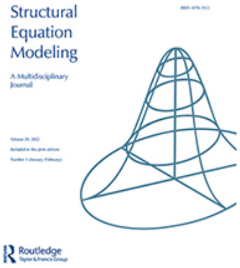

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Evaluation and Comparison of SEM, ESEM, and BSEM in Estimating Structural Models with Potentially Unknown Cross-loadings

Xiayan Wei, Jiasheng Huang, Lijin Zhang, Deng Pan & Junhao Pan

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2021.2006664

Routledge
Taylor & Francis Group

Check for updates

Evaluation and Comparison of SEM, ESEM, and BSEM in Estimating Structural Models with Potentially Unknown Cross-loadings

Xiayan Wei[a,b*], Jiasheng Huang[a*], Lijin Zhang[a], Deng Pan[c], and Junhao Pan[a]

[a]Sun Yat-sen University; [b]Zhejiang University; [c]Huazhong University of Science and Technology

ABSTRACT

Cross-loadings are common in multidimensional instruments; however, they cannot be appropriately addressed in conventional structural equation modeling (SEM) owing to the assumption of zero cross-loadings in standard confirmatory factor analysis (CFA). Although it has been proposed that exploratory structural equation modeling (ESEM) and Bayesian structural equation modeling (BSEM) can address this issue more flexibly, their performance in structural parameter estimation has not been adequately compared. This study uses simulated data to evaluate and compare SEM, ESEM, and BSEM in estimating structural models under different manipulation conditions (i.e., sample size, target loading, cross-loading, and path coefficient). The results demonstrated that the performances of these approaches were similar in the case of zero cross-loadings. SEM performed worse as cross-loadings increased, and the performance of BSEM significantly depended on the accuracy of the priors for cross-loadings. ESEM was inferior to BSEM with correctly specified prior means for cross-loadings in most evaluation measures and exhibits unstable performance in conditions with small target loadings. Recommended strategies for selecting an appropriate modeling approach are discussed based on our findings.

KEYWORDS

BSEM; cross-loadings; ESEM; structural parameter estimation

Confirmatory factor analysis (CFA) is widely adopted in the structural equation modeling (SEM) framework to evaluate the measurement model for the factor structures of latent factors. Usually, this approach, also known as the standard CFA by Kline (2015), assumes zero cross-loadings (i.e., loadings on non-targeted factors are constrained to be zero) to reflect a more parsimonious and interpretable factor structure in which each indicator is only influenced by a specific targeted factor (Asparouhov & Muthén, 2009). Such a demanding assumption has been criticized as unrealistic because it is practically impossible to perfectly design pure indicators of one construct (or factor), and most existing multi-factor instruments fail to meet these criteria (Marsh et al., 2009; Murray et al., 2019). The inappropriate constraints of zero cross-loadings on all non-targeted factors usually result in poor model fit in the applications of SEM (Marsh et al., 2014). Moreover, cross-loadings sometimes have substantive meaning. For example, psychological symptoms can reflect multiple diagnostic factors rather than a solely single factor (Murray et al., 2019). In practice, this overly restrictive assumption can lead to problematic compensating strategies, including the misuse of modification indexes (Asparouhov & Muthén, 2009; MacCallum et al., 1992). The misspecification of zero loadings may also generate inflated factor correlations, thereby distorting structural relationships or triggering biased estimates in full SEM with other outcome variables (Asparouhov & Muthén, 2009; Marsh et al., 2014, 2009).

To address the limitations of the constraint of zero cross-loadings in standard CFA, nonstandard CFA models, in which some indicators load on more than a single factor or some error terms covary, have been proposed (Kline, 2015). However, Xiao et al. (2019) suggested that the nonstandard CFA remains unsuitable for data with many cross-loadings and impractical for the unknown cross-loading or complex structure cases. Among the novel approaches that have been proposed to extend the overly restrictive CFA measurement model of the conventional SEM, exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) and Bayesian structural equation modeling (BSEM; B. Muthén & Asparouhov, 2012) are two of the most widely adopted methods.

Conventional structural equation modeling (SEM)

To explain ESEM and BSEM, we start with the conventional SEM with continuous indicators. For the i-th participant, let $\boldsymbol{y}_i$ be a random vector of p indicator variables that satisfies the following measurement model:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, i = 1, 2, \ldots, N \quad (1)$$

where $\boldsymbol{\mu}$ represents the $p \times 1$ vector of intercept, and $\boldsymbol{\Lambda}$ represents the $p \times q$ loading matrix that reflects the relationship between observed indicators in $\boldsymbol{y}_i$ and the $q \times 1$ vector of latent variables in $\boldsymbol{\omega}_i$. The $p \times 1$ vector of random errors in $\boldsymbol{\varepsilon}_i$,

CONTACT Junhao Pan panjunh@mail.sysu.edu.cn Department of Psychology, Sun Yat-sen University, Guangzhou, P.R. China; Deng Pan pand@hust.edu.cn School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, P. R. China.
*These authors contributed equally to this work.

which follows $N[\mathbf{0}, \boldsymbol{\Psi}_\epsilon]$ and $\boldsymbol{\Psi}_\epsilon$ is typically a diagonal matrix. In conventional SEMs, the structure of $\boldsymbol{\Lambda}$ is specified a priori according to substantive theory such that the measurement model is a standard CFA. To assess the interrelationships among latent variables, $\boldsymbol{\omega}_i$ is commonly partitioned into two parts $(\boldsymbol{\eta}_i^T,\ \boldsymbol{\xi}_i^T)^T$, where $\boldsymbol{\eta}_i(q_1 \times 1)$,and $\boldsymbol{\xi}_i(q_2 \times 1)$ respectively represent the vectors of endogenous and exogenous latent variables. The interrelationships between $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$ are assessed via the following linear structural model:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, i = 1, 2, \ldots, N \tag{2}$$

where $\boldsymbol{\Pi}(q_1 \times q_1)$ and $\boldsymbol{\Gamma}(q_1 \times q_2)$ are unknown structural coefficient matrices, the distribution of $\boldsymbol{\xi}_i$ is assumed to be $N[\mathbf{0}, \boldsymbol{\Phi}]$, and $\boldsymbol{\delta}_i$ is a $q_1 \times 1$ vector of residuals, which followings $N[\mathbf{0}, \boldsymbol{\Psi}_\delta]$, $\boldsymbol{\Psi}_\delta$ is diagonal.

Exploratory structural equation modeling (ESEM)

The ESEM approach was proposed to introduce modeling flexibility to conventional SEM by incorporating the optimal features of exploratory factor analysis (EFA) into the SEM framework (Asparouhov & Muthén, 2009). Integrating EFA and CFA when establishing the measurement model (Equation 1) allows cross-loadings to be freely estimated, rather than being constrained to zero, which provides more flexible and realistic estimations for the factor structure (Marsh et al., 2014). Meanwhile, it reserves distinct features of conventional SEM, such as providing goodness-of-fit statistics and the estimation of relationship among factors via a structural model formulated in the same way as Equation 2. More technical details of the ESEM can be referred to Asparouhov and Muthén (2009). ESEM is also applicable to methodological advancements associated with CFA and SEM, including multiple-group analysis, measurement invariance, and latent growth modeling (Marsh et al., 2010, 2009).

Existing studies have demonstrated that ESEM outperforms the conventional SEM approach in some aspects. ESEM can provide closer-to-reality estimates of the factor structure owing to the relaxation of the constraint of zero cross-loadings, thereby generating more accurate estimations of factor correlations and structural regression coefficients (Mai et al., 2018; Marsh et al., 2010; Xiao et al., 2019). The more realistic factor structure of the ESEM can also facilitate a better fit to the data, which guarantees subsequent advanced analyses (e.g., measurement invariance and latent growth models) to be more appropriately conducted (Marsh et al., 2011; Sánchez-Carracedo et al., 2012). Moreover, ESEM accomplishes the task of searching all cross-loadings in one step, and circumvents potential pitfalls triggered by a sequence of modifications of the CFA model. Therefore, it is suggested to be a more appropriate and simpler alternative to the highly prevalent post-hoc model searches using modification indexes in CFA when the zero cross-loading assumption is unrealistic (Asparouhov & Muthén, 2009; Schmitt, 2011).

