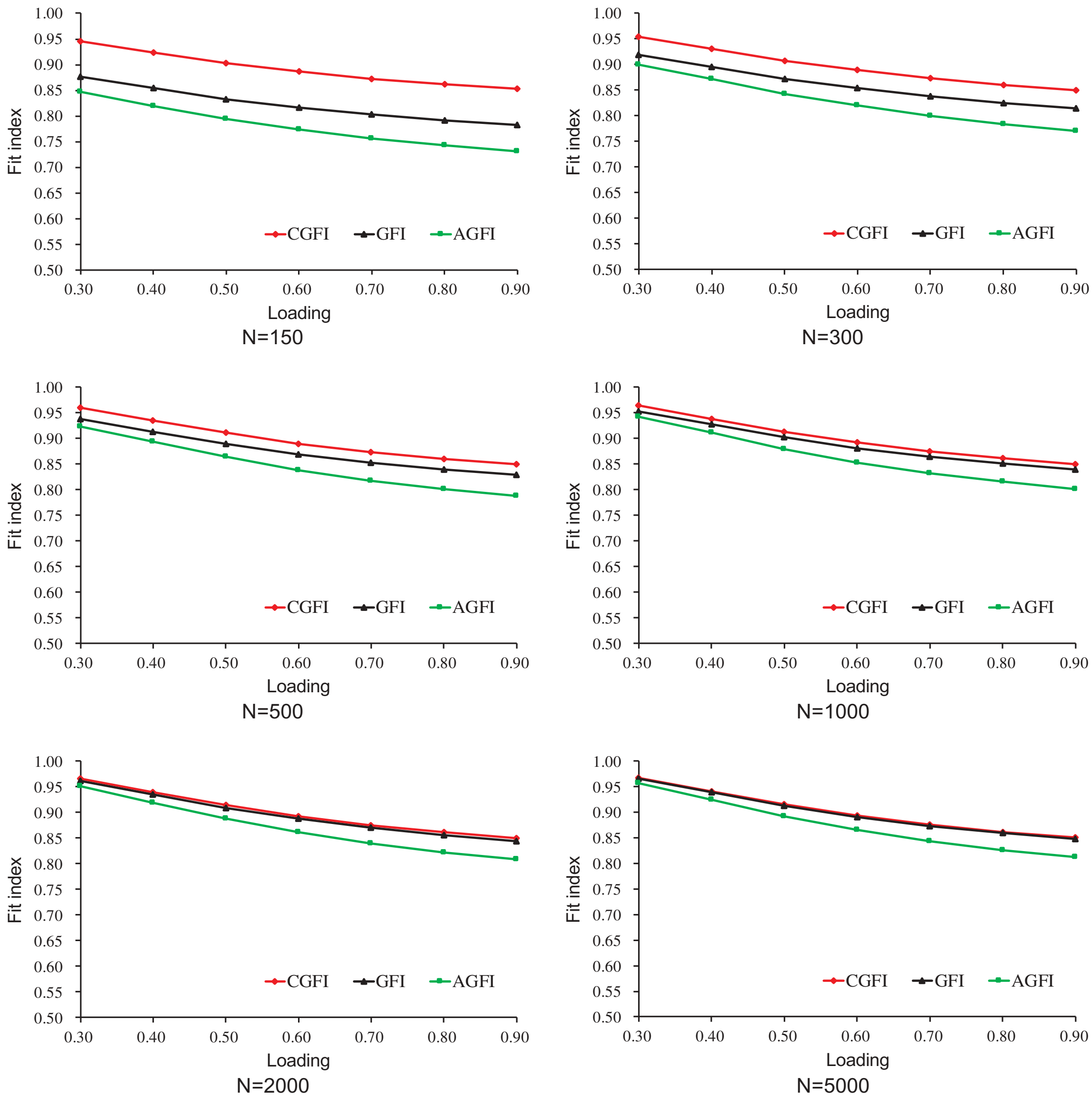


Figure 11. The trends of three fit indexes with the change of factor loadings under the severe misspecification of structural model with different sample sizes.

Table 8. The overall model fit the results of SEM model.

Study ID	Sample size(N)	Indicator(k)	df_T	Fit index			Overall evaluation
				AGFI	GFI	CGFI	
Recommend value	N/A	N/A	N/A	0.90	0.90	0.90	well/bad
Goubert et al. (2004)	122	11	40	0.85	0.91	0.95	well
Wu et al. (2007)	123	18	122	0.80	0.86	0.92	well
Yoon and Uysal (2005)	148	11	37	0.91	0.95	0.98	well
Kim (2008)	411	21	180	0.87	0.90	0.92	well
Gibbs et al. (2003)	422	9	24	0.94	0.97	0.98	well
Cook et al. (2006)	469	17	119	0.76	0.82	0.84	bad
Ko and Stewart (2002)	732	17	110	0.90	0.93	0.94	well
Papi (2010)	1011	21	170	0.95	0.96	0.97	well
Kelly (2011)	1025	7	8	0.99	1.00	1.00	well

conditions to enhance the reliability of research findings (Fan & Sivo, 2007; Hu & Bentler, 1998). The terms slight and moderate misspecification were used here exclusively to indicate different degrees of misspecification and should not be generalized beyond the present study.

