Bayesian multilevel structural equation modeling: An investigation into robust prior distributions.

Background and Purpose

SEM allows researchers to control for measurement error due to the fact that only a finite number of items are sampled

multilevel analysis takes into account sampling error due to the fact that only a finite number of individuals is sampled.

importance of multilevel model:

example: "Big-Fish-Little-Pond-Effect" at the individual (or within) level, student achievement has a positive effect on academic self-concept, whereas at the between-level, school-average achievement has a negative effect on academic self-concept

advantages of Bayesian estimation in multilevel modeling

- First, MCMC sampling does not require multi-dimensional numerical integration enabling complex models with categorical indicators to be estimated (see e.g., Asparouhov and Muthén, 2012).

- Second, estimates of transformed parameters, such as indirect structural effects, can be easily obtained by transforming the MCMC samples and credible intervals are obtained automatically (Yuan and MacKinnon,2009).
- Third,through the prior distribution, the problem of variances that are estimated to be negative(i.e., Heywood cases) can be solved
- Finally, it has been shown that Bayesian MLSEM can provide accurate estimates with a smaller number of groups compared to frequentist estimation (Hox et al., 2012; Zitzmann et al., 2016)

random effects variance parameter

decides how much pooling across observations takes place

priors:

inverse-gamma(α, α) with small α has often used as a default uninformative prior distribution

- this specification is actually very informative and highly sensitive to the exact choice of the hyperparameters (see e.g.,Gelman, 2006; Klein and Kneib, 2016; Lunn et al., 2009)
- the inverse-Gamma prior does not have support for random effects variances equal to zeroand will therefore always favor more complex solutions with less pooling

uniform improper priors

- this is in line with the "objective" Bayesian approach (Berger, 2006) since these priors will, in simple models, result in the same estimates as classical ML estimation
- main issue: at the resulting posterior might be improper as well in cases where there is a limited amount of information in the data on the higher
- Note that inverse-gamma(α, α) as α goes to zero mates the improper prior $p(\sigma^2) \propto \sigma^{-2}$. As a result, the inverse-Gamma prior can also lead to an unstable MCMC sampler

weakly-informative prior

- perform some shrinkage of the random effects variance parameter towards zero
- has been test in multilevel models, has not been extend to multilevel sem

Purpose: compare default specifications of the robust priors as well as informative (in)accurate specification

Robust Prior Distribution

In general, a prior with heavier tails is more robust since extreme values will not be shrunken towards zero as much compared to a prior with thinner

Model: Bayesian doubly latent categorical multilevel model (assume binary indicators to keep the computation time feasible in our simulation studies)

An important quantity in multilevel models is the <mark>variance partition coefficient (VPC;</mark> Goldstein et al., 2002), which is a measure of the correlation between students in the same school. In multilevel SEM, the VPC ρ_k for item k is defined as (see e.g., Muthén, 1994):

$$\rho_k = \frac{\lambda_k^{2,B}\omega_B^2 + \sigma_{B,k}^2}{(\lambda_k^{2,B}\omega_B^2 + \sigma_{B,k}^2) + (\lambda_k^{2,W}\omega_W^2 + \sigma_{W,k}^2)} \tag{6}$$

Student's t family

cauchy: special form of student't

Gelman (2006): half-Cauchy prior is a reasonable default option. Polson and Scott (2012), who have shown that it has excellent frequentist risk properties

There are some reports, however, that the heavy tails of the half-Cauchy prior can lead to numerical difficulties in MCMC sampling(Ghosh et al., 2018; Piironen and Vehtari, 2015). This issue generally arises in situations when there is not enough information in the data such that the parameter is weakly or non-identified.

The heavy tails of the (half-)Cauchy prior do not perform enough shrinkage in such cases to identify the parameter

F family

the family of F priors, also known as scaled Beta2 or generalized beta prime

can be seen as a generalization of the Student's t family.

To implement the F prior, the following Gamma mixture of inverse-Gamma parametrization can be used:

$$p(\theta^2) \sim IG(\frac{\nu_2}{2}, \tau^2) \tag{9}$$

$$p(\tau^2) \sim G(\frac{\nu_1}{2}, \sigma_0^{-2}), \tag{10}$$

Scale dependent family

Simpson et al. (2017) have proposed the use of penalized complexity priors as general default priors. The basic idea of these priors is that deviations from a simpler base model arepenalized thereby avoiding a model that overfit

horseshoe prior. Transforming the prior on the distance to the original space leads to a type-2 Gumbel prior for the precision or, equivalently, an exponential distribution with scale σ_0^2 for the standard deviation[5]:

$$p(\theta; \sigma_0^2) = \frac{1}{\sigma_0^2} \exp(-\frac{1}{\sigma_0^2}\theta) \tag{11}$$

