

One- vs. Two-Step Approach in Regularized Bayesian Structural Equation Modeling (SEM)

Research Proposal

Michael Koch (6412157)

Methodology and Statistics for the Behavioral, Biomedical, and Social Sciences

Supervisor: Dr. Sara van Erp

Email: j.m.b.koch@students.uu.nl

Word Count: 750

Intended Journal of Publication: Structural Equation Modeling

In the context of *Confirmatory Factor Analysis (CFA)* a popular model falling under the broader class of *Structural Equation Modeling (SEM)*, the desire for model simplicity manifest itself in the common practice of fitting models with a simple structure. Here, so-called “cross-loadings,” , are fixed to zero. This serves the purpose of both identifying the model, and yielding a factor structure that can be interpreted straightforwardly. However, often times, this practice leads to poor model fit. Consequently, researchers commonly free some cross-loadings based on modification indices to improve model fit. This practice, however, is flawed for a variety of reason, among which risking capitalization on chance [REFs, there’s plenty].

As solution, Muthen and Asparouhov (2012) proposed *Bayesian SEM*. For CFA models, they argued that rather than fixing all cross-loadings to zero, one should make a more realistic and flexible assumption. . . .

Formally, this is achieved by setting the so called Small Variance Prior for the cross-loadings, which is a normal distribution with mean zero and a very small variance σ (e.g.: $\sigma = 0.01$, $\sigma = 0.001$)

$$\mathcal{N}(0, \sigma)$$

By tuning the variance of the prior, one is able to fit a systematically, continuously varying sequence of models. This, in turn, allows for a more continuous process of model identification and model selection.

However, a core issue with the approach by Muthen and Asparouhov (2012) is that applying the Small Variance Prior requires a Two-step approach (Lu, Chow, & Loken, 2016). Not only the cross-loadings close to zero that can be considered irrelevant are shrunk to zero, as desired. Also the cross-loadings that are further from zero, and are therefore likely

to play a more crucial role in accurately modeling the factor structure of the data are shrunken substantially, which introduces bias. Therefore, more advanced priors need to be identified that can outperform the small-variance prior *in a single step*.

A promising candidate is the *Regularized Horseshoe Prior* [1, 2], an extension of the Horseshoe prior (Carvalho, Polson, & Scott, 2010).

Horseshoe prior (Carvalho et al., 2010) It is characterized by a *global* shrinkage component ω , shrinking all parameters towards zero, and a *local* component τ_j allowing large parameters to escape shrinkage. This, in turn results in practically no shrinkage of large coefficients, shrinkage to zero for small coefficients, within a single step. - Mixture of Gaussians.

$$\lambda_j | \tau_j^2 \sim \mathcal{N}(0, \tau_j^2)$$

$$\tau_j \sim C^+(0, \omega) \text{ for } j = 1, \dots, p$$

$$\omega | \sigma \sim C^+(0, \sigma)$$

, where we denote from the common notation of the global shrinkage parameter as λ , as that letter is commonly reserved for factor loadings in CFA.

For ever cross-loading λ_j in the model...

The *Regularized Horseshoe Prior* builds on the Horseshoe prior, by which is achieved by ... - Identification issues you get from (half-)cauchy priors (Ghosh, Li, & Mitra, 2018), especially with weakly identified parameters. - Tail (half-)cauchy - Ensure there's always a little shrinkage, by - -

$$\lambda_j | \hat{\tau}, \omega \sim \mathcal{N}(0, \hat{\tau}^2 \omega), \text{ with } \hat{\tau}^2 = \frac{c^2 t_j^2}{c^2 + \omega^2 \tau_j^2}$$

$$\omega|\omega_0^2 \sim \mathcal{N}(0, \omega_0^2), \text{ with } \omega_0 = \frac{p_0}{p - p_0} \frac{\sigma}{\sqrt{N}}$$

$$\tau_j \sim \mathcal{C}^+(0, 1)$$

$$c^2|\nu, s^2 \sim \mathcal{IG}(\nu/2, \nu s^2/2)$$

In sum, the Regularized Horseshoe Prior is characterized by a unique combination of desired traits. This raises the expectation that it would outperform the Small Variance Prior in selecting the correct loading structure of a CFA model, within a single step. However, while previous research has shown promising results with the Regularized Horseshoe Prior in the selection of the correct predictor structure of Regression models [,], no previous research has validated its performance in the selection of the factor structure in CFA.

To fill this gap, the aim of this study is to compare the Regularized Horseshoe Prior to the Small Variance Prior in their performance in selecting the true factor structure in Confirmatory Factor Analysis (CFA).

