Prior sensitivity analysis in default Bayesian structural equation modeling

March 16, 2018, SEM Working Group

Sara van Erp[1] Joris Mulder[1] Daniel L. Oberski[2]

[1]Tilburg University

[2]Utrecht University

Netherlands Organisation
for Scientific Research

Bayesian Structural Equation Modeling (BSEM)

Classical SEM

SEMs offer a good representation of substantive theories.

However..

- Nonconvergence
- Inadmissible solutions
- Large number of groups
- Computationally inconvenient

Classical SEM

SEMs offer a good representation of substantive theories.

However..

- Nonconvergence
- Inadmissible solutions
- Large number of groups
- Computationally inconvenient

Bayesian SEM thought to solve these problems.

Bayesian SEM

Prior specification is important, but difficult.

Reliance on *default* priors.

Question

Do all default priors perform similarly in BSEM?

Bayesian SEM

Prior specification is important, but difficult.

Reliance on *default* priors.

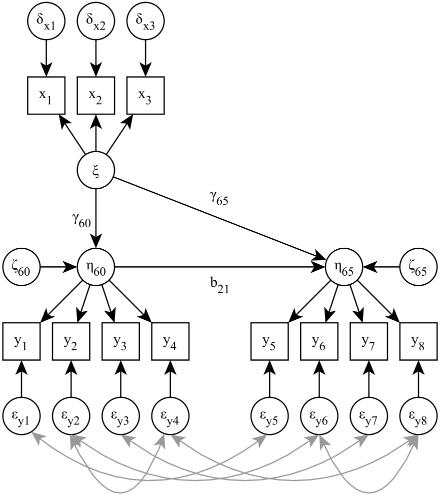
Question

Do all default priors perform similarly in BSEM?

Answer

No! So conduct prior sensitivity analyses!

Model



δx1
δx2
δx3
x1
x2
x3
ξ
γ65
γ60
ζ60
η60
η65
ζ65
b21
y1
y2
y3
y4
y5
y6
y7
y8
εy1
εy2
εy3
εy4
εy5
εy6
εy7
εy8

Priors

Priors

We consider three types of default priors:

1. Noninformative improper priors
2. Vague proper priors
3. Empirical Bayes priors

Noninformative improper priors

- $p(\sigma^2) \propto \sigma^{-2}$ implemented as inverse Gamma$(0, 0)$
- $p(\sigma^2) \propto \sigma^{-1}$ implemented as inverse Gamma$(-\frac{1}{2}, 0)$
- $p(\sigma^2) \propto 1$ implemented as inverse Gamma$(-1, 0)$ (Mplus default)

Noninformative improper priors

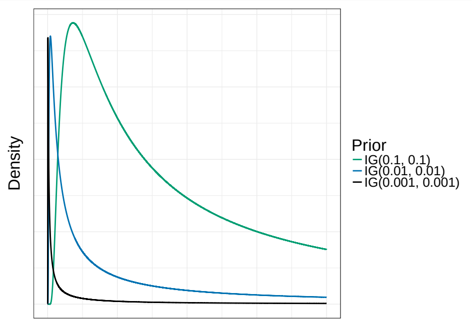
- $p(\sigma^2) \propto \sigma^{-2}$ implemented as inverse Gamma$(0, 0)$
- $p(\sigma^2) \propto \sigma^{-1}$ implemented as inverse Gamma$(-\frac{1}{2}, 0)$
- $p(\sigma^2) \propto 1$ implemented as inverse Gamma$(-1, 0)$ (Mplus default)

Problem: can result in improper posteriors

Vague proper priors

Variances: approximate $p(\sigma^2) \propto \sigma^{-2}$

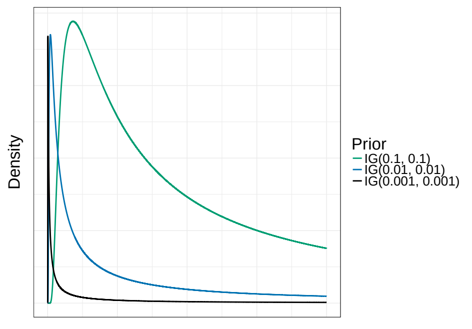
- inverse Gamma(.001, .001) (WinBUGS default)
- inverse Gamma(.01, .01)
- inverse Gamma(.1, .1)



