Comparison between Bayesian and Frequentist Regularization in Confirmatory Factor Analysis

Lijin Zhang, Graduate School of Education, Stanford University

Xinya Liang, Department of Counseling, Leadership, and Research Methods, University of Arkansas

Junhao Pan, Department of Psychology, Sun Yat-sen University

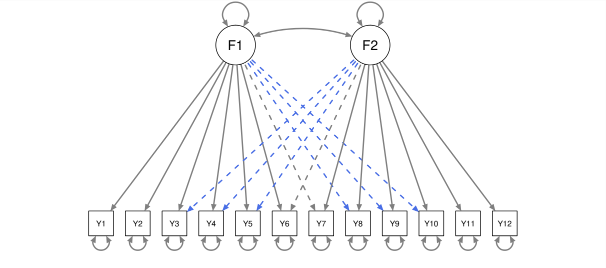
27 July 2023

Introduction

Cross-loadings in Confirmatory Factor Analysis

A general confirmatory factor analysis (CFA) model is specified as:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F}_i + \boldsymbol{\epsilon}_i, i = 1, 2, ..., N, \tag{1}$$



Dashed line: Cross-loading; Solid line: Main loading

Regularization

Ridge, Lasso, and Adaptive Lasso (alasso)

- Frequentist (Li & Jacobucci, 2021; Yuan & Liu, 2021)
- Bayesian (Chen et al., 2021; Lu et al., 2016; Muthén & Asparouhov, 2012)

Comparison

- Equivalence.
- Uncertainty quantification.
- Estimation of penalty parameters.
- Model complexity and small samples.

Frequentist Regularization for Selecting Cross-loadings

For a model with J cross-loadings $\boldsymbol{\lambda}$ to be regularized, Ridge, Lasso and Alasso can be formulated as follows:

$$
\begin{aligned}
F_{ridge} &= F_{ML} + \gamma \sqrt{\sum_{j=1}^{J} \lambda_j^2} \\
F_{lasso} &= F_{ML} + \gamma \sum_{j=1}^{J} |\lambda_j| \\
F_{alasso} &= F_{ML} + \gamma_j \sum_{j=1}^{J} |\lambda_j|
\end{aligned} \tag{2}
$$

- A greater value of penalty parameters ($\gamma, \gamma_j \geq 0$) leads to increased penalization.
- Models were selected using BIC or cross-validation.

Bayesian Regularization: Penalty Priors

For the cross-loadings:

- Ridge:

$$\lambda_j \sim N(0, \sigma^2), \sigma^2 \sim Uniform(0, 1) \tag{3}$$

- Lasso:

$$\lambda_j \sim N(0, \psi_{jj}\tau_j^2), \psi_{jj}^{-1} \sim Gamma(\alpha_j, \beta_j) \tag{4}$$

$$\tau_j^2 \sim Gamma(1, \frac{\gamma^2}{2}), \gamma^2 \sim Gamma(a_l, b_l) \tag{5}$$

  Here, $\alpha_j = \beta_j = 0.01$, $a_l = 1$, and $b_l = 0.01$.

- Adaptive Lasso:

$$\tau_j^2 \sim Gamma(1, \frac{\gamma_j^2}{2}), \gamma_j^2 \sim Gamma(a_l, b_l) \tag{6}$$

Ridge, Lasso, Adaptive Lasso

- Frequentist (Jacobucci & Grimm, 2018; Liang & Jacobucci, 2020; McNeish, 2015)
  - Ridge is advantageous in handling high collinearity.
  - Lasso performed better in shrinking nuisance parameters into zero compared to ridge.
  - Alasso can reduce the bias compared to Ridge and Lasso
- Bayesian
  - Ridge is advantageous in handling high collinearity (Zhang & Liang, 2023)
  - Lasso performed better in maintaining a simple loading structure compared to Ridge (Chen et al., 2021)
  - Alasso performed better in variable selection and estimation accuracy compared to Lasso (Feng et al., 2017)
  - Alasso performed better in detecting non-zero residual correlations compared to Lasso (Pan et al., 2021)

Purpose

- Investigate the similarity and differences in point estimates and variable selection when applying ridge, lasso, and alasso in both frequentist and Bayesian frameworks.
- Explore the strengths and limitations of various regularization methods under frequentist and Bayesian frameworks.

Simulation Study

Design

- Model structure: two- and three-factor models.

