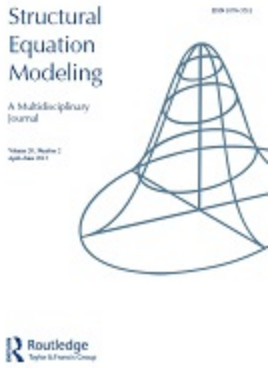

Criteria for Parameter Identification in Bayesian Lasso Methods for Covariance Analysis: Comparing Rules for Thresholding, p-value, and Credible Interval

Journal:	*Structural Equation Modeling: A Multidisciplinary Journal*
Manuscript ID	Draft
Manuscript Type:	Research Article
Keywords:	confirmatory factor analysis, cross-loading, residual correlation, regularization
Abstract:	The lasso is a commonly used regularization method that is increasing used in structural equation models (SEMs). Under the Bayesian framework, lasso is rendered more flexible and readily produces estimates of standard error and the penalty parameter. However, in practice, it remains unclear what decision rule is appropriate for parameter identification; in other words, determining what size estimate is large enough to be included into the model. The current study compared three decision rules for parameter identification—thresholding, p-value, and credible interval in confirmatory factor analysis. Specifically, two distinct parameter spaces were studied: cross-loadings and residual correlations. Results showed that the thresholding rule performed best in balancing power and Type I error rate. Different thresholds for standardized estimates were needed for different conditions. Guidelines for parameter identification and recommended thresholding values were also provided. Results of the current study have the potential to extend to a broad range of SEMs.

Criteria for Parameter Identification in Bayesian Lasso Methods for Covariance Analysis: Comparing Rules for Thresholding, *p*-value, and Credible Interval

Abstract

The lasso is a commonly used regularization method that is increasing used in structural equation models (SEMs). Under the Bayesian framework, lasso is rendered more flexible and readily produces estimates of standard error and the penalty parameter. However, in practice, it remains unclear what decision rule is appropriate for parameter identification; in other words, determining what size estimate is large enough to be included into the model. The current study compared three decision rules for parameter identification—thresholding, *p*-value, and credible interval in confirmatory factor analysis. Specifically, two distinct parameter spaces were studied: cross-loadings and residual correlations. Results showed that the thresholding rule performed best in balancing power and Type I error rate. Different thresholds for standardized estimates were needed for different conditions. Guidelines for parameter identification and recommended thresholding values were also provided. Results of the current study have the potential to extend to a broad range of SEMs.

Keywords: confirmatory factor analysis, cross-loading, residual correlation, regularization.

Introduction

Modern data analysis must often manage a large amount of data. When the number of variables is large and the sample size is insufficient, the problem of over-fitting occurs and leads to weakened model generalizability. To create a balance between model simplicity and model fit, regularization methods such as lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) have been used in many fields and increasingly in social sciences (Lindstrøm & Dahl, 2020). The idea behind lasso is to penalize models that are overly complex (i.e., that contain a large number of parameters) by using a more conservative criterion than traditional methods such as maximum likelihood or least square. To illustrate the principle of lasso, consider the following example for fitting a regression model with J predictors X to outcome data Y by minimizing the estimation function:

$$L^{\text{Lasso}}(\beta) = (|\text{Y} - \text{X}\beta|)^2 + \lambda\sum_{j=1}^{J}|\beta_j| \quad (1)$$

where $\beta(J \times 1)$ indicates the vector of regression coefficients for the J predictors, and $L^{\text{Lasso}}(\beta)$ and $(|\text{Y} - \text{X}\beta|)^2$ represent the loss function of lasso regression and least square difference, respectively. The term $\sum_{j=1}^{J}|\beta_j|$ is the lasso penalty function that consists of the sum of absolute values of regression coefficients. The tuning parameter, $\lambda(\geq 0)$, indicates the strength of the penalty. The addition of the penalty term to the least square criterion tends to result in simpler models that only include strong predictors.

Compared to models without a penalty function, models with lasso regularization have proved to have higher generalizability (Yarkoni & Westfall, 2017). Compared the ridge regularization method, which uses a quadratic penalty function $\sqrt{\sum_{j=1}^{J}\beta_j^2}$ and enforces less shrinkage strength on smaller coefficients, lasso has been found to perform better in variable selection (Hesterberg et al., 2008).

The Bayesian lasso version has been increasingly used in the social sciences (Brandt et al., 2018; Feng et al., 2015; Pan, Ip, & Dubé, 2017), and this trend is expected to continue (McNeish, 2015) for several reasons. First, lasso can be readily applied in Bayesian analyses by using the double exponential priors (Park & Casella, 2008). Second, with advances in Bayesian computational tools such as the Gibbs sampler, the Bayesian lasso method can provide estimates of standard errors that are difficult to obtain under a frequentist framework (Kyung et al., 2010). Third, the tuning parameters, which are traditionally chosen based on computationally intensive methods such as cross-validation method and multiple model comparison, can be more conveniently estimated with other coefficients simultaneously under the Bayesian lasso paradigm (Hans, 2009; Park & Casella, 2008).

Recently, novel methods that extend the lasso have been developed for structural equation modeling (SEM) and network analysis. Examples include the graphical lasso network models (Costantini et al., 2019), network analysis with an adaptive lasso method (Marcus, Preszler, & Zeigler-Hill, 2017), the Bayesian lasso confirmatory factor analysis (CFA; Chen et al., 2020; Pan, Ip, & Dubé, 2017), exploratory

mediation analysis (Serang et al., 2017), the Bayesian adaptive lasso for ordinal regression with latent variables (Feng, Wu, & Song., 2015), and regularized multiple-indicators and multiple-caused (MIMIC) models (Jacobucci, Brandmaier, & Kievit, 2019). An R-package *regsem* was also developed by Jacobucci et al. (2020) to conduct regularized SEM. Lasso and adaptive lasso methods are available under the frequentist framework in *regsem*.

Although the application of lasso has sharply increased, decision rules for parameter identification (i.e., for determining whether a coefficient is non-zero and included in the model) tend to vary greatly across studies. At least three criteria for parameter identification—the thresholding rule, the *p*-value rule, and the credible interval rule—have been used in the literature. The lack of a standard for reporting of results for parameter identification creates confusion for result interpretation and problems such as difficulty in comparing results across studies.

In this paper, we compare the three criteria for identifying parameter for inclusion into a model using the Bayesian lasso. Specifically, we focus on confirmatory factor analysis (CFA) and examine two parameter spaces that are amenable to regularization method: cross-loadings and residual correlations. To illustrate the research question, we use a data set regarding burnout in elementary school male teachers (N=372) (Byrne 1994, 2012). Participants were asked to respond to the 22-item Maslach Burnout Inventory (MBI; Maslach & Jackson, 1981, 1986). The MBI used a 7-point Likert scale (0 = feeling has never been experienced, 6 = feeling experienced daily)

and includes three dimensions of burnout. Figure 1(a) shows the factor structure specified in original articles. While it is unlikely that each item is only loaded on one factor, as specified a priori, it is also unlikely that all items significantly load on all factors. In other words, we expect sparsity in the cross-loading pattern, which contains a total of 44 possible cross-factor loading.