In our study, we generated data under a standard multivariate normal distribution and did not consider other data distributions by adopting a common assumption that data distribution does not seriously affect the performance of fit indexes (Fan & Wang, 1998; Hu & Bentler, 1998; Lei & Lomax, 2005). Improper solutions occurred more frequently for small sample sizes and for the GLS estimation method, which is consistent with the findings of Anderson and Gerbing (1984, 1987) and Boomsma (1985). The causes of an improper solution are controversial and not well understood. Restricting the range of estimates to be [0, +∞) or not is two commonly accepted methods for this problem, and we have chosen the latter one in our study, which is the same choice as Rindskof (1984).

We have shown that the GFI and AGFI are mainly affected by sample size and have a certain degree of downward bias, especially when sample sizes are less than 150, which is consistent with the conclusions of Gerbing and Anderson (1992) and Marsh et al. (1988). Both the GFI and AGFI mainly utilize the information of the covariance matrix, which is not stable in small samples and gradually becomes stable with the increase of sample size. This may explain the dependence of the GFI and AGFI on sample sizes. The results revealed that both the GFI and AGFI were not influenced by the estimation method in the correctly specified models, which is accordant with the findings of Wang, Fan, and Willson (1996). Sugawara and Maccallum (1993) also pointed out that the GFI and AGFI tend to behave relative consistently across different estimation methods, especially for well-fitting models. Although different estimation methods yield different values in discrepancy function, this has little effect on the values of the two indexes. However, as the model fit gotten worse, this discrepancy becomes more apparent and leads to distinct values for both the GFI and AGFI.

After correcting for the influence of sample size and penalizing model complexity, the CGFI is not sensitive to sample size. The CGFI is more sensitive to detect model misspecifications than the GFI and AGFI in both measurement and structural models. Moreover, the three fit indexes are not affected by the estimation method and are relatively sensitive to factor loadings and complexity of the model.

Although we concluded that the CGFI behaves much better than the GFI and AGFI, we did not compare the performance of the CGFI with other fit indexes such as RMSEA, CFI, etc. Besides, we did not give theoretical derivations in this article and the problem may be mathematically intractable. Our approach has been simulation-based and empirical. We did not investigate the estimation method on the behaviors of the CGFI under misspecified models neither. In the end, while the SEM is complex and diverse, further work to explore the limitations of the CGFI under different conditions may be needed, for example, across different types of SEM. And further theoretical studies to fully understand the dynamic nature of the multifactorial modeling may be needed. It also remains an unsolved statistical problem on how small the sample size in complex models such as SEM is.

Conclusion

Our proposed fit index, CGFI, for model evaluation in SEM, accounts for model complexity and adjusts the downward bias when the sample size is smaller (say, less than 150) since the existing GFI is artificially downshifted with smaller sample sizes. It can serve as a useful supplementary goodness-of-fit measure to existing ones.

References

Anderson, J. C., & Gerbing, D. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173. doi:10.1007/BF02294170

Bollen, K. A., Stine, R. A., Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-Fit Measures in structural equation models. *Sociological Methods & Research*, *21*, 205–229. doi:10.1177/0049124192021002004

Bone, P. F., Sharma, S., & Shimp, T. A. (1989). A bootstrap procedure for evaluating goodness-of-fit indices of structural equation and confirmatory factor models. *Journal of Marketing Research*, *26*, 105–111. doi:10.1177/002224378902600109

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation. *Psychometrika*, *50*, 229–242. doi:10.1007/BF02294248

Cook, A. J., Brawer, P. A., & Vowles, K. E. (2006). The fear-avoidance model of chronic pain: Validation and age analysis using structural equation modeling. *Pain*, *121*, 195–206. doi:10.1016/j.pain.2005.11.018

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, *2*, 119–143.

