Comparison between Different Parameters Identification Criteria using the Bayesian Lasso

Junhao Pan
Sun Yat-sen University, China, panjunh@mail.sysu.edu.cn

Joint work with Lijin Zhang
Sun Yat-sen University, China
Edward Haksing Ip
Wake Forest University School of Medicine, United States

19 July 2021

Overview

1. Bayesian Lasso
   - Bayesian Lasso Confirmatory Factor Analysis
2. Different Ways of Parameter Identification with the Lasso Method
   - The Thresholding Rule
   - The *p*-value and Interval Rule
3. Simulation Study
4. Discussion

Bayesian Lasso

Regularization

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.

Regularization has the potential to create a better balance between model simplicity and model fit in such conditions.

The lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) method has been used in many fields and increasingly in social sciences (Lindstrøm & Dahl, 2020).

Lasso in Regression Models

Consider a regression model with J predictors $\mathbf{X}$ to outcome data $\mathbf{Y}$ by minimizing the estimation function

$$L^{lasso}(\beta) = (|\mathbf{Y} - \mathbf{X}\beta|)^{\mathbf{2}} + \lambda \sum_{\mathbf{j=1}}^{\mathbf{J}} |\beta_{\mathbf{j}}| \tag{1}$$

- β: $J \times 1$ vector of regression coefficients for the J predictors.
- $L^{lasso}(\beta)$ and $(|\mathbf{Y} - \mathbf{X}\beta|)^{\mathbf{2}}$: loss function of lasso regression and least square difference, respectively.
- $\lambda \sum_{j=1}^{J} |\beta_j|$: lasso penalty function with the tuning parameter $\lambda \geq 0$ that 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.

Bayesian Lasso

The Bayesian lasso has been increasingly used in the social sciences for several reasons:

- lasso can be readily applied in Bayesian analyses by using the double exponential priors (Park & Casella, 2008).
- the Bayesian lasso can provide estimates of standard errors and intervals that are difficult to obtain under a frequentist framework (Kyung et al., 2010).
- the tuning parameters can be more conveniently estimated with other coefficients simultaneously under the Bayesian lasso paradigm (Hans, 2009; Park & Casella, 2008).

Applications of Lasso and Bayesian Lasso

Recently, novel methods that extend the lasso have been developed for structural equation modeling (SEM) and network analysis.

- graphical lasso network models (Costantini et al., 2019);
- network analysis with an adaptive lasso method (Marcus, Preszler, & Zeigler-Hill, 2017);
- Bayesian lasso confirmatory factor analysis (CFA; Chen et al., 2020; Pan, Ip, & Dubé, 2017);
- exploratory mediation analysis (Serang et al., 2017);
- Bayesian adaptive lasso for ordinal regression with latent variables (Feng, Wu, & Song, 2017);
- regularized multiple-indicators and multiple-caused (MIMIC) models (Jacobucci, Brandmaier, & Kievit, 2019).

Dilemma in Practice

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.

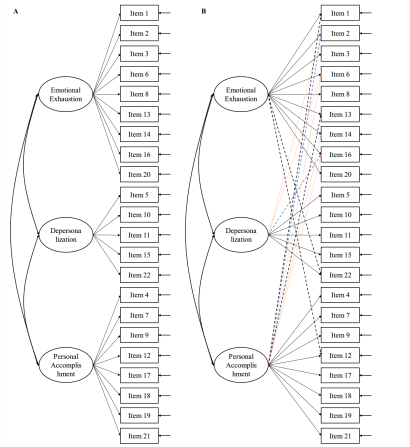
At least three criteria for parameter identification have been used in the literature.

- the thresholding rule;
- the p-value rule;
- the credible interval rule.

In this paper, we compare the three criteria for identifying parameter for inclusion into a CFA model using the Bayesian lasso.

Dilemma in Practice

To illustrate, 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, 7-point Likert scale ; Maslach & Jackson, 1981).



Research Goals

Apply the Bayesian lasso to regularize both cross-loadings or residual correlations (both within- and across-factor), to evaluate the performance of the three criteria under different conditions.

Provide further recommendations on the decision rules for using the Bayesian lasso CFA.

Confirmatory Factor Analysis

Suppose $y_1, y_2, ..., y_n$ are independent random observations, and each $y_i = (y_{i1}, y_{i2}, ..., y_{ip})^T$ satisfies the following factor analysis model:

$$y_i = \mu + \Lambda\omega_i + \epsilon_i, \quad i = 1, 2, ..., n, \tag{2}$$

- μ: intercepts.
- Λ: factor loading matrix reflecting the relationship between observed variables in y_i with the latent variables in ω_i.
- ϵ_i: measurement errors (residuals), $\sim N[0, \Psi]$, independent of ω_i.

Bayesian Lasso CFA

The theory being tested simply does not fit the data well due to the assumptions:

- no cross-loading;
- zero residual covariances;

Under the Bayesian framework, the strict assumptions are relaxed through the assignment of priors to the corresponding parameters.

Bayesian lasso:

- assigns double exponential priors for cross-loadings and residual covariances matrix.
- simultaneously identifies non-negligible cross-loadings and residual covariances in a joint estimation procedure.