Analytic Strategy

A Monte Carlo simulation study is conducted using stan (Carpenter et al., 2017). True Positives vs. False Positives in estimating truly non-zero cross-loadings as non-zero are considered as main outcome (ROC). Conditions are based on (Lu et al., 2016) and will include two factor structures (1 non-zero cross-loading, several non-zero cross-loadings), three sample sizes (100, 200, 300), and three magnitudes of the cross-loadings (0.1, 0.2, 0.3). These sizes are . . .

References

- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M.,
... Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of
Statistical Software*, 76(1), 1–32. <https://doi.org/10.18637/jss.v076.i01>
- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for
sparse signals. *Biometrika*, 97(2), 465–480.
<https://doi.org/10.1093/biomet/asq017>
- Ghosh, J., Li, Y., & Mitra, R. (2018). On the Use of Cauchy Prior Distributions for
Bayesian Logistic Regression. *Bayesian Analysis*, 13(2), 359–383.
<https://doi.org/10.1214/17-BA1051>
- 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.
<https://doi.org/10.1080/00273171.2016.1168279>
- Muthen, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation
of substantive theory, 78. <https://doi.org/10.1037/a0026802>

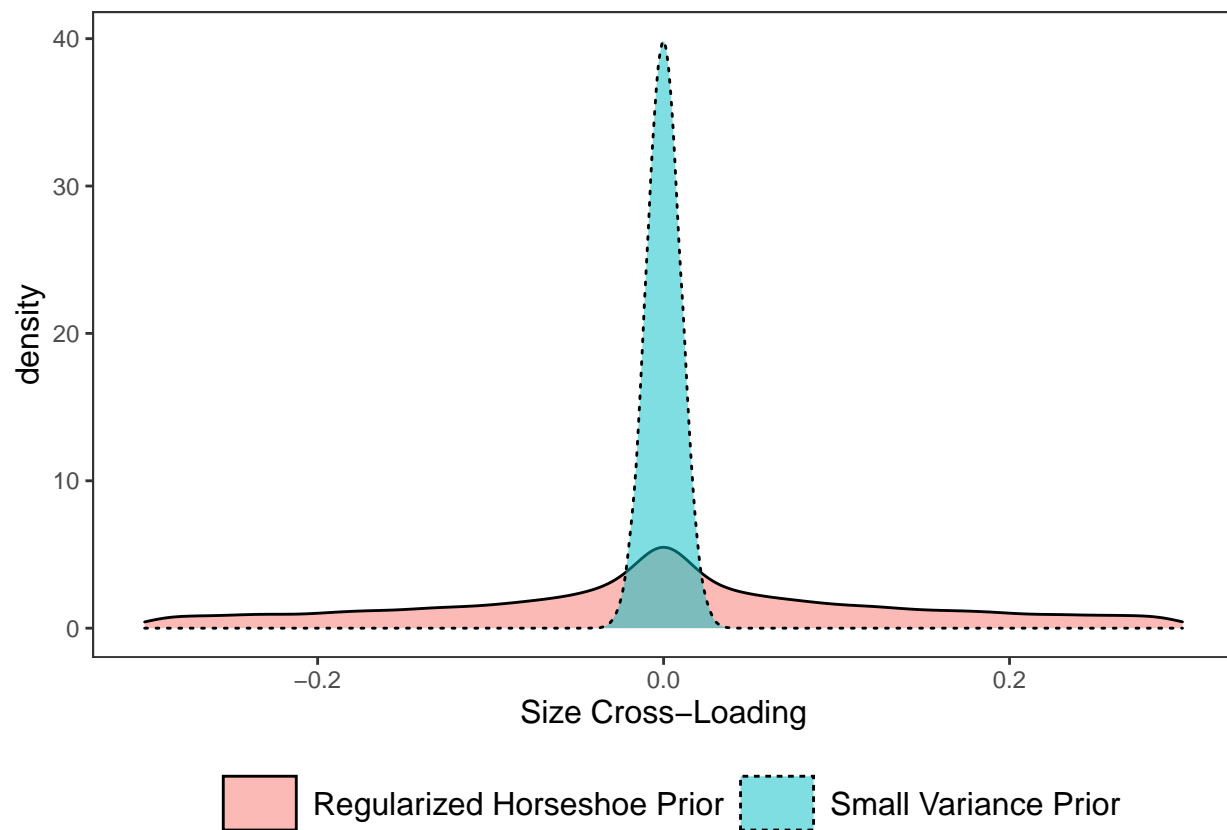


Figure 1. Density Plots of the Regularization Priors of Interest