However, the feature of freely estimating all cross-loadings also triggers possible limitations in the ESEM. For example, such an approach may confound constructs that should be separated according to the substantive theory or specific research question (Marsh et al., 2019). In addition, the increased number of free parameters can undermine the parsimony of the model, especially when the model is large and the sample size is relatively small (Marsh et al., 2019, 2014).

Bayesian structural equation modeling (BSEM)

By applying the Bayesian analysis in the SEM framework, BSEM considers unknown parameters as random variables, rather than constants, and specifies each parameter with a prior distribution that reflects the knowledge of the parameter from theoretical hypotheses, previous studies, and other resources (e.g., pilot studies and experts) (B. Muthén & Asparouhov, 2012; Zondervan-Zwijnenburg et al., 2017). Using Bayesian analysis, the posterior distribution of the parameter can then be determined based on prior and observed data (Guo et al., 2019; B. Muthén & Asparouhov, 2012). By incorporating prior information into the analysis, the Bayesian approach can better reflect substantive theories; hence, it has the potential to improve estimation efficiency (Yuan & MacKinnon, 2009).

A typical BSEM analysis can be outlined in the following steps:

(1) Define a SEM model M with unknown parameters $\boldsymbol{\Theta}$, that is, the unknown parameters in $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Pi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}_\epsilon$, $\boldsymbol{\Psi}_\delta$ and $\boldsymbol{\Phi}$ in Equations 1 and 2.
(2) Assign prior distribution of $\boldsymbol{\Theta}$ under model M, say, $p(\boldsymbol{\Theta}|M)$.
(3) Let $\mathbf{Y}$ be the observed data, construct the posterior distribution of $\boldsymbol{\Theta}$, say, $p(\boldsymbol{\Theta}|\mathbf{Y}, M)$ by combining the sample information in the likelihood function $p(\mathbf{Y}|\boldsymbol{\Theta}, M)$ and prior information in $p(\boldsymbol{\Theta}|M)$ based on Bayes' Theorem.
(4) The posterior distribution $p(\boldsymbol{\Theta}|\mathbf{Y}, M)$ is usually approximated by using a Markov chain Monte Carlo algorithm, which iteratively simulates a large number of samples from $p(\boldsymbol{\Theta}|\mathbf{Y}, M)$. After discarding the samples obtained at the burn-in phase, Bayesian inference of $\boldsymbol{\Theta}$ can be carried out on the basis of simulated samples, for example, the Bayesian estimate, the numerical standard error estimates, as well as the highest posterior density (HPD) intervals of $\boldsymbol{\Theta}$.

Regarding cross-loadings in $\boldsymbol{\Lambda}$ of Equation 1, BSEM replaces the constraint of zero cross-loadings in conventional SEM with a "softer" constraint that specifies small-variance informative priors concentrated around zero for the cross-loadings. Let λ_{jk} be the factor loading in $\boldsymbol{\Lambda}$ for the j-th indicator associated with the k-th latent variables. As suggested by B. Muthén and Asparouhov (2012), the normal prior (also known as Bayesian ridge prior) is assigned to λ_{jk}, that is, $\lambda_{jk} \sim N\left(0, \sigma_{jk}^2\right)$, where the prior variance σ_{jk}^2 is given by researchers. If the λ_{jk} represents the cross-loading on non-targeted factor, σ_{jk}^2 is given small value, which makes the posterior distribution of λ_{jk} generally do not deviate largely from zero. Hence, the small-variance informative priors concentrated around zero for cross-loadings better reflect the underlying substantive theories and detects possible misspecifications than CFA (Asparouhov et al., 2015). Such an approach retains the CFA model while allowing cross-loadings to be nonzero (Murray et al., 2019). More technical details of BSEM including specification of priors for other unknown parameters can be referred to B. Muthén and Asparouhov (2012).

Owing to prior information, BSEM is particularly beneficial and flexible in applications where the model specification is not identified in a likelihood-based approach (B. Muthén & Asparouhov, 2012; Pan et al., 2017; Scheines et al., 1999). Moreover, model modification in conventional SEM with modification indexes can solely free one parameter at a time, which often triggers a long series of modifications, BSEM can free all parameters simultaneously; hence, it only requires a single-step modification analysis, which significantly reduces the risk of capitalizing on chance. BSEM is also appealing in studies with small samples because Bayesian analysis does not depend on large sample approximations, as conventional SEM using maximum likelihood (ML) estimation does (Yuan & MacKinnon, 2009). In addition, it can provide less inadmissible solutions, convergence issues, and more precise estimations when taking the prior information into consideration (Zondervan-Zwijnenburg et al., 2017). A large amount of evidence has demonstrated that Bayesian analysis performs better than that of ML methods including conventional SEM in small samples (Van de Schoot et al., 2017).

Because prior information is a key component in Bayesian analysis, the specification of priors is a potentially challenging issue in BSEM (MacCallum et al., 2012). The choice of priors substantially influences the estimates (Depaoli & Clifton, 2015; Van Erp et al., 2018). The influence of priors can be more prominent in small samples, where the application of Bayesian analysis is highly valued (Van de Schoot et al., 2017). Although it is time-consuming and challenging to construct proper informative priors, the possible bias resulting from vague or improper priors in BSEM deserves to be noted (Van Erp et al., 2018; Zondervan-Zwijnenburg et al., 2017).

Analysis of cross-loadings using SEM, ESEM, and BSEM

Conventional SEM with standard CFA as a measurement model has a strict assumption of zero cross-loadings, which has been determined to be overly restrictive in practice. To address this issue, ESEM allows all cross-loadings to be freely estimated, while BSEM specifies small-variance priors centered around zero for cross-loadings to relax such a constraint. Based on this notion, previous studies with both simulated and empirical data consistently found that ESEM and BSEM outperformed conventional SEM when non-ignorable cross-loadings existed (Asparouhov & Muthén, 2009; Mai et al., 2018; Marsh et al., 2010, 2014; B. Muthén & Asparouhov, 2012).

A few studies have compared the performance of ESEM and BSEM. Reis (2017) applied ESEM and BSEM to the empirical data of a multidimensional assessment and found that both approaches consistently revealed substantive cross-loadings in the instrument, which helps to improve the understanding of multidimensional constructs. Comparing the performance of BSEM and ESEM on simulated and empirical data, Guo et al. (2019) reported that ESEM outperformed BSEM with normal priors of cross-loadings (mean = 0, variance = .01) in unbalanced factor structures (only positive cross-loadings exist); however, the reverse was true for balanced factor structures (the sum of cross-loadings was zero for each factor). Another study found that with correct priors of cross-loadings (mean = the true value of cross-loadings, variance = .01), BSEM exhibited better performance than ESEM on simulated data (Xiao et al., 2019). However, in models with ordinal indicators, it was found that compared to BSEM, ESEM yielded less biased estimates for both measurement and structural parameters when the sample size was relatively large; however, it produced a higher rate of inadmissible solutions in small samples (Liang et al., 2020).

Juxtaposing ESEM and BSEM, although the studies above provide information about their performance under different conditions, these approaches have their inherent limitations. There is insufficient research to provide general guidelines on selecting the appropriate analytical approach among SEM, ESEM, and BSEM for analyzing structural regression models when the measurement model involves potentially unknown cross-loadings. To estimate the structural regression model where cross-loadings were present, Mai et al. (2018) compared the performance of SEM, ESEM, and manifest regression analysis, and ESEM was recommended. However, BSEM was not considered in their research; hence, we have insufficient knowledge on the relative performance between ESEM and BSEM when the structural regression model was considered. Liang et al. (2020) focused on the performance of ESEM and BSEM, and only priors with zero mean and smaller variance were considered for cross-loadings. However, the benefits of BSEM rely on correctly specified priors (Xiao et al., 2019; Zondervan-Zwijnenburg et al., 2017), and adopting a smaller variance prior indicates that a researcher was more confident about the cross-loading being close to zero. Priors with zero mean and relatively larger variance, which indicates less confidence, were not involved. However, as discussed in the study by B. Muthén and Asparouhov (2012), when the prior variance becomes larger, the model moves closer to being nonidentified.