Insert Figure 1 about here

By applying the Bayesian lasso to the data, we obtained estimates for all cross-loadings. Figure 1(b) shows the results of applying the three-parameter identification criteria—threshold 0.1, *p*-value and 95% highest posterior density (HPD) interval—to the estimated values. Although there is some agreement of identified cross-loadings among the criteria, the numbers of identified significant cross-loadings were different —i.e., 11, 6, and 4 cross-loadings were respectively identified. Without knowing the true loading values, it is not possible to tell which method works best.

In this paper, our goal is to evaluate, through extensive simulation experiments, the performance of the three criteria under different conditions. We apply the Bayesian lasso to regularize both cross-loadings and residual correlations (both within- and across-factor). Based on the results of the evaluation, we provide further recommendations on the decision rules for using the Bayesian lasso CFA.

The reminder of the paper is organized as follows. First, we provide background for the Bayesian lasso for CFA and the three criteria and decision rules for parameter identification. We next describe the design for the simulation study. Results are reported by two sets of parameters—cross-loading and residual correlation, and three metrics—power, Type I error, and percentage of correct identification. Finally, we provide a discussion of the findings.

Bayesian Lasso Confirmatory Factor Analysis

The Bayesian lasso method has been used in CFA to detect possible cross-loadings of indicators (Chen et al., 2020) as well as to identify non-zero residual covariances (Pan, Ip, & Dubé, 2017; Zhang, Pan, Dubé, & Ip, in press). Suppose there are *J* items and *N* participants, the CFA is specified as follows:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \quad i = 1,2,...,N \tag{2}$$

where $\boldsymbol{y}_i$ $(J \times 1)$ represents the observed values of *J* items for the i-th participant. The vector $\boldsymbol{\mu}$ represents the intercept, and $\boldsymbol{\Lambda}$ represents the loading matrix that encode the relationship between observed variables in $\boldsymbol{y}_i$ and latent variables in $\boldsymbol{\omega}_i$. The error term $\boldsymbol{\varepsilon}_i$, which follows $N[0,\boldsymbol{\Psi}_\epsilon]$ **,** represents residuals of observed variables. In traditional CFA models, the following two assumptions are made: (1) the structure of loadings matrix is specified a priori according to substantive theory such that each item is loaded on one and only one latent factor, and (2) the residual variance-covariance matrix $\boldsymbol{\Psi}_\epsilon$ is diagonal. In other words, all cross-loadings and residual covariances are assumed to be zero. These two assumptions are rather

restrictive and not likely to be entirely satisfied in real-world applications. If non-zero parameters exist and are not identified and properly handled, it could result in an inferior model fit (Muthén & Asparouhov, 2012), and the structural estimate could be biased. For example, Brandt et al. (2020) found that even missing a single cross-loading can cause an unacceptably large bias of interaction effect. Similarly, it has been shown that correctly identified non-zero residual covariances lead to improved precision in the estimates of structural parameters (Pan, Ip, & Dubé, 2017).

Under the Bayesian framework, the strict assumptions of no cross-loading and zero residual covariance are relaxed through the assignment of priors to the corresponding parameters. Specifically, the Bayesian lasso assigns double exponential priors to the parameter, thus allowing cross-loadings and residual covariances to fluctuate around zero. The Bayesian lasso method simultaneously identifies non-negligible cross-loadings and residual covariances in a joint estimation procedure.

Because regularization methods work by shrinking many small parameters toward zero and only retain significantly large non-zero entries, the technique is especially useful when one expects the parameter space to be sparse (e.g., when only a few non-zero cross-loadings exist). Note that subsequent to using the Bayesian lasso or another regularized method, it is quite common to reanalyze the data with the identified freed (non-zero) parameters to improve fit and avoid bias and underestimation (Muthén & Asparouhov, 2012; Serang et al., 2017; Serang & Jacobucci, 2020).

The identification of "significantly large" non-zero entries require careful operationalization. While the Bayesian lasso can produce probabilistic measures such as a credible interval or *p*-value as evidence of a non-zero parameter value, a user is still required to determine whether a parameter should be included into the model, a process we call "parameter identification." The process sets parameters that are deemed small to zero.

Different Ways of Parameter Identification with the Lasso Method

In general, there are three common ways of parameter identification in lasso: (1) the thresholding rule; (2) the *p*-value rule; and (3) the credible interval rule.

When standard error is not available or difficult to derive (e.g., in a frequentist estimation), it is common to apply a simple thresholding rule to the (absolute) value of the estimated parameter for parameter identification. Apparently, an intuitive cutoff value is zero or a small value. Note that in practice, using a zero-thresholding rule is equivalent to including all parameters, however small the values. In a frequentist setting, Serang et al. (2017) compared lasso with the zero-thresholding rule and other *p*-value based methods (such as multivariate delta method) in an exploratory mediation analysis. Results showed that lasso outperformed *p*-value-based methods in correctly identifying the mediators. Liang and Jacobucci (2020) also chose zero as the threshold in detecting measurement bias in maximum likelihood estimation. Yuan and Liu (2020) adopted the same threshold to select loadings in CFA and found that lasso did not perform well under this situation. Unlike the other studies, Serang and

Jacobucci (2020) chose 0.001 instead of 0 as threshold when selecting mediators to exclude the negligible effect.

Within the Bayesian estimation, some researchers have opted to use the cutoff of 0.1 ($|\beta| > 0.1$). This cutoff value is especially popular with the Bayesian lasso method (Guo et al., 2012; Hoti and Sillanpää, 2006; Feng et al., 2015; Peterson et al., 2014). The cutoff value can also be justified from a substantive standpoint. Muthén and Asparouhov (2012) suggested that cross-loading of less than 0.1 can be considered to have little practical importance. Cohen (1988) also pointed out that the correlation coefficient as 0.1 is a typical value of low-effect size. Although substantively the cutoff value of 0.1 seems reasonable, the criterion may not have sufficient power to detect truly non-zero parameters. This is due partly to the shrinkage effect of regularization methods—that is, it tends to deflate parameters in achieving a high degree of parsimony.

In frequentist lasso applications when the standard error is available, the *p*-value (with cutoff at 0.05) is the conventional standard for parameter identification. For lasso regression, the *p*-value can be obtained using the R-package *covTest* (Lockhart et al., 2014). For the Bayesian lasso method, it is relatively straightforward to obtain *p*-values. Additionally, credible intervals such as the HPD interval can also be calculated using Markov Chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1996) algorithms. In network analysis, Epskamp et al. (2018) demonstrated how frequentist-lasso regularization can provide confidence intervals

using a bootstrap method. The authors further argued that because the distribution of Lasso regularized parameters is far from normal (Pötscher & Leeb, 2009), thus using the bootstrapped confidence interval is more appropriate than using the *p*-value.