Vague proper priors

Variances: approximate $p(\sigma^2) \propto \sigma^{-2}$

- inverse Gamma(.001, .001) (WinBUGS default)
- inverse Gamma(.01, .01)
- inverse Gamma(.1, .1)



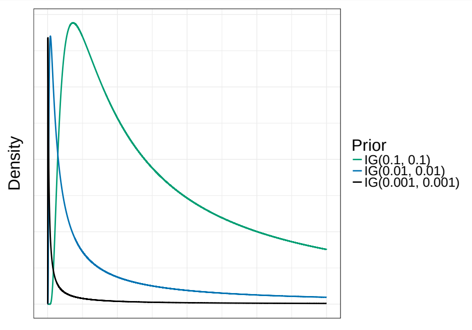
Location parameters: approximate flat prior

- Normal$(0, 10^{10})$ (Mplus default)
- Normal$(0, 1000)$ & Normal$(0, 100)$ (Blavaan default)

Vague proper priors

Variances: approximate
$p(\sigma^2) \propto \sigma^{-2}$

- inverse Gamma$(.001, .001)$ (WinBUGS default)
- inverse Gamma$(.01, .01)$
- inverse Gamma$(.1, .1)$



Location parameters: approximate flat prior

- Normal$(0, 10^{10})$ (Mplus default)
- Normal$(0, 1000)$ & Normal$(0, 100)$ (Blavaan default)

Problem: choice of hyperparameters can arbitrarily influence the results.

Empirical Bayes priors

Variances: inverse Gamma $\left(\frac{1}{2}, \hat{\sigma}^2 \cdot Q^{-1}(\frac{1}{2}, \frac{1}{2})\right)$

- Information of one data point
- Median equals ML estimate

Empirical Bayes priors

Variances: inverse Gamma $\left(\frac{1}{2}, \hat{\sigma}^2 \cdot Q^{-1}(\frac{1}{2}, \frac{1}{2})\right)$

- Information of one data point
- Median equals ML estimate

Location parameters: Normal$(0, \hat{\mu}^2 + \hat{\sigma}^2)$

- e.g. Normal$(0, \hat{\nu}_y^2 + \hat{\sigma}_y^2)$
- Minimal information
- Positive support where the likelihood is concentrated

Simulation study

Data generating conditions

- Political democracy and industrialization model
- $N \in \{35, 75, 150, 500\}$
- Population values equal to ML estimates
- Varied population values for: γ_{65}, γ_{60}, λ_{y4}, λ_{y8}, $\boldsymbol{\Omega_D}$
- Analyzed with the different default priors & ML

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$
- *Relative bias*: ML performed best. Mplus default for location parameters & EB1 for variances

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$
- *Relative bias*: ML performed best. Mplus default for location parameters & EB1 for variances
- *MSE*: Methods perform similarly except for the vague proper priors

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$
- *Relative bias*: ML performed best. Mplus default for location parameters & EB1 for variances
- *MSE*: Methods perform similarly except for the vague proper priors
- *Coverage*: $\pi(\sigma^2) \propto \sigma^{-1}$ and IG$(.1, .1)$ performed best

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$
- *Relative bias*: ML performed best. Mplus default for location parameters & EB1 for variances
- *MSE*: Methods perform similarly except for the vague proper priors
- *Coverage*: $\pi(\sigma^2) \propto \sigma^{-1}$ and IG$(.1, .1)$ performed best
- *Quantiles*: Noninformative improper priors and vague normal prior performed best

Results

- *Convergence*: Nonconvergence for $\pi(\sigma^2) \propto \sigma^{-2}$
- *Relative bias*: ML performed best. Mplus default for location parameters & EB1 for variances
- *MSE*: Methods perform similarly except for the vague proper priors
- *Coverage*: $\pi(\sigma^2) \propto \sigma^{-1}$ and IG$(.1, .1)$ performed best
- *Quantiles*: Noninformative improper priors and vague normal prior performed best

Conclusion

Performance of default priors varies greatly across conditions and outcomes. Important to conduct a prior sensitivity analysis.