The present investigation

The objective of the present investigation is to compare the performance of conventional SEM, ESEM, and BSEM in estimating structural models with potentially unknown cross-loadings. Specifically, we adopted simulated data to evaluate the performance of these three approaches under different manipulated conditions (i.e., sample size, target loading, cross-loading, and path coefficient).

Simulation

Design

The population model was adapted from the investigation conducted by Mai et al. (2018). It contains three latent variables, with two being exogenous (ξ_1 and ξ_2) and one being endogenous (η). The factor correlation between ξ_1 and ξ_2 is set as ρ = .30, which represents a medium correlation size (Cohen, 1988). In this model, a) the expectation of each error term is assumed to be zero, b) the error terms are independent of each other, and c) any of the error terms is independent of the latent variables ξ_1, ξ_2 and η. This study did not assume homoscedasticity for error terms. More details regarding the population model are presented in Figure 1.

This simulation study manipulated five variables, which include:

(1) The standardized values of target loadings are λ_T = .55, .70, .84, and .95, which represent different sizes of target loadings. These values approximate different levels of reliability of the indicator to the factor, which equals the square of the target loadings (i.e., .30, .50, .70, and .90) (Mai et al., 2018).
(2) The standardized values of the cross-loadings are λ_C = 0, .10, .20, and .30, with the latter three representing cross-loadings of little, some, and real importance (B. Muthén & Asparouhov, 2012). The standardized target loading and cross-loading for the same indicator should satisfy the restriction that $\lambda_T{}^2 + \lambda_C{}^2 + 2\rho\lambda_T\lambda_C \leq 1$.[1] Hence, when $\lambda_T = .95$, λ_C can only be 0 or .10.
(3) The standardized values of the path coefficients γ_1 and γ_2 can be 0, .14, .36, or .51, with the latter three representing small, medium, and large effect sizes (Preacher & Kelley, 2011) and $\gamma_2 \geq \gamma_1$. Therefore, there are 10 combinations of (γ_1, γ_2), that is (0, 0), (0, .14), (0, .36), (0, .51), (.14, .14), (.14, .36), (.14, .51), (.36, .36), (.36, .51), and (.51, .51).
(4) The sample size was N = 100, 200, and 500, representing small, medium, and large samples, respectively. A sample size of 100 is considered to be the minimum for SEM (Boomsma, 1982), and 500 is considered to be large enough to provide unbiased estimates for most applied studies (Kyriazos, 2018).
(5) Modeling approaches: Conventional SEM, ESEM, and BSEM. Both SEM and ESEM use maximum likelihood (ML) for estimation. For SEM, the models were estimated without cross-loadings or further adaptations based on the modification index. For ESEM, Geomin rotation, the default rotation method of Mplus, was used. Although other rotation methods can be considered, Geomin rotation is frequently adopted in ESEM and EFA, and it performs well for a two-factor EFA (Asparouhov & Muthén, 2009). For BSEM, we considered four conditions for informative priors of cross-loadings:
   a. BSEM-(CL, .01): The correctly specified prior of cross-loadings, in which the prior mean is the true value of cross-loading, and the prior variance is .01.
   b. BSEM-(0, .01): A prior with a small variance, in which the prior mean is 0 and the prior variance is .01.
   c. BSEM-(0, .04): A prior with medium variance, in which the prior mean is 0 and the prior variance is .04.
   d. BSEM-(0, .09): A prior with a large variance, in which the prior mean is 0 and the prior variance is .09. Although .09 is usually not considered a large prior variance in traditional Bayesian analysis, in the context of standardized cross-loadings in BSEM, a prior of zero mean and variance of 0.09 allows prior distribution to lie in [−0.59, 0.59]. Therefore, BSEM-(0, .09) was used to represent a condition of a large prior variance compared with the variances of 0.01 and 0.04 in this study.

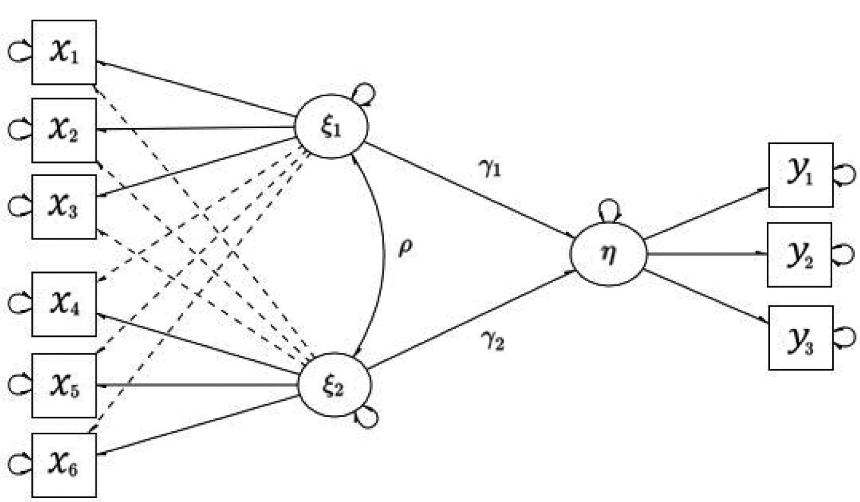

Figure 1. Population model of the simulated data.

[1] For an indicator, for example, x_1, loads concurrently on two latent variables (ξ_1 and ξ_2) $Var(x_1) = Var(\lambda_T \cdot \xi_1 + \lambda_C \cdot \xi_2) + Var(\varepsilon) = \lambda_T{}^2 Var(\xi_1) + \lambda_C{}^2 Var(\xi_2) + 2\lambda_T \cdot \lambda_C \cdot \rho \cdot \sqrt{Var(\xi_1)Var(\xi_2)} + Var(\varepsilon)$. Because the variances of the factors and indicators were equal to 1.0, $\lambda_T{}^2 + \lambda_C{}^2 + 2\lambda_T \cdot \lambda_C \cdot \rho = 1 - Var(\varepsilon)$. Because $Var(\varepsilon) \geq 0$, $\lambda_T{}^2 + \lambda_C{}^2 + 2\lambda_T \cdot \lambda_C \cdot \rho \leq 1$.

The priors in b), c), and d) mimic the situation in empirical studies where zero mean indicates the expectation of no cross-loadings, and different variances indicate the varied levels of confidence in the expectation. Note that when $\lambda_C = 0$, BSEM-(CL, .01), and BSEM-(0, .01) are the same.

In summary, these experimental factors resulted in a total of 2520 ($14 \times 10 \times 3 \times 6$) conditions. We simulated 500 data sets for each condition and fitted SEM, ESEM, and BSEM with each simulated data set separately.

Implementation

Monte-Carlo simulations were conducted using Mplus 7 (L. Muthén & Muthén, 1998). The SEM and ESEM models were estimated based on the default ML estimator and the default ML estimator and GEOMIN rotation, respectively. In BSEM, the default non-informative priors in Mplus are assigned to unknown parameters other than cross-loadings. Two parallel chains were used to check the convergence of the BSEM model with potential-scale reduction (PSR). For each replication, BSEM models were estimated where PSRs were verified to be less than 1.10 for all parameters. The Mplus scripts are available in Appendix 1 of the online supplemental materials.