When applying lasso to SEM, multiple testing of parameter significance is common. There does not appear to be consensus as to whether the α level should be corrected. Epskamp et al. (2018) pointed out the α level would be corrected to 0.000003 even with a small 20-node network if the Bonferroni correction is used. Such stringent criterion is not practical and thus not recommended by the authors. Pan et al. (2017) adopted the nominal α level at 0.05 when calculating the HPD intervals of residual covariances and demonstrated that a 95% HPD interval could maintain low Type I error rates. In this paper, the nominal α level at 0.05 is used throughout.

Simulation study

The Monte Carlo simulation study was conducted to investigate the performance of three criteria—thresholding, p-value, and credible interval—as well as their associated decision rules for detecting significant cross-loading/residual correlation under a range of experimental conditions. For each condition, 100 replications were generated and each sample was analyzed by Bayesian lasso CFA.

Method

Data generation

The following factors were selected as conditions for the Monte Carlo study for cross-loading: sample size (3 levels), model size (2 levels), and the magnitude of cross-loading (4 levels). Similar conditions were used for residual correlations (Table 1). As such, we had a total of $3 \times 2 \times 4 = 24$ conditions for each set of parameters or a total of 48 conditions for cross-loading and residual correlation. For model size, two measurement models of different number of factors were considered: a two-factor model with 5 items per factor and a three-factor model with 6 items per factor. These model sizes were commonly used in previous empirical and simulation studies (Chen et al., 2020; Khong, Onyemeh, & Chong, 2013; Muthén & Asparouhov, 2012). The variances of factors and items were set at one and the main loadings were set at 0.7.

Apart from the conditions of zero cross-loading and residual correlation, we considered three levels of non-zero effect size. Denote the factor loading between item a and factor b $\lambda_{a,b}$. Following Lu, Chow, & Loken (2016) and Muthén & Asparouhov (2012), two or 20% cross-loadings ($\lambda_{6,1}$,$\lambda_{5,2}$) in the two-factor model, and six or 16.67% cross-loadings ($\lambda_{7,1}$,$\lambda_{18,1}$,$\lambda_{6,2}$,$\lambda_{13,2}$,$\lambda_{1,3}$,$\lambda_{12,3}$) in the three-factor model (Fig. 2) were set at values of 0.1 / 0.2 or 0.3. Similarly, 2 or 3.64% out of 55 ($= C(10,2)$) off-diagonal elements in the residual variance-covariance matrix were non-zero in the two-factor model. Six or 3.51% out of 171 ($= C(18,2)$) residual covariances were non-zero in the three-factor model. The residual correlations (Fig. 2)

were set at 0.1 / 0.3 or 0.7 to represent zero, low, medium, and high correlation (Muthén & Asparouhov, 2012; Pan et al., 2017).

To avoid the possible confounding effect between non-trivial cross-loadings and residual correlations, the conditions of non-zero cross-loadings and non-zero residual correlations were separately analyzed. In other words, when the purpose was to detect significant cross-loadings, the non-diagonal parameters of residual variance-covariance matrix were set to zero in both data generation and model estimation and vice versa.

Insert Table 1 and Figure 2 about here

Model Estimation

The specific models used for the simulation studies were: (1) M1: model with some non-zero cross-loadings (Fig. 2) and diagonal residual covariance matrix, and (2) M2: model with some non-zero, off-diagonal residual covariance entries but no cross-loading.

Using the MCMC algorithm, the posterior distribution of parameters can be estimated based on the specified priors and the data. Suppose there are J items and K factors for the confirmatory factor analysis, the priors for the Bayesian lasso confirmatory factor analysis are issued as follows:

$$\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \mathbf{H}_{\mu 0}), \boldsymbol{\Phi}^{-1} \sim \text{Wishart}(\mathbf{R}_0, \rho_0) \quad (3)$$

For M1, the loadings for the j-th item $\mathbf{\Lambda}_j = \begin{pmatrix} \mathbf{\Lambda}_j^m \\ \mathbf{\Lambda}_j^c \end{pmatrix}$, where $\mathbf{\Lambda}_j^m$ and $\mathbf{\Lambda}_j^c$ respectively represent the component of the loading on the designated factor and the component of cross-loading. The hierarchical priors for loadings are specified by:

$$\mathbf{\Lambda}_j^m \sim N(\mathbf{\Lambda}_{0j}, \mathbf{H}_{0j}),\ \mathbf{\Lambda}_j^c \mid \psi_{jj} \sim N(0, \psi_{jj}\boldsymbol{D}_{\tau_j}),\ \psi_{jj}^{-1} \sim Gamma(a_{0j}, b_{0j}),$$

$$\boldsymbol{D}_{\tau_j} = diag(\tau_{j1}^2, \ldots, \tau_{jk}^2),\ \tau_{jk}^2 \sim Gamma\left(1, \frac{\delta^2}{2}\right),\ \delta^2 \sim Gamma(a_{lj}, b_{lj}) \quad (4)$$

For M2, the prior for the loadings is $\mathbf{\Lambda}_j \sim N(\mathbf{\Lambda}_{0j}, \mathbf{H}_{0j})$. Following Khondker et al. (2013) and Wang (2012), graphical lasso priors are specified for the residual variance-covariance matrix $\mathbf{\Psi}$:

$$\mathbf{\Psi}^{-1} = \mathbf{\Sigma} = (\sigma_{jj'})_{J \times J} \quad (5)$$

where independent exponential priors $(\frac{\lambda}{2}\exp(-\lambda|\sigma_{jj'}|), j < j'$ and the double exponential priors $(\frac{\lambda}{2}\exp\left(-\frac{\lambda}{2}\sigma_{jj}\right))$ are assigned for the diagonal and the off-diagonal elements of $\mathbf{\Psi}^{-1}$, respectively. Moreover, $\lambda \sim Gamma(a_{\lambda 0}, b_{\lambda 0})$ where $a_{\lambda 0} = 1$ and $b_{\lambda 0}$ set at a small value.

For sensitivity analysis, we adopted two sets of values for hyperparameters $\boldsymbol{\mu}_0$, $\mathbf{H}_{\mu 0}$,$\mathbf{R}_0$,ρ_0,$\mathbf{\Lambda}_{0j}$, $\mathbf{H}_{0j}$,a_{0j},b_{0j},a_{lj},b_{lj}, $a_{\lambda 0}$, $b_{\lambda 0}$ (Table 2).

The factor variances were set at 1.0 for model identification, and the MCMC algorithms proposed by Liu (2008) and Liu and Daniels (2006) was used to estimate the factor correlation matrix. Other parameters were estimated using the Gibbs sampler (Geman & Geman, 1984). We conducted several test runs to select the

number of MCMC iterations and found that the algorithm converged in less than 8,000 iterations as judged by the value of the estimated potential scale reduction (EPSR) values (<1.2; Gelman, 1996). Based on the result from the test runs, all final estimates were obtained using 10,000 draws from the posterior distributions after a burn-in phase of 10,000 iterations. Data generation and estimation used R 3.5.3 (R Development Core Team, 2020).