Bayesian Lasso CFA

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 without regularization (Muthén & Asparouhov, 2013; Serang et al., 2017; Serang & Jacobucci, 2020).

The identification of “significantly large” non-zero entries require careful operationalization.

Different Ways of Parameter Identification with the Lasso Method

The Thresholding Rule

Frequentist Lasso:

- 0, 0.001 (Liang & Jacobucci, 2020; Serang et al., 2017; Serang & Jacobucci, 2021; Yuan & Liu, 2020).

Bayesian Lasso:

- 0.1 ($|\beta| \geq 0.1$, Guo et al., 2012; Hoti & Sillanpää, 2006; Feng, Wu, & Song, 2017).

The cutoff value can also be justified from a substantive standpoint.

- cross-loading of less than 0.1 can be considered to have little practical importance (Muthén & Asparouhov, 2012).
- correlation coefficient as 0.1 is a typical value of low-effect size (Cohen, 1988).

The p-value and Interval Rule

Frequentist Lasso:

- In regression models, R-package covTest (Lockhart et al., 2014) can be used to obtain p-values.
- In network analysis, Epskamp et al. (2018) demonstrated how frequentist-lasso regularization can provide confidence intervals using a bootstrap method.

Bayesian Lasso:

- it is relatively straightforward to obtain p-values.
- credible intervals such as the HPD interval can also be calculated using Markov Chain Monte Carlo (MCMC; Gilks, Richardson, & Spiegelhalter, 1996) algorithms.

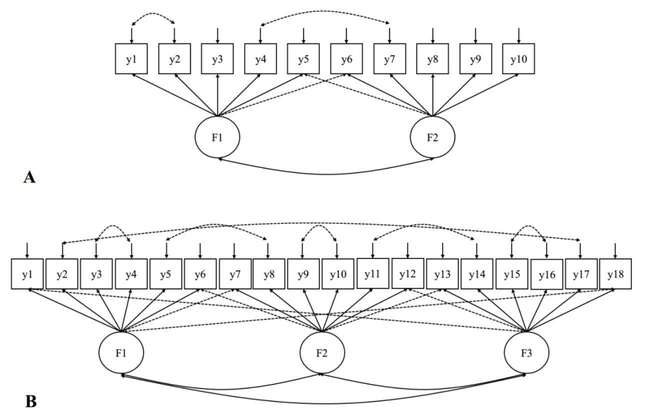
Simulation Study

Simulation Study

To investigate the performance of three criteria, as well as their associated decision rules for detecting significant cross-loading/residual correlation under a range of experimental conditions.

- Sample size: 200, 500, 1000.
- Model Size: 2 factors and 10 items, 3 factors and 18 items.
- Effect Size: 0, 0.1, 0.2, 0.3 for cross-loadings; 0, 0.1, 0.3, 0.7 for residual correlations.

Simulation Study



Model Estimation

To avoid the possible confounding effect, the conditions of non-zero cross-loadings and non-zero residual correlations were separately analyzed.

- M_1: model with some non-zero cross-loadings and diagonal residual covariance matrix.
- M_2: model with some non-zero off-diagonal residual covariance entries but no cross-loading.

Priors for the intercepts and factor covariance matrix:

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

For M_1, the loadings for the j-th item $\Lambda_j = \binom{\Lambda_j^m}{\Lambda_j^c}$ where Λ_j^m and Λ_j^c respectively represent main loadings and cross-loadings

$$\Lambda_j^m \sim N(\Lambda_{0j}, \mathbf{H}_{0j}), \Lambda_j^c|\psi_{jj} \sim N(0, \psi_{jj}\mathbf{D}_{\tau_j}) \tag{4}$$

$$\psi_{jj}^{-1} \sim Gamma(a_{0j}, b_{0j}), \mathbf{D}_{\tau_j} = diag(\tau_{j1}^2, ..., \tau_{jK}^2) \tag{5}$$

$$\tau_{jk}^2 \sim Gamma(1, \frac{\gamma^2}{2}), \gamma^2 \sim Gamma(a_{lj}, b_{lj}) \tag{6}$$

Model Estimation

For M_2,

$$\Lambda_j \sim N(\Lambda_{0j}, \mathbf{H}_{0j}), \tag{7}$$

Following Khondker et al. (2013) and Wang (2012), graphical lasso priors are specified for the inverse of the residual variance-covariance matrix, $\Psi^{-1} = \Sigma = (\sigma_{jj'})$:

$$\sigma_{jj} \sim \frac{\lambda}{2} \exp(-\frac{\lambda}{2} \sigma_{jj}), \tag{8}$$

$$\sigma_{jj'} \sim \frac{\lambda}{2} \exp(-\lambda |\sigma_{jj'}|), \quad j < j' \tag{9}$$

$$\lambda \sim Gamma(a_{\lambda 0}, b_{\lambda 0}). \tag{10}$$

Sensitivity Analysis

We adopted three sets of values for hyperparameters, results were similar.