Method

The performance of different modeling approaches was evaluated in terms of the following measures:

(1) Model convergence rate: The ratio of the number of converged replications to the total number of replications (500) under each experimental condition.
(2) Model rejection rate: The proportion of the number of replications where the model was rejected in the total number of converged replications under each experimental condition. The ML Chi-square tests were used for the SEM and ESEM, and a *p*-value less than 0.05 triggered a rejection of the model (Kline, 2015). The posterior predictive *p* (PP*p*) value was adopted to evaluate the model fit in BSEM, which quantified the discrepancy between the original data and the replicated data at Markov chain Monte Carlo (MCMC) iterations. A PP*p* value smaller than 0.05 indicates poor model fit, which triggers a rejection of the model (B. Muthén & Asparouhov, 2012).
(3) Relative bias of estimation (RBEST): The ratio of bias (i.e., the average difference between the estimated and true values across converged replications) to the true value, is a quantitative measure of estimation accuracy. A positive RBEST value indicates overestimation, whereas a negative value indicates underestimation. Generally, an acceptable RBEST ranges from −.1 to .1 and an ignorable RBEST ranges from −.05 to .05 (Hoogland & Boomsma, 1998).
(4) Relative bias for standard error (RBSE): The calculation is similar to RBEST with the standard deviation of estimates across converged replications considered as the true value, which evaluates the precision of estimations. An acceptable RBSE ranges from −.1 to .1 (Hoogland & Boomsma, 1998).
(5) Mean-square error (MSE): The average of the squared deviations of estimates, in which the deviations are calculated by subtracting the true value from the parameter estimates under each condition, which represents the trade-off between estimation accuracy and precision. A smaller MSE value indicates a more efficient parameter recovery.
(6) Statistical power: The rate of statistically significant coefficients in the totally converged replications under each condition. This calculation was conducted under the condition that the path coefficient was non-zero in the population.
(7) Type I error rate: The calculation is similar to statistical power, but was conducted under the condition that the true value of path coefficients was zero. We calculated the type I error rate for γ_2 in the conditions $\gamma_1 = 0$ and $\gamma_2 = 0$. The acceptable range of type I error rate is [.025, .075] ($\alpha = .05$) (MacKinnon et al., 2004).

Results

Because the results under different conditions of path coefficients were similar, as well as the results for γ_1 and γ_2, we selectively present the evaluation performance for γ_2 under the condition that $\gamma_1 = .14$ and $\gamma_2 = .36$, unless otherwise specified. The detailed results of the model evaluation under all conditions are available in Appendices 2–7 of the supplemental materials. Moreover, we only present the evaluation performance in terms of path coefficients in the main text because this study focuses on comparing different approaches for estimating structural models. Readers can refer to supplemental materials for the results regarding factor correlations.

Model convergence rate

Figure 2 illustrates the convergence rate of the model using different modeling approaches and conditions. As illustrated in Figure 2, BSEM-(CL, .01), BSEM-(0, .01), BSEM-(0, .04), and BSEM-(0, 09) demonstrate excellent convergence rates under all conditions (> .99). Although SEM exhibits slightly poorer convergence than BSEM under some conditions with small target loadings ($\lambda_T = .55$) and relatively large cross-loadings ($\lambda_C = .20,\ .30$), the lowest convergence rates of SEM are larger than .96. However, under some conditions of path coefficients (e.g., $\gamma_1 = .36$, $\gamma_2 = .36$), the convergence of SEM was generally better than that of BSEM when the target loadings were small and cross-loadings were relatively large. The detailed results of the model convergence rate under all conditions are presented in Appendix 2. ESEM demonstrated a salient decrease in the convergence rate with increasing cross-loadings, dropping to a convergence rate as low as .56, which was alleviated in larger samples and in data with larger target loadings.

Model rejection rate

Figure 3 presents the model rejection rates based on different modeling approaches and conditions. SEM consistently demonstrated higher rejection rates than the BSEM approach, followed by ESEM. The model rejection rates of SEM were

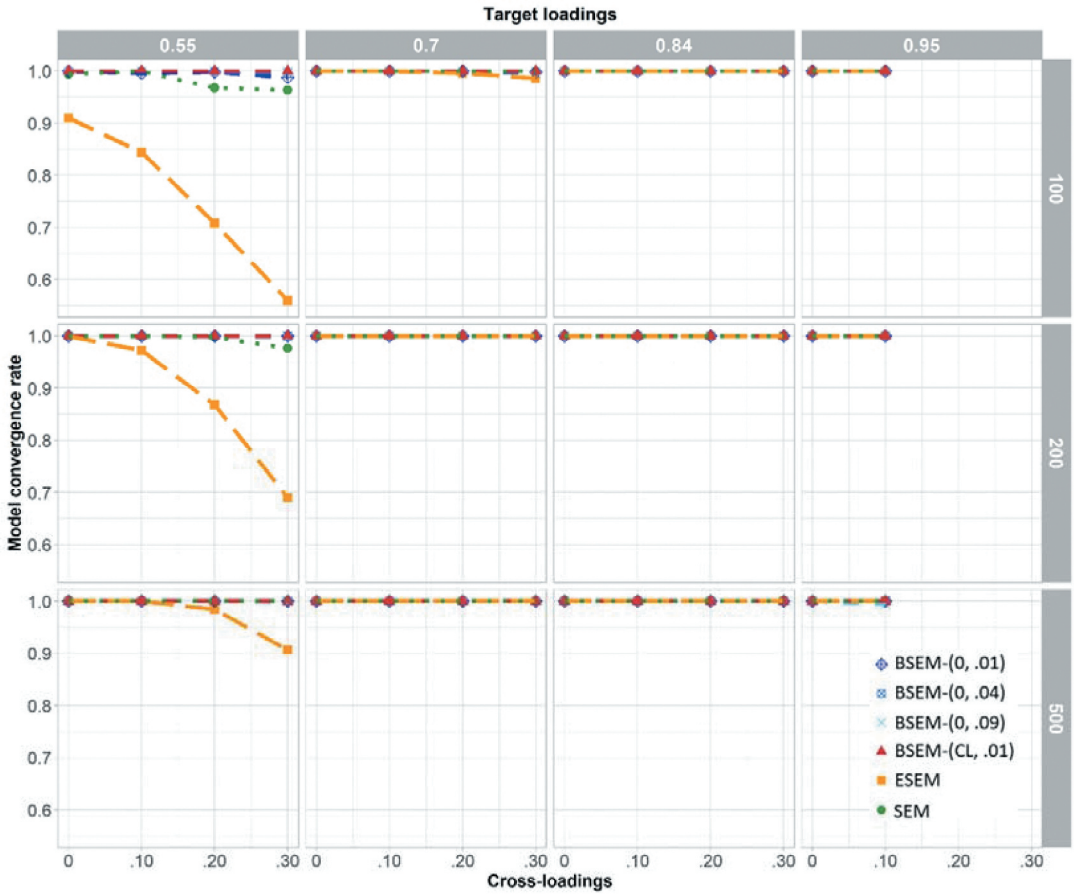

Figure 2. Model convergence rate with $\gamma_1 = .14$ and $\gamma_2 = .36$.

alleviated for larger sample sizes. Because SEM estimated models without any cross-loadings, it was logical that the resulting model exhibited relatively higher rejection rates in conditions with non-zero cross-loadings. Using a cutoff of .05 for PP*p*, the rejection rates of BSEM-(0, .01) were slightly enhanced as the cross-loadings increased in small samples with small target loadings ($\lambda_T = .55$). In comparison, BSEM-(CL, .01), BSEM-(0, .04), and BSEM-(0, .09) demonstrated stably low rejection rates close to zero under most conditions. Because PP*p* cutoffs of .10 and .01 are also reasonable (B. Muthén & Asparouhov, 2012), we conducted additional analyses for BSEM approaches using multiple cutoffs and found that the results were similar to those using a cutoff of .05 (Appendix 3, Table A3.11–A3.30).

Relative bias of estimation (RBEST)

Figure 4 presents the RBEST of γ_2 using different modeling approaches and conditions. In general, the RBEST of SEM and BSEM with cross-loading prior means of 0 substantively elevated as cross-loadings increased, which was alleviated by increased target loadings and sample sizes. A similar pattern was also observed in the performance of ESEM, except that the RBEST of ESEM exhibited relatively large fluctuations in the conditions of a small target loading with a medium sample size ($\lambda_T = .55$, $N = 200$). Compared with those of the other approaches, the RBEST of BSEM-(CL, .01) was consistently lower under most conditions. Except for the conditions of a small target loading with a small sample size ($\lambda_T = .55$, $N = 100$), BSEM-(CL, .01) could provide acceptable RBEST (i.e., [−.10, .10]), despite increased cross-loadings.

Relative bias for standard error (RBSE)

Figure 5 presents the RBSE of γ_2 using different modeling approaches and conditions. In conditions with concurrent small target loadings and small or medium sample sizes ($\lambda_T = .55$, $N = 100, 200$), all approaches fail to stably provide acceptable RBSE, regardless of the level of cross-loading. Under other conditions, SEM outperformed the other modeling approaches in providing precise path coefficients, thus exhibiting a more acceptable RBSE of γ_2, followed by BSEM-(CL, .01) and ESEM. BSEM with cross-loading prior means of 0 did not demonstrate evident advantages over other approaches, regardless of different prior variances.