Parameter Identification

The following criteria for parameter identification were used in the current study:

1) Thresholds of magnitude 0, 0.05, 0.1, and 0.15 with the decision rule to include if the absolute value of the parameter estimate is larger than the cutoff. The thresholding rule applies to the estimates of residual correlations and standardized cross-loadings.

2) A p-value with $\alpha = 0.05$, with the decision rule to include if $p < 0.05$. The p-value was based on MCMC samples rather than the z-test.[1] As defined by Muthén (2010), the p-value is the proportion of the negative/positive posterior samples for or a positive/negative estimate, respectively.

[1] We also checked the performance of the z-test, and results were similar to but slightly worse than the 95% HPD interval. Results based on the z-test are not reported in the current paper. Note that the worse performance of z-test may be due to the non-normal property of Lasso regularized parameters (Pötscher & Leeb, 2009).

3) A 95% HPD interval, with the decision rule to include if the point 0.0 is outside the 95% HPD interval.

Note that for the thresholding criterion, the values of 0.0 and 0.1 were commonly used in previous frequentist and Bayesian studies. Here we also included two additional cutoff values of 0.05 and 0.15 to fully investigate the effect of different thresholds.

Evaluation criteria

Three metrics were considered for a performance comparison of the decision rules:

1) Power: the probability of correctly identifying the cross-loadings/residual correlations when the parameters are non-zero (Muthén & Asparouhov, 2012). Acceptable power is above 80% (Muthén & Muthén, 2002).
2) Type-I error rates: the probability of erroneously identifying the cross-loadings/residual correlations when the parameters are zero. The acceptable Type I error rate is within the 95% confidence interval of a binomial distribution [0.007, 0.093] ($[0.05 \pm 1.96 \times \sqrt{0.05 \times (1-0.05)/number\ of\ replications}$; Cham et al., 2012).
3) The ratio of correct identification to the total number of identified parameters (Yuan and Liu, 2020): There is no established criterion for this metric.

We focus on these three metrics in the current paper, the estimation results of Bayesian Lasso CFA were reported in the supplementary materials. The results of

these metrics were similar in two sets of hyperparameters, the results of the first set of hyperparameters were reported in the paper and the others were also shown in the supplementary materials.

Results

Power

Figure 3 shows the performance of the three criteria for decision rules on power. Compared to the other two rules, the HPD interval rule performed worst in detecting non-zero cross-loadings when the sample size was small or when the target was within-factor residual correlations. For example, even when the sample size was 1,000 and the residual correlations were 0.7, the power of the 95% HPD interval rule in detecting within-factor residual correlations in the two-factor model was close to zero. The performance of the *p*-value decision rule was better than HPD interval rule in detecting cross-loadings and between-factor residual correlations, but was worse than the thresholding rule with threshold values of 0.05 and 0.1. However, the *p*-value rule had similar power problems as the HPD interval rule. For example, in the two-factor model, the *p*-value rule did not have sufficient power even when the sample size was 1,000 and the residual correlation was 0.3.

As expected, the zero-threshold produced 100% power in all the conditions. Under most conditions, power was maintained at approximately 80% by increasing the threshold to 0.05. The cutoff value of 0.1 only produced acceptable power when

the estimated value was larger than 0.1. We also found that the power was higher for residual correlations than for cross-loadings with the same threshold and the same level of actual parameter values (0.1, 0.3). The threshold of 0.15 was acceptable in power for residual correlations but not for cross-loadings.

Additionally, we found an increasing sample size significantly improved the performance of the HPD interval and the *p*-value decision rules. However, the increasing sample size only led to a small improvement in the performance of the thresholding rules in general.

Insert Figure 3 about here

Type I Error Rate

Figure 4 displays the performance of the three criteria for decision rules on Type I error rate. The general pattern was almost a mirror image of that of power. For example, the HPD interval rule produced the lowest Type I error rate and outperformed the thresholding rule in most conditions. The *p*-value rule also maintained the Type I error rate well in all the conditions. For the thresholding rule, Type I error rates were higher for residual correlations than for cross-loadings. As expected, the Type I error rate decreased with increasing cutoff value. The 0.05 cutoff produced Type I error rates higher than 0.093 for cross-loadings in some conditions and for residual correlations in almost all conditions. The cutoff of 0.1 maintained

Type I error rates well at the nominal level for cross-loadings. However, for residual correlations conditions, the 0.1 cutoff produced high Type I error rates (> 0.3 in some conditions). Using the 0.15 cutoff value instead significantly improved performance.

Although the abovementioned results suggest using the thresholding rule at the 0.1 and 0.15 cutoffs respectively for cross-loading and residual correlation for maintaining the Type I error, using a lower cutoff may be appropriate for other purposes, such as exploring factor and residual covariance structures (Jacobucci, Brandmaier, & Kievit, 2019; Serang et al., 2017). In the case in which power is emphasized, the two respective thresholds (0.05 for cross-loading and 0.1 for residual correlation) are favored. Results showed that for parameters with a true value of 0.1 or higher, these thresholds provided high power for detection than the 0.1-threshold for loadings and the 0.15-threshold for residual correlations (Fig. 3).

Insert Figure 4 about here

The Ratio of Correct Identification

Figure 5 summarizes the results for the metric in the correct identification ratio. The ratio increased with an increasing sample and decreasing Type I error rate (Fig. 5). The metric was higher for cross-loadings compared to residual correlations, which was partly caused by the relatively large proportion of zero residual correlations in the generative models.

Results showed the more conservative the rule, the higher the ratio of correct identification. The most conservative rule—a 95% HPD interval—performed best in terms of this metric in most conditions. The *p*-value rule also performed better than the thresholding rule for detecting within-factor residual correlations. Cutoff at zero performed worst, followed by the 0.05 cutoff. For correctly identifying cross-loading, cutoff values of 0.1 and 0.15 performed the best, and their performances were similar. The 0.15 cutoff outperformed the 0.1 cutoff in detecting residual correlations.

Insert Figure 5 about here

Discussion

The current study offers a comprehensive comparison between the thresholding rule, the *p*-value rule, and the HPD interval rule in parameter identification within the context of the Bayesian lasso CFA. In general, we found no universally superior criterion or decision rule. For example, criteria based on thresholding rules with cutoff value > 0 showed a more robust performance than the HPD interval in balancing power and Type I error rate. However, the thresholding rules did not perform well in terms of percentage of correct identification. As different criteria suit different purposes (e.g., for exploratory purposes power is more important, whereas for formal hypothesis testing, Type I error rate is more important), we cannot recommend universal simple rules-of-thumb for parameter identification. Instead, we summarize

our recommendations for which criterion to consider in a table that separates out major circumstances—metric to consider, sample size, and model parameter type (Table 3).

In general, power and Type I error rates for rules using thresholding criteria were higher for residual correlations than for cross-loadings. For cross-loading, the threshold of 0.1, which has been commonly used by previous studies (Guo et al., 2012; Hoti and Sillanpää, 2006; Feng et al., 2015; Peterson et al., 2014) provides a good balance—i.e., sufficient power and acceptable Type I error rates. The cutoff of 0.1, however, produced unacceptably high Type I error rates for residual correlations; as such, the cutoff value of 0.15 is instead recommended.