The loading matrices for the two- and three-factor models:

$$\Lambda'_{true} = \begin{pmatrix} M & M & M & M & M & 0 & 0 & 0 & C_2 & C_1 \\ 0 & 0 & 0 & C_2 & C_1 & M & M & M & M & M \end{pmatrix}$$

$$\Lambda'_{true} = \begin{pmatrix} M & M & M & M & M & 0 & 0 & 0 & 0 & C_1 & 0 & 0 & 0 & C_2 & 0 \\ 0 & 0 & 0 & C_2 & 0 & M & M & M & M & M & 0 & 0 & 0 & 0 & C_1 \\ 0 & 0 & 0 & 0 & C_1 & 0 & 0 & 0 & C_2 & 0 & M & M & M & M & M \end{pmatrix}$$

- Major loadings: 1
- Number of non-zero cross-loadings per factor: 1 or 2
- Magnitude of non-zero cross-loadings: 0.1, 0.2, 0.3
- Factor correlation: 0.3, 0.5, 0.7
- Sample size: 200, 500, 800

Frequentist Estimation

The regsem package (Jacobucci, 2017): models were selected using BIC for γ, γ_j from 0 to 0.29 with a .01 increment.

Penalty cannot be assigned to all cross-loadings, otherwise the model would not converge.

So we assume that there are at least one item per factor loaded on only one factor, and assign penalty on the cross-loadings with $C-$marks:

$$\mathbf{\Lambda}'_{est} = \begin{pmatrix} 1 & M & M & M & M & 0 & C_0 & C_0 & C_2 & C_1 \\ 0 & C_0 & C_0 & C_2 & C_1 & 1 & M & M & M & M \end{pmatrix}$$

$$\mathbf{\Lambda}'_{est} = \begin{pmatrix} 1 & M & M & M & M & 0 & C_0 & C_0 & C_0 & C_1 & 0 & C_0 & C_0 & C_2 & C_0 \\ 0 & C_0 & C_0 & C_2 & C_0 & 1 & M & M & M & M & 0 & C_0 & C_0 & C_0 & C_1 \\ 0 & C_0 & C_0 & C_0 & C_1 & 0 & C_0 & C_0 & C_2 & C_0 & 1 & M & M & M & M \end{pmatrix}$$

Bayesian Estimation

For the main loadings, intercepts, and factor variance-covariance matrix, diffuse priors were assigned:

$$\lambda_j^m \sim N(\lambda_0, H_{\lambda 0}), \mu_j \sim N(\mu_0, H_{\mu 0}), \mathbf{\Phi}^{-1} \sim Wishart(\mathbf{R}_0, \rho_0) \tag{7}$$

where $\lambda_0 = \mu_0 = 0$, $H_{\lambda 0} = H_{\mu 0} = 1000$, $\mathbf{R}_0$ is the identity matrix and ρ_0 is the number of factors plus one.

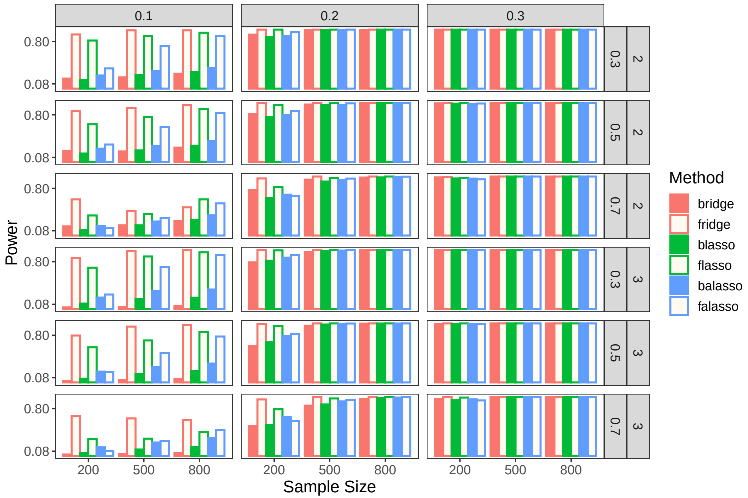
- Number of burn-in iterations were set between 5,000 - 20,000.
- Model Convergence Criteria: EPSR value $<$ 1.05.
- Software: R, JAGS (Plummer, 2003)

Evaluation Criteria

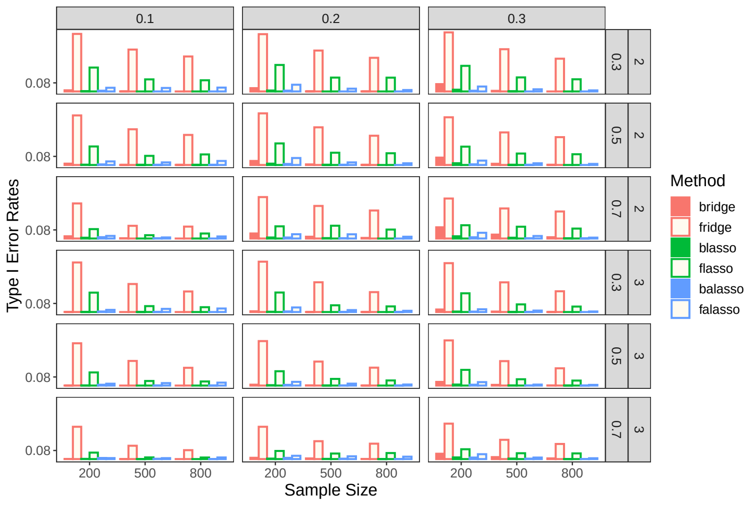
200 datasets per condition.