Mean-square error (MSE)

Figure 6 presents the MSE of γ_2 using different modeling approaches and conditions. When the target loading was small ($\lambda_T = .55$), the MSE of SEM and BSEM with cross-loading prior means of 0 exhibited sharp enhancement as

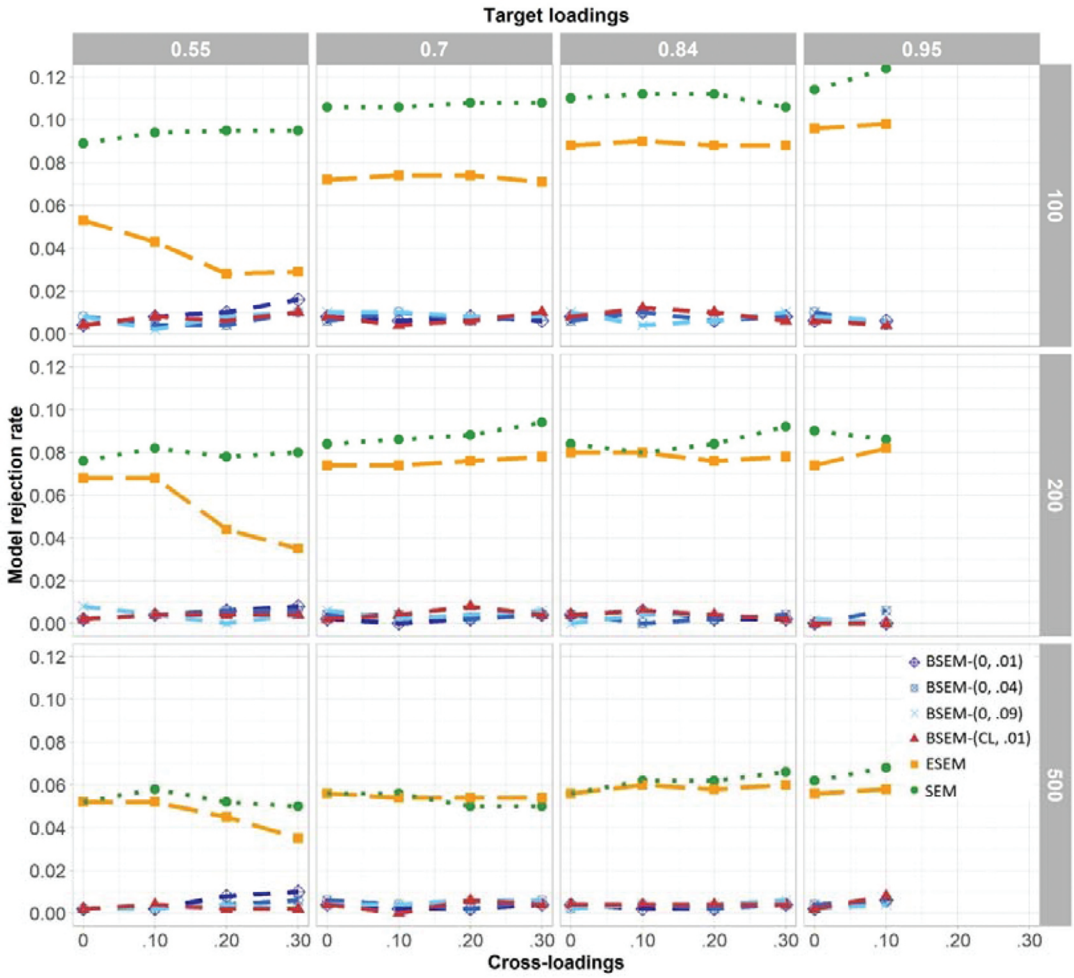

Figure 3. Model rejection rate with γ_1 = .14 and γ_2 = .36.

cross-loadings increased, some of which reached an extreme level. The MSE of the ESEM exhibited large fluctuations when the target loading was small, especially in the medium sample size (N = 200). However, as the target loadings increased, the performance in terms of the MSE of all modeling approaches was substantially improved. The MSE of BSEM-(CL, .01) was more stable than that of the other approaches, which increased occasionally; however, it remained low under most conditions, ranging from 0.003 to 0.141.

Statistical power

Figure 7 presents the statistical power of γ_2 using different modeling approaches and conditions. In general, the statistical power demonstrates a declining trend with increasing cross-loadings, and an increasing trend with increasing sample sizes and target loadings. In the case of zero cross-loadings, all modeling approaches demonstrated relatively similar statistical power. When cross-loadings existed, BSEM-(CL, .01) outperforms the other approaches under most conditions.

Type I error rate

Figure 8 presents the type-I error rate of γ_2 using different modeling approaches and conditions when γ_1 = 0 and γ_2 = 0. Detailed type I error rates are available in Appendix 8 of the supplemental materials. BSEM-(CL, .01) outperformed the other approaches and demonstrated acceptable type I error rates under all conditions, followed by BSEM-(0, .01), BSEM-(0, .04), and ESEM. When the target loadings are small, the type I error rate of the SEM substantially decreased as cross-loadings increased, reaching a lower-than-acceptable level. Compared with other approaches, BSEM-(0, .09) provided lower-than-acceptable type-I error rates under more conditions.

Discussion

In psychological research, it is important to investigate and understand the true factor structure before elucidating the relationships among different constructs via full SEM analysis. The potential cross-loadings complicate this process; hence, they may affect the estimation of the structural parameters in the full SEM model. Based on simulation studies covering different sample sizes, target loadings, cross-loadings, and structural regression coefficients, this study found that ESEM and BSEM provide flexibility in estimating all possible cross-loadings under different frameworks compared to conventional SEM. Our findings advance the existing literature by providing insights into the performance of conventional SEM, ESEM, and BSEM for estimating a structural model. These findings are summarized and discussed as follows

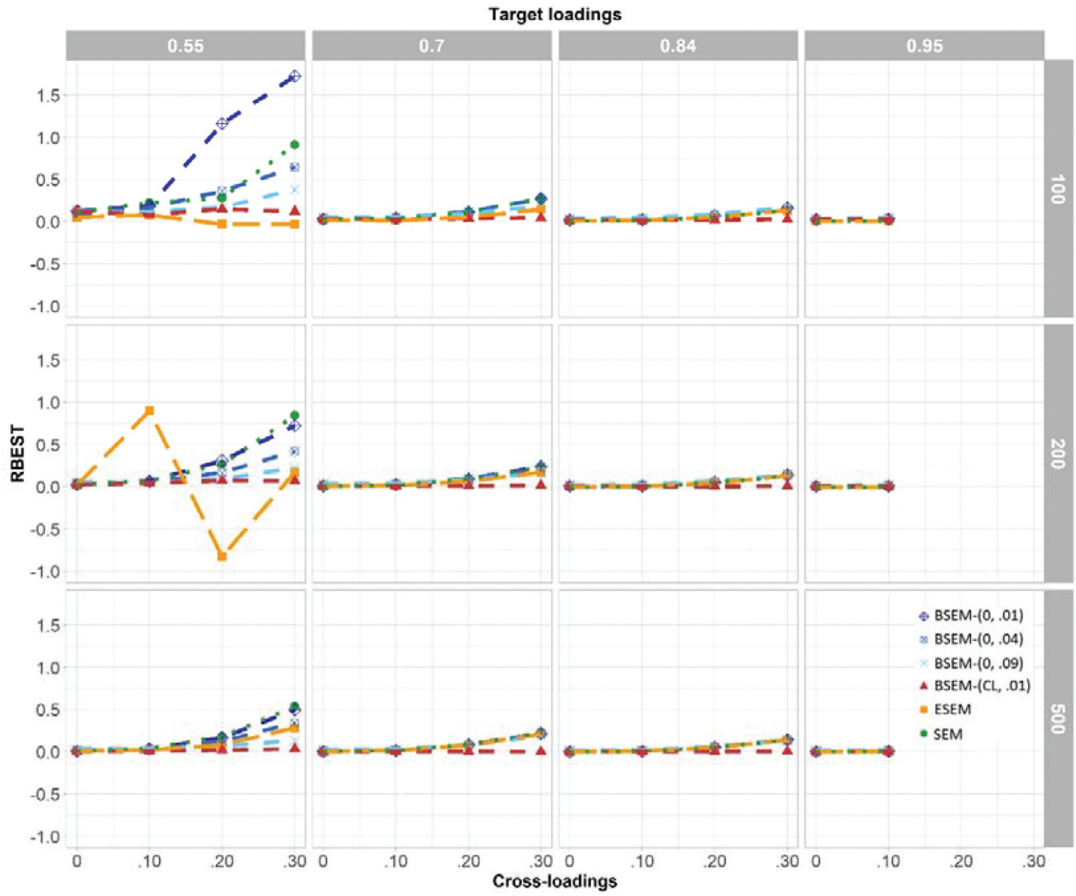

Figure 4. Relative bias of estimation (RBEST) of γ_2 with $\gamma_1 = .14$ and $\gamma_2 = .36$.