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	$4\mathbf{I}$	$\mathbf{I}+0.1$	K+2	0	$4\mathbf{I}$	1	0.01	1	0.01	1	0.01
2	0	$4\mathbf{I}$	$\mathbf{I}+0.5$	K+8	0	$\mathbf{I}$	1	0.1	1	0.1	1	0.1
3	0	$100\mathbf{I}$	$\mathbf{I}$	K+2	0	$100\mathbf{I}$	1	0.01	1	0.01	1	0.01

Parameter Identification

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

- Thresholds of magnitude 0, 0.05, 0.1, and 0.15, with the decision rule to include if the absolute value of the standard estimate is larger than the cutoff.
- A *p*-value with $\alpha = 0.05$, with the decision rule to include if $p < 0.05$. The *p*-value can be different from the frequentist *p*-value, it is one-tailed and is based on MCMC samples rather than the z-test.
- A 95% HPD interval, with the decision rule to include if the point 0.0 is outside the 95% HPD interval.

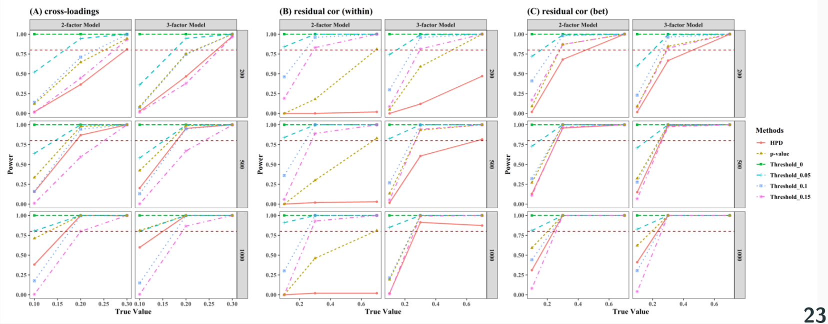
Evaluation Criteria

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

- Power: the probability of correctly identifying the cross-loadings/residual correlations when the parameters are non-zero (Muthén & Asparouhov, 2012).
- Type-I error rates: the probability of erroneously identifying the cross-loadings/residual correlations when the parameters are zero.
- The ratio of correct identification to the total number of identified parameters (Yuan & Liu, 2020).

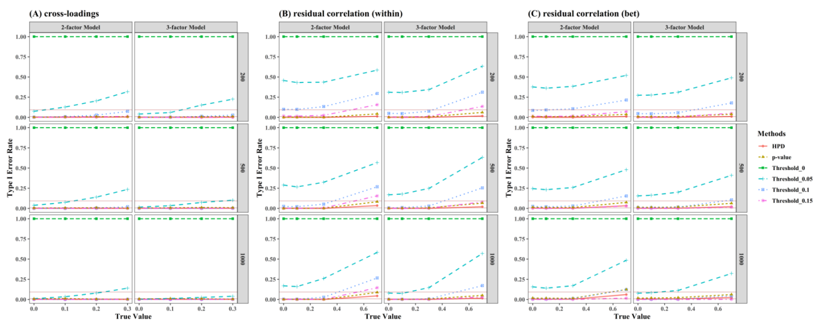
Results: Power

- 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.
- The p-value rule had similar power problems as the HPD interval rule.
- The thresholding rule is more robust to sample sizes.



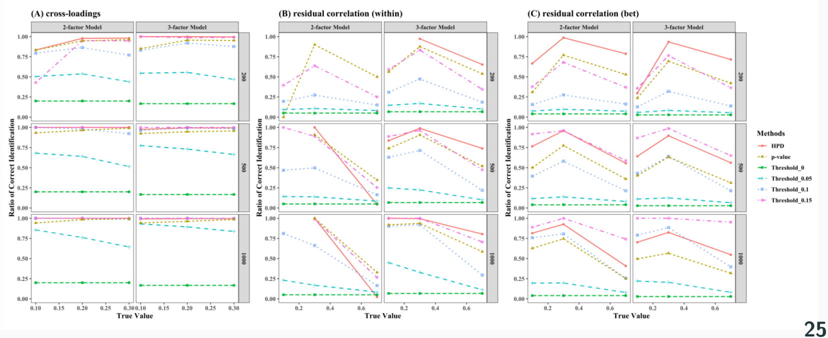
Results: Type I Error Rate

The general pattern was almost a mirror image of that of power.



Results: The Ratio of Correct Identification

- The more conservative the rule, the higher the ratio of correct identification.
- 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.



Detect Cross-loadings & Residual Covariances Simultaneously

- The co-existence of cross-loadings and residual correlations may be expected in practice.
- The phenomenon of low power of p-value and HPD interval methods was still present for within-factor residual correlations.
- Compared to performance when only one kind of parameter was present, we found the HPD interval and p-value methods were more sensitive to model size and provided lower power.

Discussion

Recommendation

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.

Usage of The Thresholding Rule

- 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.
- We conjecture that using a cutoff of 0.15 for other correlation parameters would be appropriate if the Bayesian lasso model is adopted.
- 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.

Limitations and Future Research

- Limited model sizes.
- Extension of the lasso method (e.g., adaptive lasso).
- Comparison with one-step approach such as the spike-and-slab method.

Thanks for listening!