

One- vs. Two-Step Approach in Regularized Bayesian Confirmatory Factor Analysis (CFA)

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: 817

Intended Journal of Publication: Structural Equation Modeling

FETC permission obtained on 27'th September 2021

A common model-constraint in *Confirmatory Factor Analysis (CFA)* is fixing all cross-loadings to zero. This is done both to identify the model, and to achieve an easy-to-interpret factor structure. However, this practice often 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 reasons, among which risking capitalization on chance (Jacobucci, Grimm, & McArdle, 2016; Muthen & Asparouhov, 2012).

As solution, Muthen and Asparouhov (2012) proposed that rather than fixing all cross-loadings to zero, one should assume that *most* cross-loadings are zero. 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 (e.g.: $\sigma = 0.01$, $\sigma = 0.001$). Such prior gives cross-loadings of zero large prior density, but cross-loadings that are far from zero almost zero prior density (see Figure 1 for an example with $\sigma = 0.01$). Consequently, all cross-loadings in the model are shrunk substantially, those close to zero to zero, and those far from zero towards zero. The prior's variance is thus the deciding factor in how admissive the model is of cross-loadings closer to zero. Hence, by tuning the variance of the prior, one is able to fit a systematically and continuously varying sequence of models allowing for a much more continuous process of model identification and model selection than classical CFA.

However, an issue with this method is that not only the cross-loadings close to zero that are considered irrelevant are shrunk to zero, as desired. Also the cross-loadings that are further from zero, are shrunk towards zero, which introduces bias. The method thus requires a two-step approach. First, the model is estimated with shrinkage. Then the model is re-estimated where the cross-loadings that have been shrunk to zero in the previous step are fixed to zero, and the remaining cross-loadings are estimated without shrinkage. To overcome this, alternative regularization priors need to be identified that can outperform the small-variance prior *in a single step*.

Lu, Chow, and Loken (2016) pointed out that the Bayesian CFA approach by Muthen and Asparouhov (2012) can be viewed as a form of *Regularization*, where rather than selecting variables that are relevant as predictors in a regression model, cross-loadings that are relevant in modeling a factor structure are selected. The literature on regularization in a regression context (see Van Erp, Oberski, & Mulder, 2019 for an overview) can thus provide alternative priors. Here, the *Regularized Horseshoe Prior* (Piironen & Vehtari, 2017), an extension of the *Horseshoe prior* (Carvalho, Polson, & Scott, 2010), appears as particularly promising. Both priors are characterized by a *global* shrinkage component ω , shrinking all parameters towards zero (explaining the steep peak $\lambda_c = 0$ in Figure 1), and a *local* component τ_j , which gives the prior its heavy tails and thereby allows large parameters to escape the shrinkage entirely¹. Although being one of the core qualities of the original Horseshoe Prior, in practice not shrinking large coefficients at all can lead to computational issues in the MCMC sampling of the model (Ghosh, Li, & Mitra, 2018; Piironen & Vehtari, 2017). To avoid this, the *Regularized Horseshoe Prior*² is designed to ensure that there is always a little shrinkage, even for large parameters (Piironen & Vehtari, 2017):

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

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

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

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

where p_0 represents a prior guess of the number of relevant cross-loadings.

¹ We deviate from the common notation of the global shrinkage parameter as λ , as that letter is commonly reserved for factor loadings in CFA.

² Following previous research, we set the hyper-parameters $\nu = 4$ and $s^2 = 2$ (Piironen & Vehtari, 2017; Van Erp et al., 2019).

While the Regularized Horseshoe Prior has been shown to perform well in the selection of relevant predictors in regression (Piironen & Vehtari, 2017; Van Erp et al., 2019), no previous research has validated its performance in the context of selecting relevant cross-loadings 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. 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), yielding a total of $2 \times 3 \times 3 = 18$ main conditions. All models will be sampled using the NUTS-sampler (Betancourt, 2018), where each time four chains are sampled for 1000 iterations, following a burn-in-period of 1000.

References

- Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [Stat]*. Retrieved from <http://arxiv.org/abs/1701.02434>
- 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>
- Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4), 555–566. <https://doi.org/10.1080/10705511.2016.1154793>
- 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>
- Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2), 5018–5051. <https://doi.org/10.1214/17-EJS1337SI>
- Van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, 89, 31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>

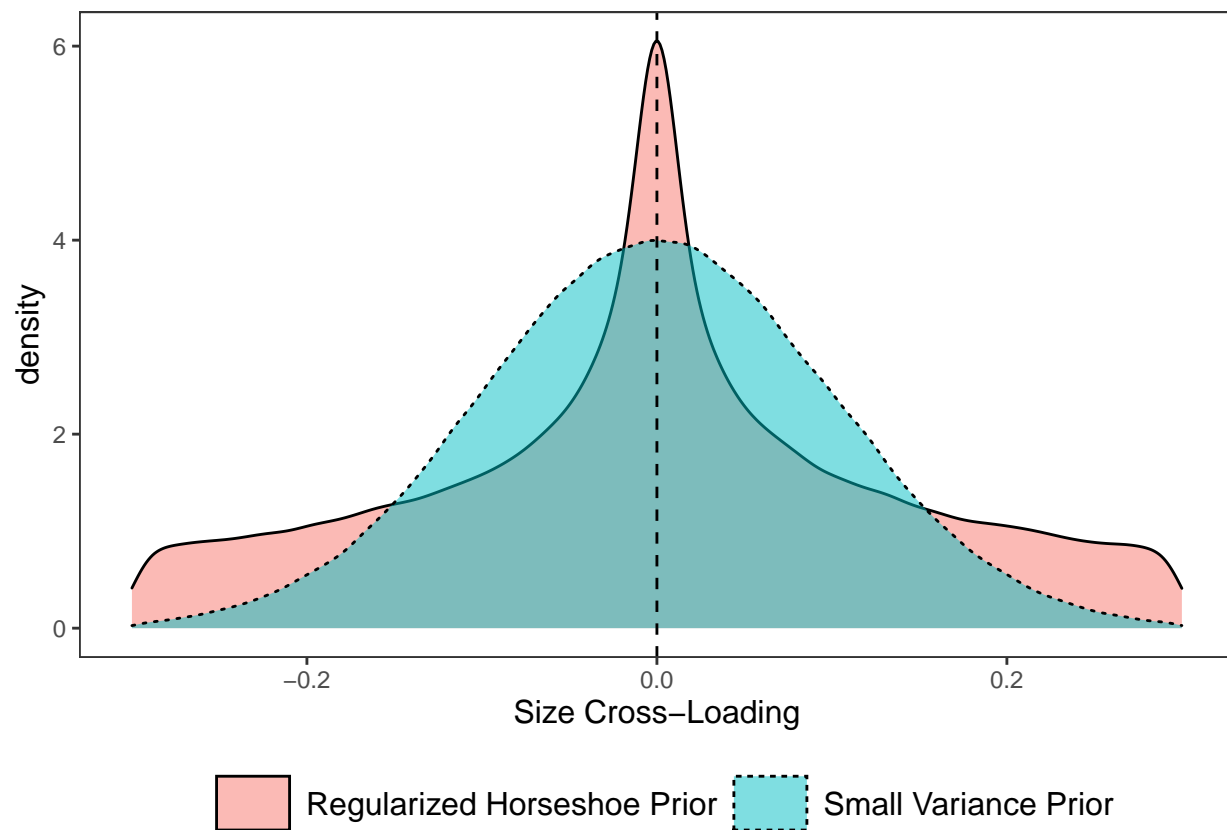


Figure 1. Density Plots of the Regularization Priors of Interest