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: 750

Intented Journal of Publication: Structural Equation Modeling

FETC permission obtained on 27'th September 2021

*Confirmatory Factor Analysis (CFA)* is an essential tool for modeling measurement structures. Usually, all cross-loadings, factor loadings that relate items to factors that they theoretically do not belong to, are fixed to zero to identify the model. This often leads to poor model fit, and forces researchers to free some cross-loadings again to improve fit, a practice that 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 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 no prior density (see Figure 1 for an example with $\sigma = 0.01$). Consequently, all cross-loadings in the model are shrunken. The larger the prior's variance, the more admissive the model is in the amount of deviation from zero it allows. Hence, by tuning the variance of the prior, one is able to fit a systematically and continuously varying sequence of models. This allows for a much more continuous process of model identification than classical CFA.

An issue with Muthen and Asparouhov (2012)'s approach is that not only the cross-loadings close to zero that are considered irrelevant are shrunken to zero, as desired, but also the ones further from zero are shrunken 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, with cross-loadings that have been shrunken to zero in the previous step fixed to zero, and the remaining cross-loadings estimated without shrinkage. To overcome this, alternative 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 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 regression (see Van Erp, Oberski, & Mulder, 2019 for an overview) can thus provide alternative priors. 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 $\omega$, shrinking all parameters towards zero (explaining the steep peak at zero in Figure 1), and a *local* component $\tau_j$, which gives the prior its heavy tails and thereby allows large parameters to escape the shrinkage entirely[1]. Although being one of the core qualities of the original Horseshoe Prior, in practice not shrinking large coefficients at all leads estimation issues (Ghosh, Li, & Mitra, 2018; Piironen & Vehtari, 2017). To avoid this, the *Regularized* Horseshoe Prior[2] is designed to ensure that even for large parameters there is at least a little shrinkage (Piironen & Vehtari, 2017):

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

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

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

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

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

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 selecting relevant cross-loadings in CFA.

---

[1] We deviate from the common notation of the global shrinkage parameter as $\lambda$, 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).

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 (Stan Development Team, 2021). 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 (one non-zero cross-loading, two 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 No-U-Turn-sampler (Betancourt, 2018; Homan & Gelman, 2014), 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

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

Homan, M. D., & Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, *15*(1), 1593–1623.

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

Stan Development Team. (2021). Stan User Guide. Retrieved from https://mc-stan.org/docs/2_27/stan-users-guide-2_27.pdf

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

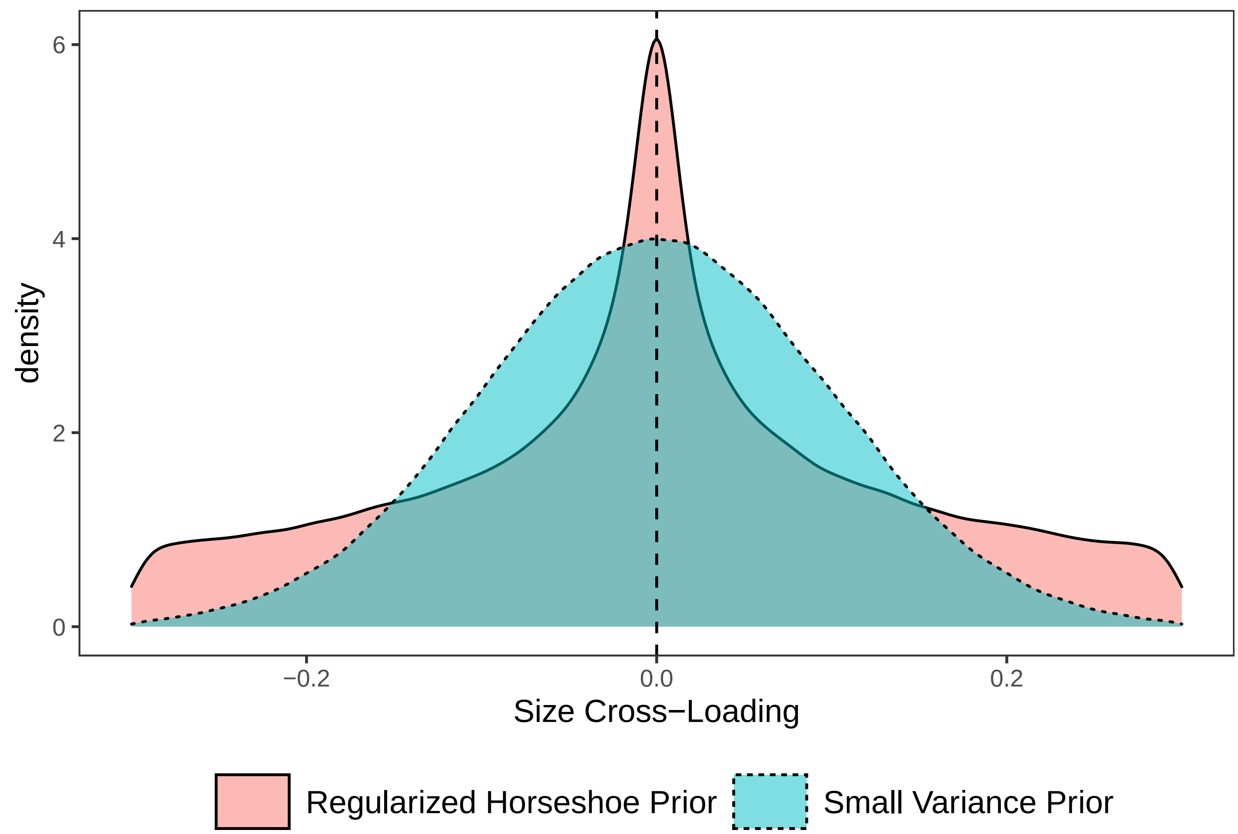


*Figure 1*. Density Plots of the Regularization Priors of Interest