Prior Sensitivity Analysis

Steps

1. Decide which parameters to investigate
2. Decide which priors to include
3. Technical implementation
4. Interpretation of the results:
   no differences; differences between informative & default; differences between all priors.

Conclusions

Conclusions

- BSEM is useful and popular
- Prior specification can be difficult
- "Default" priors vary in performance
- Prior sensitivity analysis is essential
- More robust priors should be used (or better EB priors?)

Questions?

Paper: http://dx.doi.org/10.1037/met0000162
Code prior sensitivity analysis:
http://dx.doi.org/10.1037/met0000162.supp
Preprint: https://osf.io/zt8e9/

s.vanerp_1@tilburguniversity.edu
@SaravanErp
https://saravanerp.com

Model

The structural model (for $i = 1, \ldots, n$) is given by:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\xi_i + \boldsymbol{\zeta}_i \quad \text{with } \xi_i \sim N(\mu_\xi, \omega_\xi^2),$$
$$\text{and } \boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\zeta}})$$

The measurement model is given by:

$$\boldsymbol{y}_i = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i + \boldsymbol{D}_i + \boldsymbol{\epsilon}_i^y \quad \text{with } \boldsymbol{D}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{D}}),$$
$$\text{and } \boldsymbol{\epsilon}_i^y \sim N(\mathbf{0}, \boldsymbol{\Sigma}_y)$$
$$\boldsymbol{x}_i = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\xi_i + \boldsymbol{\delta}_i^x \quad \text{with } \boldsymbol{\delta}_i^x \sim N(\mathbf{0}, \boldsymbol{\Sigma}_x)$$

Data generating conditions

Variable	Levels	Specifications
Sample size	4	$N \in \{35, 75, 150, 500\}$
Direct effect	3	$\gamma_{65} = 0$
		$\gamma_{65} = 1$
		$\gamma_{65} = 2$
Indirect effect	3	$\gamma_{60} \times b_{21} = 0 \times 0.837$
		$\gamma_{60} \times b_{21} = 1 \times 0.837$
		$\gamma_{60} \times b_{21} = 2 \times 0.837$
Loadings	3	$\lambda_{y4} = \lambda_{y8} = 0$
		$\lambda_{y4} = \lambda_{y8} = 1$
		$\lambda_{y4} = \lambda_{y8} = 2$
Error covariances	2	$\boldsymbol{\Omega_D} = 0$
		$\boldsymbol{\Omega_D} = 1$

Analysis conditions

Prior type	Prior specification
Noninformative improper	$\pi(\sigma^2) \propto 1$ & $\mathsf{N}(0, 10^{10})$
	$\pi(\sigma^2) \propto \sigma^{-1}$ & $\mathsf{N}(0, 10^{10})$
	$\pi(\sigma^2) \propto \sigma^{-2}$ & $\mathsf{N}(0, 10^{10})$
Vague proper	$\mathsf{IG}(.001, .001)$ & $\mathsf{N}(0, 10^{10})$
	$\mathsf{IG}(.01, .01)$ & $\mathsf{N}(0, 10^{10})$
	$\mathsf{IG}(.1, .1)$ & $\mathsf{N}(0, 10^{10})$
Vague normal:	$\mathsf{N}(0, 1000)$ & $\mathsf{N}(0, 100)$ & $\pi(\sigma^2) \propto 1$
Empirical Bayes	
EB1:	$\mathsf{IG}\left(\frac{1}{2}, \hat{\sigma}^2 * Q^{-1}(\frac{1}{2}, \frac{1}{2})\right)$ & $\mathsf{N}(0, \hat{\mu}^2 + \hat{\sigma}^2)$
EB2:	$\pi(\sigma^2) \propto 1$ & $\mathsf{N}(0, \hat{\mu}^2 + \hat{\sigma}^2)$