Fan, X., & Fan, X. (2005). TEACHER'S CORNER: Using SAS for Monte Carlo SIMULATION RESEARch in SEM. *Structural Equation Modeling*, *12*, 299–333. doi:10.1207/s15328007sem1202_7

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*, 343–367. doi:10.1207/s15328007sem1203_1

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*, 509–529.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56–83. doi:10.1080/10705519909540119

Fan, X., & Wang, L. (1998). Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational & Psychological Measurement*, *58*, 701–735. doi:10.1177/0013164498058005001

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika*, *52*, 99–111. doi:10.1007/BF02293958

Gerbing, D. W., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods & Research*, *21*, 132–160. doi:10.1177/0049124192021002002

Gibbs, J. J., Giever, D., & Higgins, G. E. (2003). A test of Gottfredson and Hirschi's general theory using structural equation modeling. *Criminal Justice & Behavior*, *30*, 441–458. doi:10.1177/0093854803253135

Goubert, L., Crombez, G., & Van Damme, S. (2004). The role of neuroticism, pain catastrophizing and pain-related fear in vigilance to pain: A structural equations approach. *Pain*, *107*, 234–241. doi:10.1016/j.pain.2003.11.005

Hays, W. L. (1988). *Statistics*. New York, NY: Holt, Rinehart & Winston.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal on Business Research Methods*, *6*, 141–146.

Hu, L., & Bentler, P. (1998). Fit indices in covariance structure modelling: Sensitivityto underparameterization model misspecification. *Psychological Methods*, *3*, 424–453. doi:10.1037/1082-989X.3.4.424

Joreskog, K. G., & Sorbom, D. (1981). LISREL V: Analysis of linear structural relationships by the method of maximum likelihood. *National Educational Resources*.

Joreskog, K. G., & Sorbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*, 404–416. doi:10.1177/002224378201900402

Joreskog, K. G., & Sorbom, D. (1984). *LISREL VI user's guide* (3rd ed.). Mooresville, IN: Scientific Software.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation ma- trices from an arbitrary population correlation matrix. *Psychometrika*, *27*, 179–182. doi:10.1007/BF02289635

Kelly, S. (2011). Do homes that are more energy efficient consume less energy?: A structural equation model of the English residential sector. *Energy*, *36*, 5610–5620. doi:10.1016/j.energy.2011.07.009

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, *10*, 333–351. doi:10.1207/S15328007SEM1003_1

Kim, K. (2008). Analysis of structural equation model for the student pleasure travel market: Motivation, involvement, satisfaction, and destination loyalty. *Journal of Travel & Tourism Marketing*, *24*, 297–313. doi:10.1080/10548400802156802

Ko, D. W., & Stewart, W. P. (2002). A structural equation model of residents' attitudes for tourism development. *Tourism Management*, *23*, 521–530. doi:10.1016/S0261-5177(02)00006-7

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification?. *Sociological Methods & Research*, *41*, 124–167. doi:10.1177/0049124112442138

La Du, T. J., & Tanaka, J. S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology*, *74*, 625–635. doi:10.1037/0021-9010.74.4.625

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling*, *12*, 1–27. doi:10.1207/s15328007sem1201_1

Marsh, H. W., & Balla, J. (1994). Goodness of fit in confirmatory factor analysis: The effects of sample size and model parsimony. *Quality & Quantity*, *28*, 185–217. doi:10.1007/BF01102761

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-Fit Indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410. doi:10.1037/0033-2909.103.3.391

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*, 430–445. doi:10.1037/0033-2909.105.3.430

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, *38*, 467–479. doi:10.1016/j.system.2010.06.011

Rindskof, D. (1984). Structural equation models: Empirical identification, heywood cases, and related problems. *Sociological Methods & Research*, *13*, 109–119. doi:10.1177/0049124184013001004

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, *17*, 105–129. doi:10.2307/271030

Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy*, *4*, 83–97. doi:10.1016/j.sapharm.2007.04.003

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*, 323–338. doi:10.3200/JOER.99.6.323-338

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*, 85–90.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*, 267–288. doi:10.3200/JEXE.74.3.267-288

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the Annual meeting of the Psychometric Society.

Sugawara, H. M., & Maccallum, R. C. (1993). Effect of estimation method on incremental fit indexes for covariance structure models. *Applied Psychological Measurement*, *17*, 365–377. doi:10.1177/014662169301700405

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure under arbitrary GLS estimation. *Journal of British Journal of Mathematical Statistical Psychology*, *38*, 197–201.

Thompson, B. (2000). *Ten commandments of structural equation modeling*.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research*, *17*, 371–388. doi:10.1207/s15327906mbr1703_5

Wang, L., Fan, X., & Willson, V. L. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling*, *3*, 228–247. doi:10.1080/10705519609540042

Wu, J. H., Wang, S. C., & Lin, L. M. (2007). Mobile computing acceptance factors in the healthcare industry: A structural equation model. *International Journal of Medical Informatics*, *76*, 66–77. doi:10.1016/j.ijmedinf.2006.06.006

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. *Tourism Management*, *26*, 45–56.

Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2015). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling*, *23*, 1–12.

Yuan, K. H., Yang, M., & Jiang, G. (2017). Empirically corrected rescaled statistics for sem with small n and large p. *Multivariate Behavioral Research*, *52*, 1–26. doi:10.1080/00273171.2017.1354759