The findings in this paper are based on the context of relaxed CFA assumptions and the Bayesian lasso tool for estimating the relaxed model. In the broader context of SEM, relaxing similar constraints using regularized methods has become an emerging and important trend (Jacobucci & Grimm, 2018; Lu et al., 2016; Muthén & Asparouhov, 2012). The 0.1 cutoff value, which we recommend for detecting cross-loading, can be a candidate for application to parameter identification for path coefficients in SEM, given the similar roles of path coefficient and factor loading as regression parameters. The simple 0.1 cutoff rule, pending further investigation, could be especially useful in helping researchers to specify complex structural models in which causal relationships are difficult to fully delineate from theoretical consideration. On the other hand, some SEM applications require conditional

correlations to be properly modeled for improving model fit. For example, in a SEM with two time points, the same indicator variables of the same latent factor across the time points may be modeled as correlated. From the results in this paper, we conjecture that using a cutoff of 0.15 for such correlation would be appropriate if the Bayesian lasso model is adopted. As a caveat, the current study did find that the 0.1 and 0.15 cutoffs (respectively for cross-loading and residual correlation) were unable to provide sufficient power in some cases when the true value of the parameter is 0.1, although power is acceptable when the true value of the parameter is > 0.1. In situations in which path coefficients (standardized) and correlations less than 0.1 are not deemed clinically meaningful, the respective cutoffs at 0.1 and 0.15 represent reasonable choices. Our recommendation can serve as a preliminary guideline for future investigation into the use of simple thresholding rules for broader use in SEM.

For exploratory analysis in SEM in which the purpose is to extract as many potentially important relationships as possible, the cutoff value can be lowered; we suggest using 0.05 for path coefficients and 0.1 for correlations. Further study is needed to fully assess the performance of these rules for exploratory SEM.

The current study has limitations. The number of factors investigated in the simulation study is relatively small, and the method is limited to the Bayesian lasso. Another future direction would be to investigate the criteria performance and decision rules in a broader setting, including CFA with a larger number of factors, as well as other extensions of the lasso method such as the adaptive lasso (Zou, 2006).

In summary, the current paper provides guidelines on the use of different decision rules for parameter identification in a CFA when the Bayesian lasso method is applied. The guidelines are based on a comparison of several commonly used criteria. As pointed out by Hindman (2015), regularized approaches have proven to outperform traditional methods, and methods such as penalized regression (e.g., the Bayesian lasso) should be more widely applied for theory building and hypothesis testing in social sciences. The current study represents an important contribution that facilitates such application.

References

Brandt, H., Cambria, J., & Kelava, A. (2018). An Adaptive Bayesian Lasso Approach with Spike- and-Slab Priors to Identify Multiple Linear and Nonlinear Effects in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(6), 946-960.

Brandt, H., Umbach, N., Kelava, A., & Bollen, K. A. (2020). Comparing estimators for latent interaction models under structural and distributional misspecifications. *Psychological Methods, 25*(3), 321-345.

Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research, 29*, 289–311.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences, 136*, 68–78.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed. pp 79-80)*. NJ: Erlbaum: Hillsdale.

Cham, H., West, S. G., Ma, Y., and Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research, 47*(6), 840-876.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212.

Feng, X., Wu, H., Song, X. (2015) Bayesian Adaptive Lasso for Ordinal Regression with Latent Variables. *Sociological Methods & Research 46*(4): 1-28

Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelharter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 131–144). Chapman & Hall.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice.* London, UK: Chapman & Hall.

Guo, R., H. Zhu, S. M. Chow, and J. G. Ibrahim. (2012). Bayesian Lasso for Semi-parametric Structural Equation Models. *Biometrics 68*:567-77.

Hans, C. (2009). Bayesian Lasso regression. *Biometrika, 96*(4), 835–845.

Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and l1 penalized regression: A review. *Statistics Surveys, 2*, 61–93.

Hindman M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *The Annals of the American Academy of Political and Social Science*. 659, 48-62.

Hoti, F. and M. J. Sillanpää. (2006). Bayesian Mapping of Genotype × Expression Interactions in Quantitative and Qualitative Traits. *Heredity 97*:4-18.

Jacobucci, R., & Grimm, K. J. (2018). Comparison of frequentist and Bayesian regularization in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*, 639–649.

Jacobucci, R., Brandmaier, A., & Kievit, R. (2019). A practical guide to variable selection in structural equation models with regularized MIMIC models. *Advances in Methods and Practices in Psychological Science, 2*, 55-76.

Jacobucci, R., Grimm, K., Brandmaier, A., Serang, S., Kievit, R., Scharf, F., Li, X. (2020). *regsem: regularized structural equation models. R package version 1.5.2.* Retrieved from https://cran.r-project.org/web/packages/regsem/index.html

Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis, 5*(2): 369–411.

Khong, K. W., Onyemeh, N. C. , & Chong, Y. L. . (2013). Bsem estimation of network effect and customer orientation empowerment on trust in social media and network environment. *Expert Systems with Applications, 40*(12): 4858-4870.

Khondker, Z. S., Zhu, H., Chu, H., Lin, W., & Ibrahim, J. G. (2013). The Bayesian covariance lasso. *Statistics and Its Interface, 6*, 243–259.

Liu, X. (2008). Parameter expansion for sampling a correlation matrix: An efficient GPX-RPMH algorithm. *Journal of Statistical Computation and Simulation, 78*, 1065–1076.

Liu, X., & Daniels, M. J. (2006). A new efficient algorithm for sampling a correlation matrix based on parameter expansion and re-parameterization. *Journal of Computational and Graphical Statistics, 15*, 897–914.

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: alternative priors and consequences. *Multivariate Behavioral Research, 51*(4): 519–539.

Lindstrøm, J.C., & Dahl, F.A. (2020). Model Selection with Lasso in Multi-group Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1):33-42.

Liang, X., & Jacobucci, R. (2019). Regularized Structural Equation Modeling to Detect Measurement Bias: Evaluation of Lasso, Adaptive Lasso, and Elastic Net. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(5): 722-734.

Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso. *The Annals of Statistics, 42*, 413–468.

Marcus, D. K., Preszler, J., & Zeigler-Hill, V. (2017). A network of dark personality traits: What lies at the heart of darkness? *Journal of Research in Personality, 73*, 56–62.

Maslach, C., & Jackson, S. E. (1981). *Maslach Burnout Inventory manual.* Palo Alto, CA: Consulting Psychologists Press.

Maslach, C., & Jackson, S. E. (1986). *Maslach Burnout Inventory manual* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

McNeish, D. M. (2015). Using Lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research, 50*(5), 471–484.

Muthén, B. O. (2010). *Bayesian analysis in Mplus: A brief introduction.* Los Angeles, CA: Muthén & Muthén. https://www.statmodel.com/download/IntroBayesVersion%203.pdf.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620.

Park, T., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association, 103*(482), 681–686.

Pan, J.H., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods, 22*(4), 687–704.