Lower values for σ_0^2 will result in an increased penalisation for deviating from the base
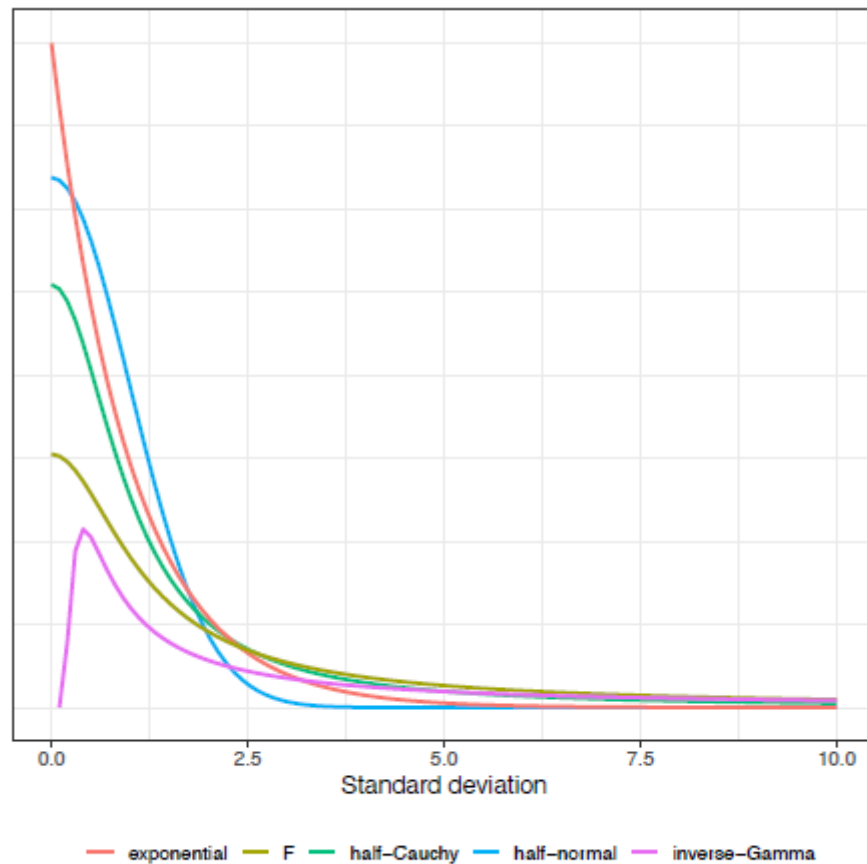

Figure 2: Densities of the priors on the standard deviations.

Simulation Study

Study 1: compare robust priors with inverse-gamma

The main difference between the inverse-Gamma prior and the more robust alternatives is that the inverse-Gamma prior has zero density at zero.

Manipulation: VPC, contextual effect $(\beta^B - \beta^W)$ with group-mean centering

standardized contextual effect:

$$ES = \frac{2\beta^C \sigma_x^B}{\sigma_x^W (\beta^W)^2 + \omega^W},$$ (12)

where β^C denotes the contextual effects which is equal to $\beta^B - \beta^W$ in the case of group-mean centering, and σ_x^B and σ_x^W denote the standard deviations of the predictor at the between and within level, respectively. We consider a small negative standardized contextual effect and no contextual effect, i.e., $ES = c(-0.2, 0)$. We choose these

groups: 20, sample size per group: 20

Hox et al. (2012) who concluded that 20 groups are sufficient for multilevel SEM.

Monte Carlo SEs

computed the Monte Carlo SE for every outcome measure to quantify the uncertainty inthe simulation results (Morris et al., 2019). All Monte Carlo SEs were sufficiently small to conclude that 500 replications was enough (the maximum MCSE was 0.03 for the coverage percentages)

strict convergence criteria: split PSRF (cutoff: 1.05). present the results for those conditions with at least 50% convergence

Main Results

the inverse-Gamma prior is most problematic with multiple conditions resulting in less than 50% convergence

Study 2: Influence prior misspecifications

In the second study, we focus on informative prior distributions. Specifically, we choose the prior scale σ_0 in such a way that $P(\sigma_0 > \sigma_{true}) = \alpha$, where σ_{true} equals the population value under which that parameter was simulated. We consider two settings: in the "correct" informative prior (i.e., HNinf, HCinf, Finf, and EXPinf), $\alpha = 0.5$ such that the resulting prior has sufficient probability around the population value. In the "incorrect" informative prior (i.e., HNinc, HCinc, Finc, and EXPinc), on the other hand, $\alpha = 0.05$ such that the prior has only very little mass on the population value. Since we never know the true value

Conclusion：

although the robust priors outperform the traditional inverse-Gamma prior, their hyperparameters do require careful consideration

The differences between the robust prior distributions and the uniform prior were gen-erally small. The main differences in results in the simulation studies appear to arise dueto the population conditions rather than the prior distribution

My Thoughts

limited findings and limited research conditions

-