Major findings

In general, as expected by several researchers, the most commonly used SEM performed worse as cross-loadings increased, producing more biased structural path coefficient estimates and lower statistical power, which was substantially alleviated by enhanced sample sizes and target loadings. These findings are in line with the literature on the disadvantages of conventional SEM in analyzing models with cross-loadings (Asparouhov & Muthén, 2009; Marsh et al., 2014; B. Muthén & Asparouhov, 2012). However, under conditions with large target loadings ($\lambda_T \geq .84$) and medium-to-large sample sizes (N ≥ 200), the performance of SEM in estimating path coefficients is comparable to that of ESEM and BSEM. In this case, using SEM benefits the development of parsimonious instruments (Asparouhov & Muthén, 2009).

Although both approaches were proposed to appropriately address cross-loadings, ESEM produced lower convergence rates as cross-loadings increased and rejected models more frequently than BSEM in most cases. It is worth noting that when the target loadings were small ($\lambda_T = .55$), ESEM exhibited large fluctuations in the estimation performance of path coefficients, reaching an extreme level of bias under some conditions. When target loadings were not small ($\lambda_T \geq .70$), the performance of ESEM in estimating path coefficients was substantially improved; however, it still provides unacceptable estimation accuracy in some conditions with large cross-loadings ($\lambda_C = .30$), which was inferior to BSEM-(CL, .01). Xiao et al. (2019) found that BSEM outperformed ESEM when the prior means assigned to cross-loadings matched the true values. We infer that this conclusion can be generalized to the performance of ESEM in estimating structural models. Our findings are also consistent with previous findings by Liang et al. (2020).

Regarding BSEM, its performance in estimating structural regression parameters significantly depends on the accuracy of priors for cross-loadings in the present investigation, which is contrary to the conclusion of Liang et al. (2020) for BSEM with ordinal indicators. With a correctly specified prior mean and small prior variance, BSEM-(CL, .01) exhibited optimal performance (e.g., optimal model convergence and rejection, higher statistical power) under most conditions. Specifically, BSEM-(CL, .01) outperformed other approaches in stably providing precise and accurate structural path estimates, despite increasing cross-loadings. In comparison, with zero-mean cross-loading priors, BSEM performs better than SEM and ESEM in medium target loadings and a relatively small sample size. As discussed in the study by Guo et al. (2019), under the framework of the factor analysis model, the performance of BSEM improves substantially when informative priors for cross-loadings are set close to the population values. Xiao et al. (2019) also pointed out that the cross-loading mean was more influential than the prior variance in the BSEM estimation precision. Under the conditions considered in the present

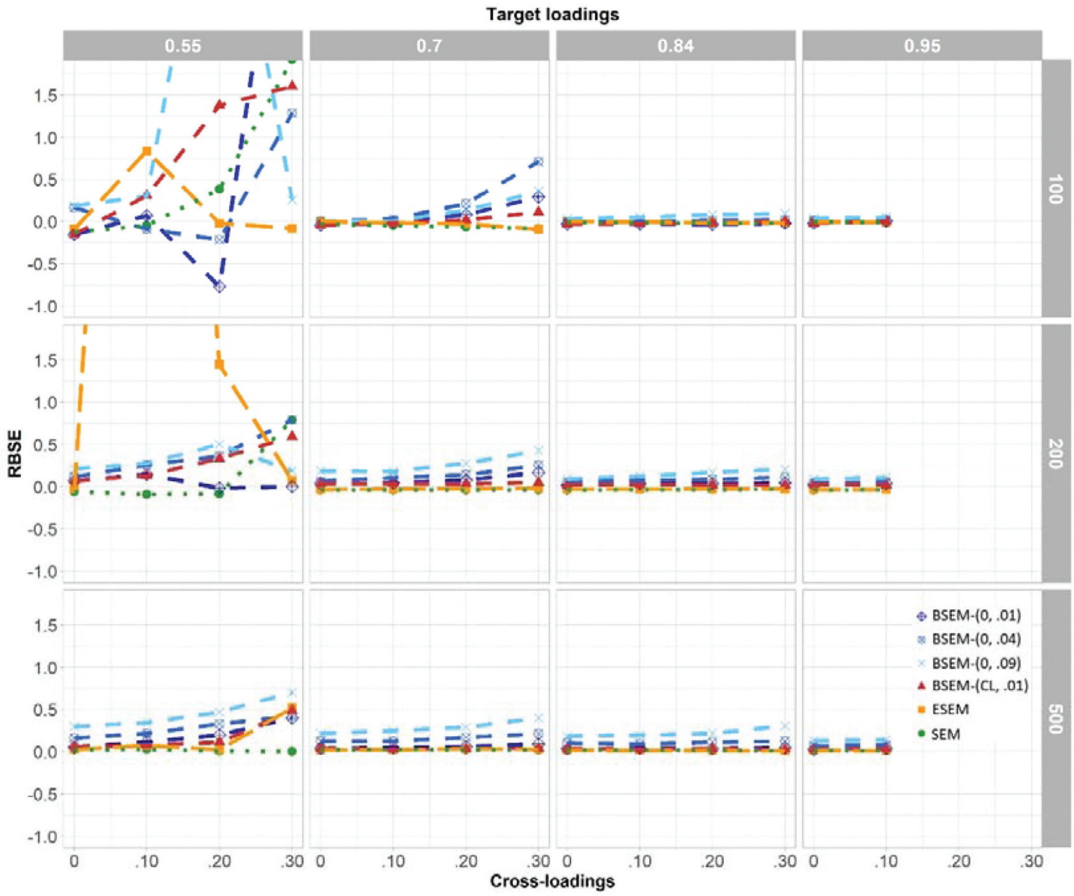

Figure 5. Relative bias for standard error (RBSE) of γ_2 with γ_1 = .14 and γ_2 = .36. The values of RBSE under certain conditions exceed the depicting range of the y-axis, which are displayed in the supplemental material.

study, we find that the above-mentioned conclusions hold for the estimation of structural regression parameters, which is also in line with previous findings regarding the importance of prior accuracy (Van Erp et al., 2018; Zondervan-Zwijnenburg et al., 2017).

Recommended strategies for method selection

Based on the performance of conventional SEM, ESEM, and BSEM under different conditions of target loadings, cross-loadings, and sample sizes in this investigation, we recommended the following strategies to select the appropriate modeling approach in practice for the applied researchers' consideration:

(1) Researchers are advised to gather prior information about target loadings and cross-loadings before choosing a modeling approach because the performance of different approaches largely depends on the conditions of target loadings and cross-loadings in the model. Available strategies include existing empirical studies, meta-analyses, and pilot data. Researchers may also conduct preliminary analyses with conventional SEM and make subsequent methodological decisions based on the results. A larger sample size is also recommended because the performance of most approaches can be substantively improved as sample size increases.

(2) When there were no cross-loadings in the model or the target loadings were substantially large (e.g., λ_T = .95), the performances of conventional SEM, ESEM, and BSEM are relatively similar. In this case, SEM is recommended to consider parsimony.