Peterson, C., Vannucci, M., Karakas, C. Choi, W., Ma, L., Maletic-savatic, M. (2014). Inferring metabolic networks using the Bayesian adaptive graphical lasso with informative priors. *Statistical and Its Interface, 6*(4): 547-558.

Pötscher, B.M., & Leeb, H. (2009). On the distribution of penalized maximum likelihood estimators: the lasso, scad, and thresholding. *Journal of Multivariate Analysis, 100*(9), 2065–2082.

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Serang, S., & Jacobucci, R. (2020). Exploratory Mediation Analysis of Dichotomous Outcomes via Regularization. *Multivariate Behavioral Research, 55*(1), 69–86.

Serang, S., Jacobucci, R., Brimhall, K. C., & Grimm, K. J. (2017). Exploratory mediation analysis via regularization. *Structural Equation Modeling: A Multidisciplinary Journal, 24*. 733–744.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, 58*(1), 267—288.

Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis, 7*, 867–886.

Yuan, K-H., & Liu, F. (2020). Which method is more reliable in performing model modification: lasso regularization or lagrange multiplier test? *Structural Equation Modeling: A Multidisciplinary Journal.* Advance online publication.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science A Journal of the Association for Psychological Science, 12*(6), 1100–1122.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of American Statistical Association, 101*, 1418–1429.

Zhang, L., Pan, J., Dube, L., Ip, E.H. (accepted). blcfa: An R Package for Bayesian Model Modification in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal.*

Table 1. Design of Simulation Study

Factors	Conditions
Sample Size	200, 500, 1000
Model Size	2 factors and 10 items, 3 factors and 18 items
Effect Size of Parameters	0, 0.1, 0.2, 0.3 for cross-loadings
	0, 0.1, 0.3, 0.7 for residual correlations

Table 2. Two Sets of Values for Hyperparameters

Set	$\boldsymbol{\mu}_0$	$\mathbf{H}_{\mu 0}$	$\mathbf{R}_0$	ρ_0	$\boldsymbol{\Lambda}_{0j}$	$\mathbf{H}_{0j}$	a_{0j}	b_{0j}	a_{lj}	b_{lj}	$a_{\lambda 0}$	$b_{\lambda 0}$
1	0	4**I**	**I** + 0.1	K+2	0	4**I**	1	0.01	1	0.01	1	0.01
2	0	4**I**	**I** + 0.5	K+8	0	**I**	1	0.1	1	0.1	1	0.1

Note: **I**: identity matrix, **I** + 0.1: diagonal elements = 1 and non-diagonal elements = 0.1, **I** + 0.5: diagonal elements = 1 and non-diagonal elements = 0.5, K: number of factors.

Table 3. Summary of Recommendations

Sample Size	Parameters	Threshold 0.1		Threshold 0.15		HPD Interval		*p*-value	
		Power	Type I	Power	Type I	Power	Type I	Power	Type I
200	Cross-loadings[1]	×	√	×	√	×	√	×	√
	Residual correlations	√	×	√	√	×	√	×	√
500	Cross-loadings	√	√	×	√	√	√	√	√
	Residual correlations	√	√	√	√	×	√	×	√
1000	Cross-loadings	√	√	√	√	√	√	√	√
	Residual correlations	√	√	√	√	×	√	×	√

Note: Type I: Type I Error Rate; √: acceptable in most conditions, ×: unacceptable in many conditions, shaded: the best criterion in the corresponding condition.

[1]None of the criteria can provide sufficient power under this condition.

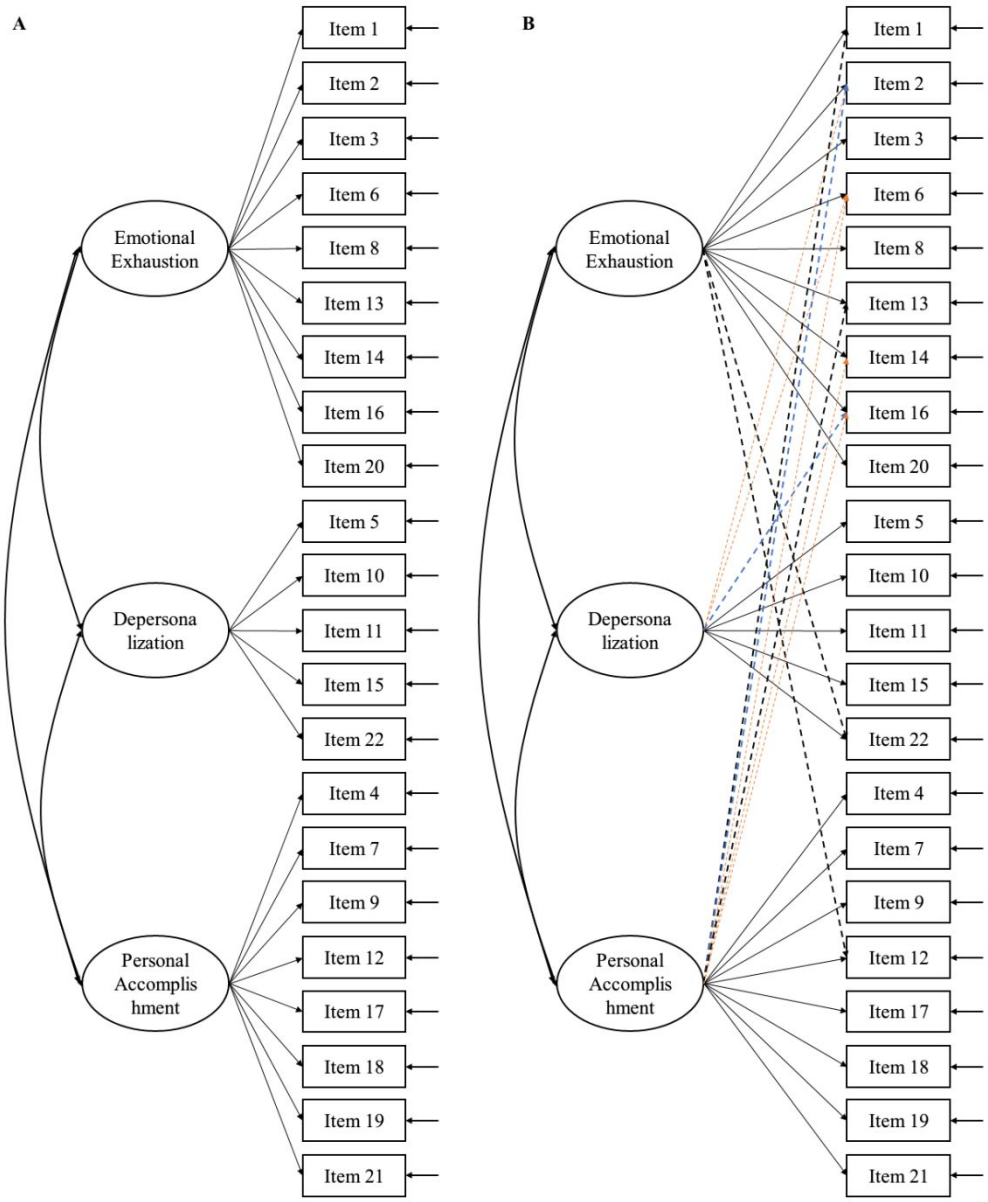

Figure 1. Measurement model of the Maslach Burnout Inventory. Fig. 1A: Three dimensions of the Maslach Burnout Inventory. Fig. 1B: Results of the Bayesian Lasso CFA. Black solid line: main loadings, black dotted line: cross-loadings identified by all three criteria, blue dotted line: cross-loadings identified by threshold and *p*-value, orange dotted line: cross-loadings identified only by threshold.