- Parameter selection (Thresholding rule: for frequentist regularization: $|\beta| > 0$; for Bayesian regularization: $|\beta| > 0.1$; Zhang, Pan, & Ip, 2021):
  - Power
  - Type I error rate
- Parameter Estimation (For Bayesian estimation, median of posterior samples was used as point estimate):
  - Relative bias
  - Root mean square error

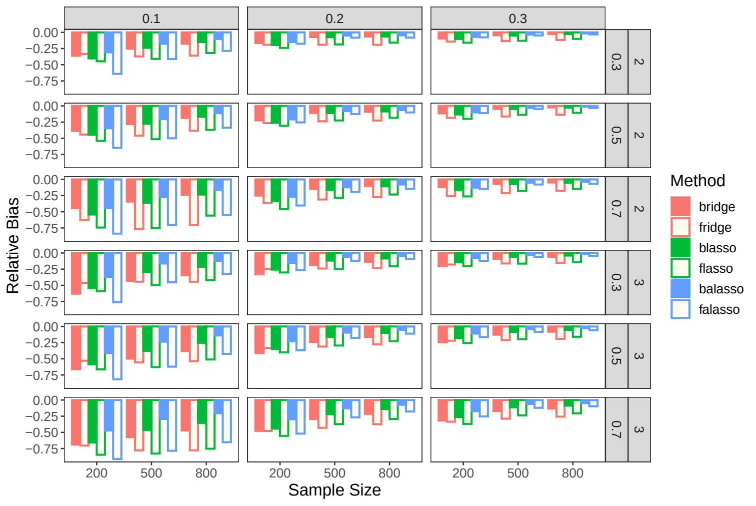
Results: Power



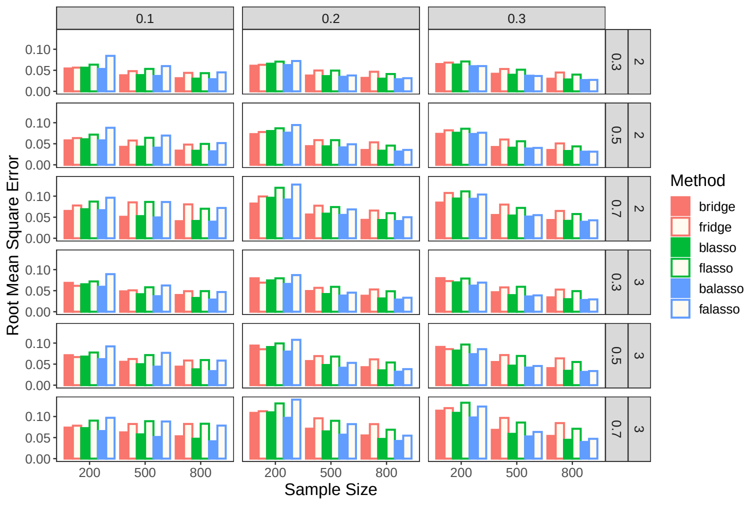
Results: Type I Error Rates



Results: Relative Bias



Results: Root Mean Square Error



Discussion

Method Comparison

- Frequentist methods generally provide greater power but may also encounter inflated type I error rates, particularly for ridge.
  - However, note that different thresholds were used for frequentist and Bayesian methods.
- Bayesian regularization generally performed better in parameter estimation (lower bias and RMSE).
- Frequentist alasso exhibited lower power in general, while showed improved parameter estimates with increasing sample size, compared to ridge and lasso.
- Bayesian alasso outperformed Bayesian ridge and lasso in most conditions, particularly in the three-factor model.

Takeaways

- Frequentist methods generally offer higher power in detecting cross-loadings, while Bayesian methods excel at controlling type I errors
- Use Bayesian Alasso for Lower Bias and RMSE
- Exercise caution about Type I error rates when using Frequentist Ridge and Lasso.
- Findings may vary depending on sample size, model complexity, and the chosen threshold.

Thank you for listening!

Lijin Zhang: lijinzhang@stanford.edu
Xinya Liang: xl014@uark.edu
Junhao Pan: panjunh@mail.sysu.edu.cn