(3) If there is adequate information to specify appropriate priors for cross-loadings (e.g., well-supported evidence in the literature including meta-analyses, results of the pilot data), BSEM is recommended for its superior performance to other approaches. Specifically, if prior knowledge suggested that cross-loading was substantial, it might be more appropriate to specify a non-zero mean with small variance. However, it should be noted that it is difficult to guarantee the correctness of priors under the Bayesian framework, while existing studies including ours did not provide information regarding the impact of incorrect cross-loading priors. Researchers are advised to conduct prior sensitivity analyses and compare the results of different priors with caution.

(4) In the absence of adequate information on priors for cross-loadings, while target loadings are relatively large (e.g., $\lambda_T \geq$ .70), ESEM is recommended, provided the

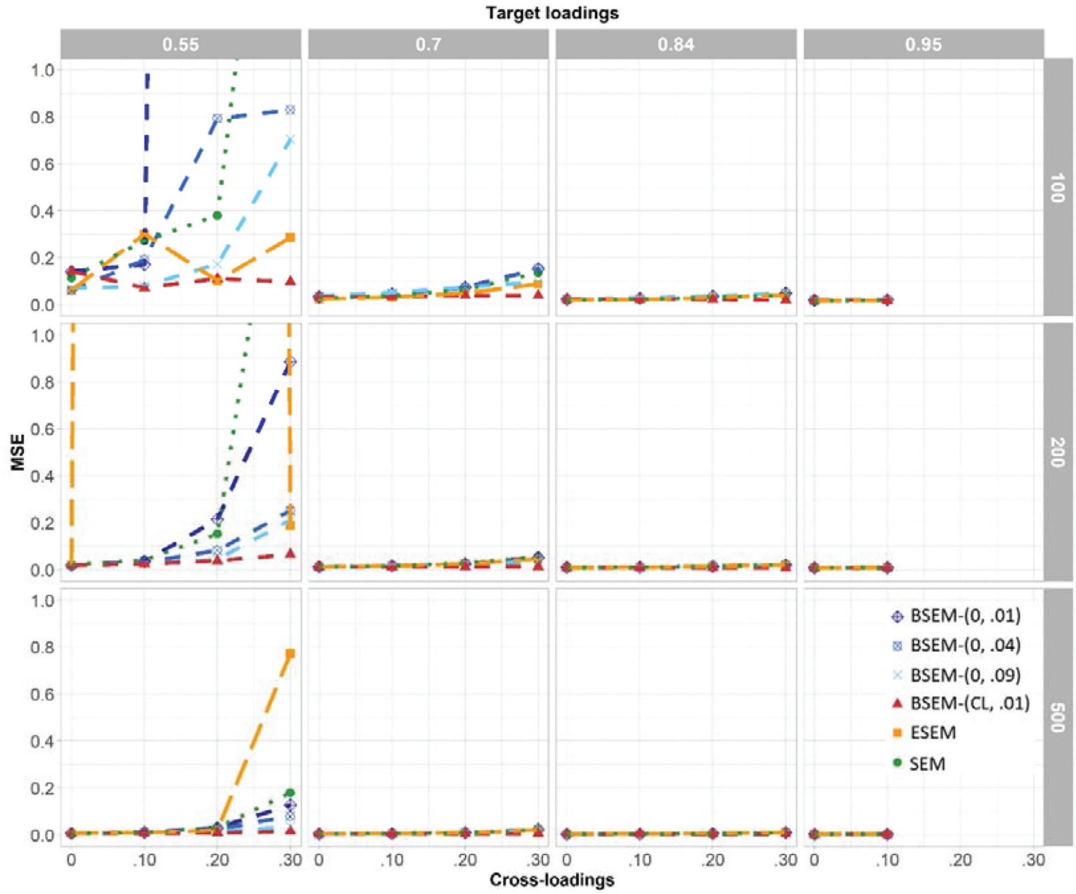

Figure 6. MSE of γ_2 with γ_1 = .14 and γ_2 = .36. The values of MSE under certain conditions exceed the depicting range of the y-axis, which are displayed in the supplemental material.

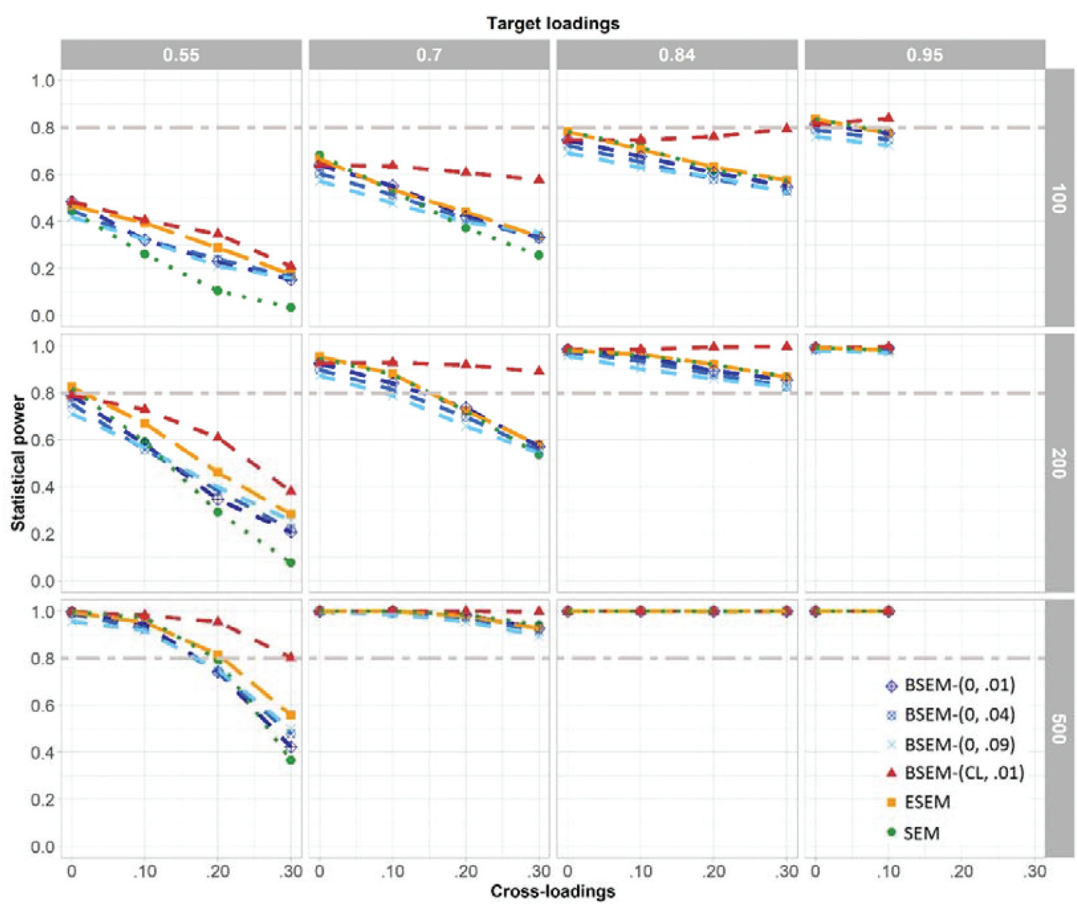

Figure 7. Statistical power of γ_2 with γ_1 = .14 and γ_2 = .36.