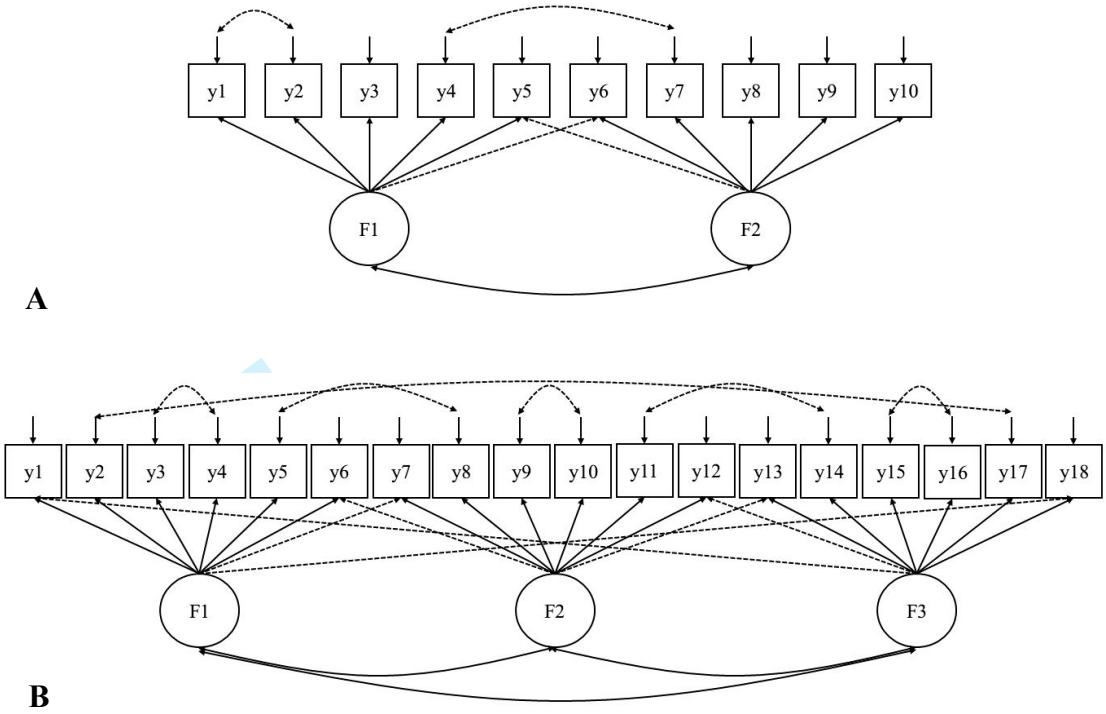

Figure 2. Measurement models in the simulation study. Fig. 1A: measurement model with 2 factors and 10 items, Fig. 1B: measurement model with 3 factors and 18 items. Dotted line with single arrow: cross-loadings, dotted line with double arrow: residual co-variances, solid line with single arrow: main loadings and residual variances, solid line with double arrow: factor co-variances.

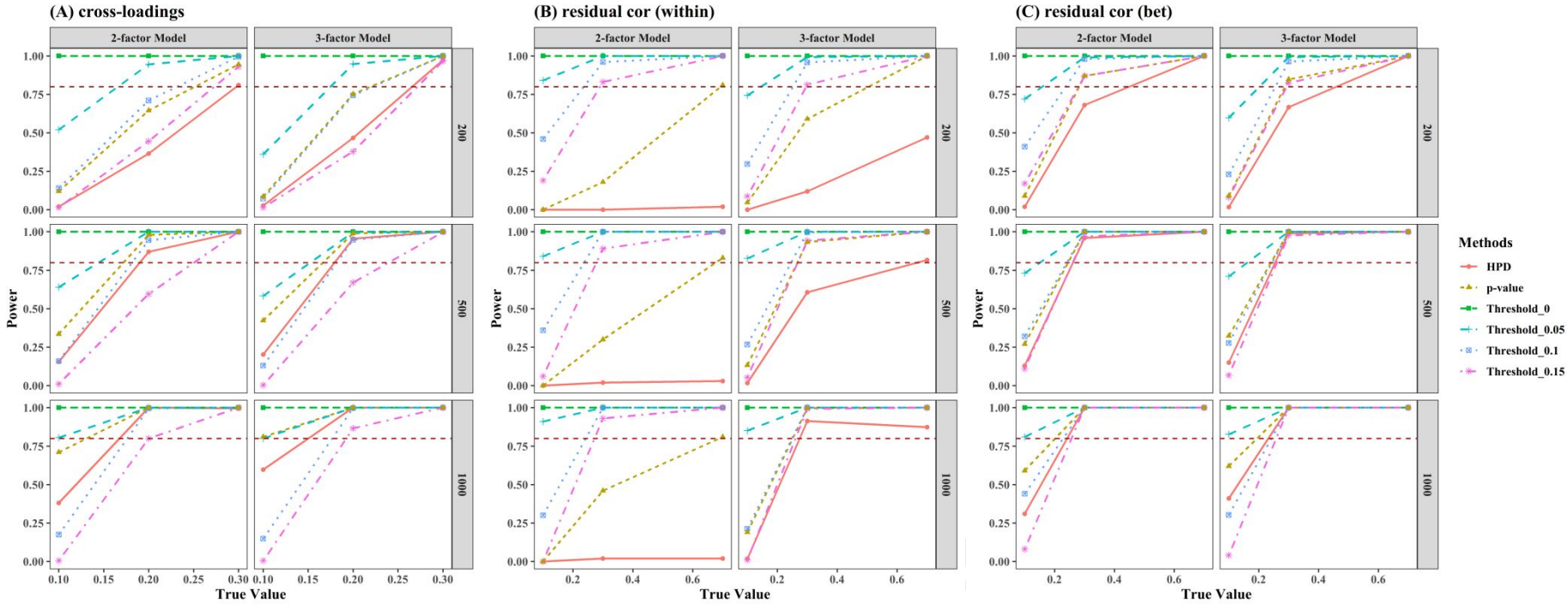

Figure 3. Power.

Note: True value: 0.1, 0.2 and 0.3 for cross-loadings and 0.1, 0.3, 0.7 for residual correlations. Methods: methods of variable selection, thresholds include 0, 0.05, 0.1, and 0.15, HPD: 95% highest posterior density interval. Residual cor (with): within-factor residual correlations. residual cor (bet): between-factor residual correlations. The horizontal red dotted line indicates the threshold of acceptable power (0.8).