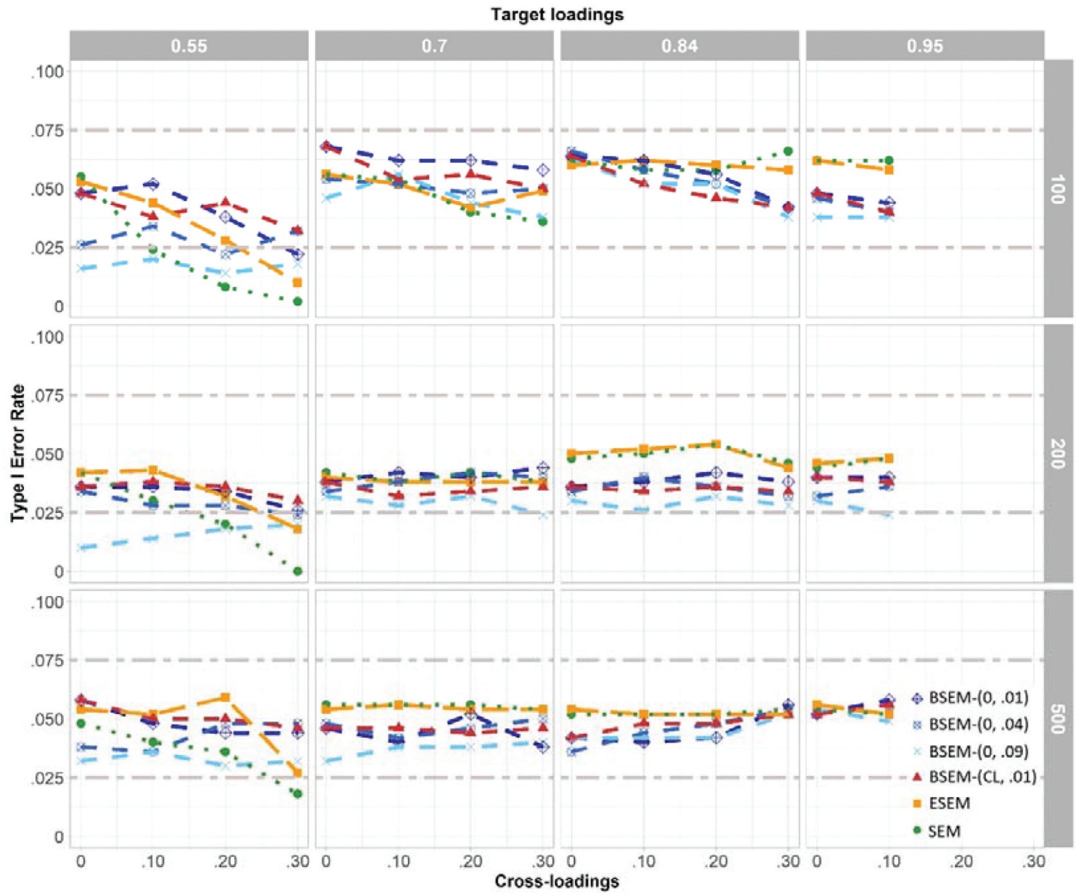

Figure 8. Type I error rate of γ_2 with γ_1 = .0 and γ_2 = .0. The acceptable range is [.025, .075].

convergence problem does not occur, owing to its practical convenience and relatively stable performance in estimating path coefficients under most conditions. If researchers are interested in identifying cross-loadings, ESEM should be preferable, despite the comparable performance of SEM in estimating path coefficients, because analyzing cross-loadings in SEM relies on modification index and might lead to capitalizing on chance. However, ESEM should be avoided when the target loadings are small (e.g., λ_T = .55), owing to the possible extreme biases in estimating the path coefficients.

Limitations and future directions

There are some limitations to this study. First, the data in this simulation study was based on a normal distribution. Future research should further compare the performance of SEM, ESEM, and BSEM on non-normal data. Second, this investigation did not evaluate each approach's performance under different levels of model complexity (e.g., the number of factors and number of indicators for each factor). Future studies should consider the impact of model complexity because it has been verified to influence model convergence and estimation accuracy (Xiao et al., 2019). Third, this investigation solely considered ESEM with Geomin rotation. Further studies are required to compare the performance of SEM, BSEM, and ESEM with other rotation methods. For example, ESEM with target rotation and Geomin rotation exhibited different patterns of performance in estimating the measurement models (Guo et al., 2019; Xiao et al., 2019). Fourth, the posterior predictive *p* (PP*p*) value was used to evaluate the model fit in BSEM in our study, considering implementation convenience in Mplus and the extensive use of PP*p* in previous BSEM studies. However, one of the reviewers suggested the prior-posterior predictive *p* value may provide more accurate model evaluation in the context of small-variance priors (Hoijtink & van de Schoot, 2018). Further studies are needed to evaluate the performance of the prior-posterior predictive *p* value in the BSEM framework. Finally, it is difficult to specify a perfectly correct BSEM in empirical applications. Zondervan-Zwijnenburg et al. (2017) provided strategies to specify priors based on existing knowledge. More work is required to investigate how we can obtain more accurate priors in the SEM framework. In addition, while existing studies including ours implicated the importance of cross-loading prior means, the impact of incorrect while non-zero prior means for cross-loadings on BSEM remains largely unknown. Future studies may also consider addressing this issue in the context of cross-loadings.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded by the National Natural Science Foundation of China [31871128] and the MOE (Ministry of Education) Project of Humanities and Social Science of China [18YJA190013].

ORCID

Lijin Zhang http://orcid.org/0000-0002-4222-8850
Junhao Pan http://orcid.org/0000-0001-7156-6077

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et. *Journal of Management*, *41*, 1561–1577. https://doi.org/10.1177/0149206315591075

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 149–173). North-Holland.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. https://doi.org/10.1080/10705511.2014.937849

Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Lüdtke, O., & Diallo, T. M. O. (2019). A systematic evaluation and comparison between exploratory structural equation modeling and Bayesian structural equation modeling. *Structural Equation Modeling*, *26*, 529–556. https://doi.org/10.1080/10705511.2018.1554999

Hoijtink, H., & Van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive p values. *Psychological Methods*, *23*, 561–569. https://doi.org/10.1037/met0000131

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*, 329–367. https://doi.org/10.1177/0049124198026003003

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *09*, 2207. https://doi.org/10.4236/psych.2018.98126

Liang, X., Yang, Y., & Cao, C. (2020). The performance of ESEM and BSEM in structural equation models with ordinal indicators. *Structural Equation Modeling*, *27*, 1–14. https://doi.org/10.1080/10705511.2020.1716770

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological Methods*, *17*, 340–345. https://doi.org/10.1037/a0027131

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504. https://doi.org/10.1037/0033-2909.111.3.490

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling*, *25*, 737–749. https://doi.org/10.1080/10705511.2018.1444993

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2019). Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and Set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research*, *55*, 1–18. https://doi.org/10.1080/00273171.2019.1602503

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29*, 322–346. https://doi.org/10.1177/0734282911406657

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. https://doi.org/10.1037/a0019227

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*, 439–476. https://doi.org/10.1080/10705510903008220

Murray, A. L., Booth, T., Eisner, M., Obsuth, I., & Ribeaud, D. (2019). Quantifying the strength of general factors in psychopathology: A comparison of CFA with maximum likelihood estimation, BSEM, and ESEM/EFA bifactor approaches. *Journal of Personality Assessment*, *101*, 631–643. https://doi.org/10.1080/00223891.2018.1468338

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. https://doi.org/10.1037/a0026802

Muthén, L., & Muthén, B. (1998). *Mplus 7*. Muthén & Muthén.

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods*, *22*, 687. https://psycnet.apa.org/doi/10.1037/met0000112

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93–115. https://doi.org/10.1037/a0022658

Reis, D. (2017). Further insights into the German version of the multidimensional assessment of interoceptive awareness (MAIA): Exploratory and Bayesian structural equation modeling approaches. *European Journal of Psychological Assessment*, *35*, 317–325. https://doi.org/10.1027/1015-5759/a000404

Sánchez-Carracedo, D., Barrada, J. R., López-Guimerà, G., Fauquet, J., Almenara, C. A., & Trepat, E. (2012). Analysis of the factor structure of the sociocultural attitudes towards appearance questionnaire (SATAQ-3) in Spanish secondary-school students through exploratory structural equation modeling. *Body Image*, *9*, 163–171. https://doi.org/10.1016/j.bodyim.2011.10.002

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37–52. https://doi.org/10.1007/BF02294318

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*, 304–321. https://doi.org/10.1177/0734282911406653

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*, 217–239. https://doi.org/10.1037/met0000100

Van Erp, S., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*, 363–388. https://doi.org/10.1037/met0000162

Xiao, Y., Liu, H., & Hau, K.-T. (2019). A comparison of CFA, ESEM, and BSEM in test structure analysis. *Structural Equation Modeling*, *26*, 665–677. https://doi.org/10.1080/10705511.2018.1562928

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322. https://doi.org/10.1037/a0016972

Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & Van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*, 305–320. https://doi.org/10.1080/15427609.2017.1370966