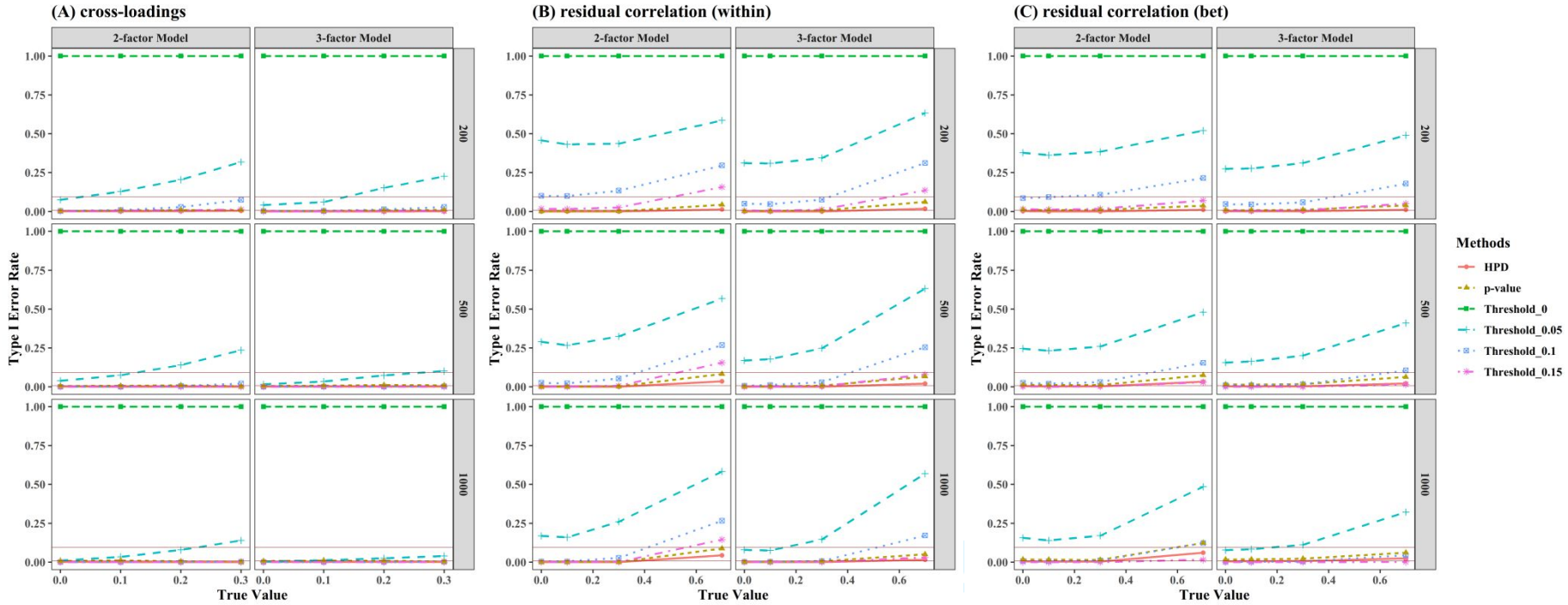

Figure 4. Type I error rate.

Note: True value: 0, 0.1, 0.2 and 0.3 for cross-loadings and 0, 0.1, 0.3, 0.7 for residual correlations. Methods: methods of variable selection, thresholds include 0, 0.05, 0.1, and 0.15, HPD: 95% highest posterior density interval. The two horizontal red lines indicate the bounds of acceptable Type I error rate (0.007-0.093).

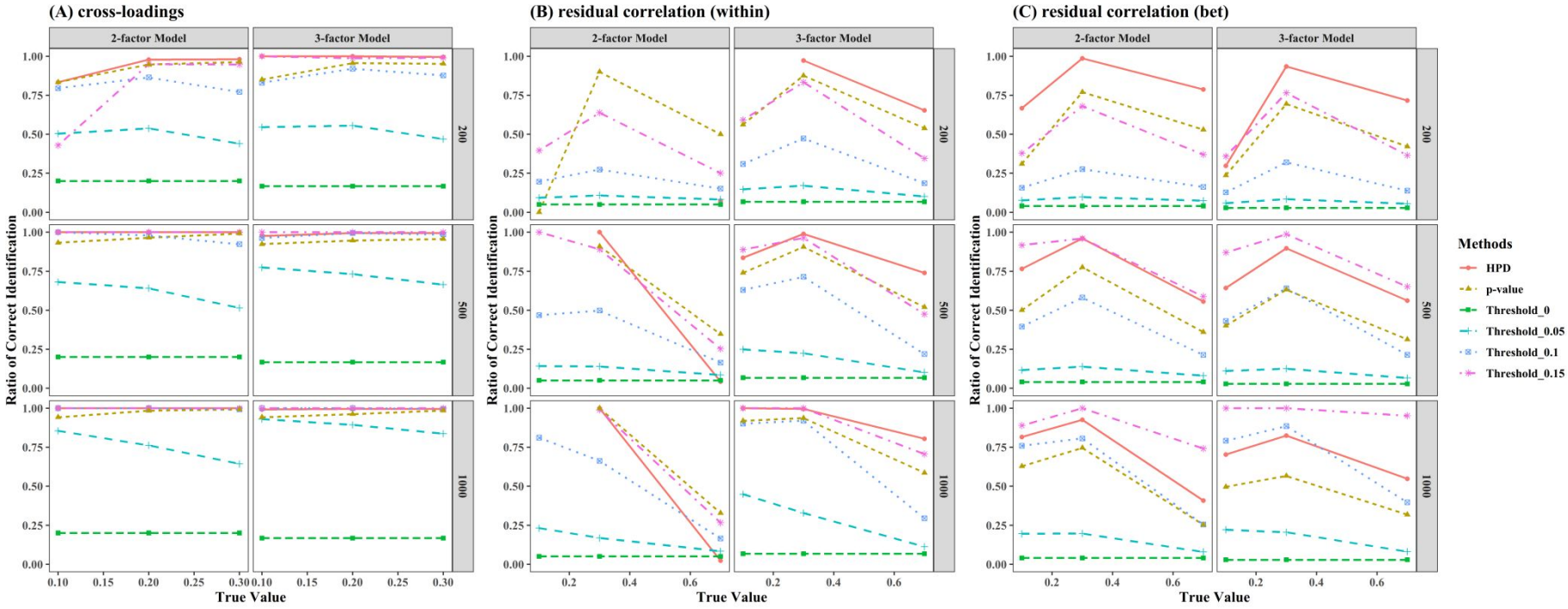

Figure 5. Ratio of correct Identification.

Note: True value: 0.1, 0.2 and 0.3 for cross-loadings and 0.1, 0.3, 0.7 for residual correlations. Methods: methods of variable selection, thresholds include 0, 0.05, 0.1, and 0.15, HPD: 95% highest posterior density interval. If a line is missing, it indicates that no cross-loading/residual correlation was identified by the method in that condition.