Bayesian SEM in blavaan: Estimation and model comparison results

Ed Merkle

Acknowledgments

- Acknowledgments:
  - Yves Rosseel (blavaan development)
  - Daniel Furr, Sophia Rabe-Hesketh (information criteria results)
  - NSF grant 1460719

Overview

- blavaan: Translating lavaan syntax to JAGS/Stan code; estimating/summarizing Bayesian models.
- Goals for today:
  - Comparison of JAGS vs Stan estimation of SEMs (Stan is a relatively new development).
  - Discussion of computational methods for DIC/WAIC/LOOIC in models with latent variables.

Overview

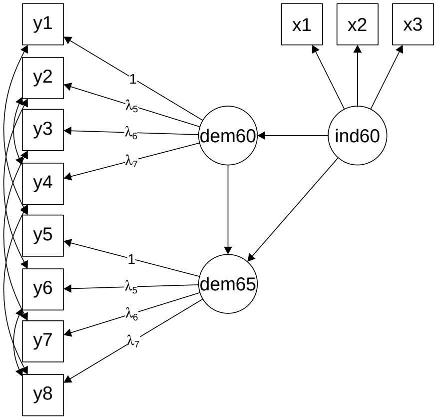
- JAGS and Stan differ in MCMC samplers employed: JAGS uses traditional samplers, whereas Stan uses Hamiltonian Monte Carlo (improved sampling via likelihood derivatives).

- To sample from SEMs more quickly and efficiently, we can use some tricks to define the models. JAGS tricks differ from Stan tricks.

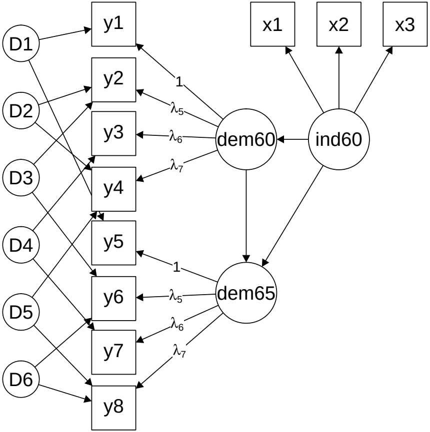
Tricks

- To improve sampling speed and efficiency, we focus on multivariate distributions in the model (typically, inverse and determinant of multivariate normal covariance matrix)
  - JAGS: Overparameterize the model to obtain conditional independence
  - Stan: Computational shortcuts in evaluating multivariate normal distribution of latent variables (inverse & determinant)

JAGS Tricks


y1
y2
y3
y4
y5
y6
y7
y8
x1
x2
x3
dem60
dem65
ind60
1
λ5
λ6
λ7
1
λ5
λ6
λ7

JAGS Tricks



y1
y2
y3
y4
y5
y6
y7
y8
x1
x2
x3
D1
D2
D3
D4
D5
D6
dem60
ind60
dem65
1
λ_5
λ_6
λ_7
1
λ_5
λ_6
λ_7

Stan Tricks

- Stan works better with identified likelihoods (the JAGS tricks have completely failed so far).
    - Speed can be gained through economic evaluation of the latent variable distribution (multivariate normal).
    - Stan provides functionality to easily specify a new likelihood. So we can define a new multivariate normal likelihood that makes use of the SEM framework.

Stan Tricks

- Typical SEM distribution of latent variables (η):

$$\boldsymbol{\eta} \sim N(\boldsymbol{\alpha}, (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{B}')^{-1})$$

Stan Tricks

- Inverse covariance matrix often can be written to avoid inverses:

$$\begin{aligned}\boldsymbol{\Sigma} &= (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{B}')^{-1} \\ \boldsymbol{\Sigma}^{-1} &= (\boldsymbol{I} - \boldsymbol{B}')\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{B})\end{aligned}$$

Stan Tricks

- Determinant of the covariance matrix can also be simplified:

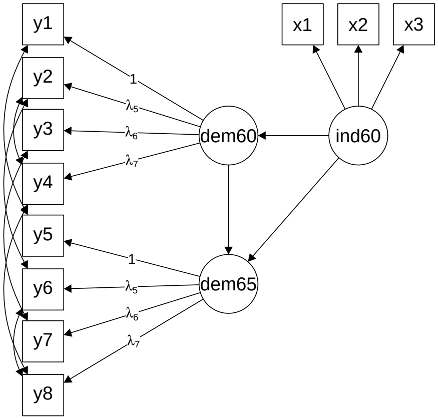
$$\det(\boldsymbol{\Sigma}) = (\det(\boldsymbol{I} - \boldsymbol{B}))^{-1} \det(\boldsymbol{\Psi}) (\det(\boldsymbol{I} - \boldsymbol{B}))^{-1}.$$

  - $\boldsymbol{B}$ is often lower triangular (recursive models), so $\det((\boldsymbol{I} - \boldsymbol{B}))$ is the product of diagonal entries.
  - $\boldsymbol{\Psi}$ is often diagonal, so $\det(\boldsymbol{\Psi})$ is the product of diagonal entries.

Stan Tricks

- For Stan, blavaan looks at the model syntax and employs the matrix tricks where it can.
- There is potential for further improvements for specific types of models (e.g., path analysis).

Comparison 1



Comparison 1

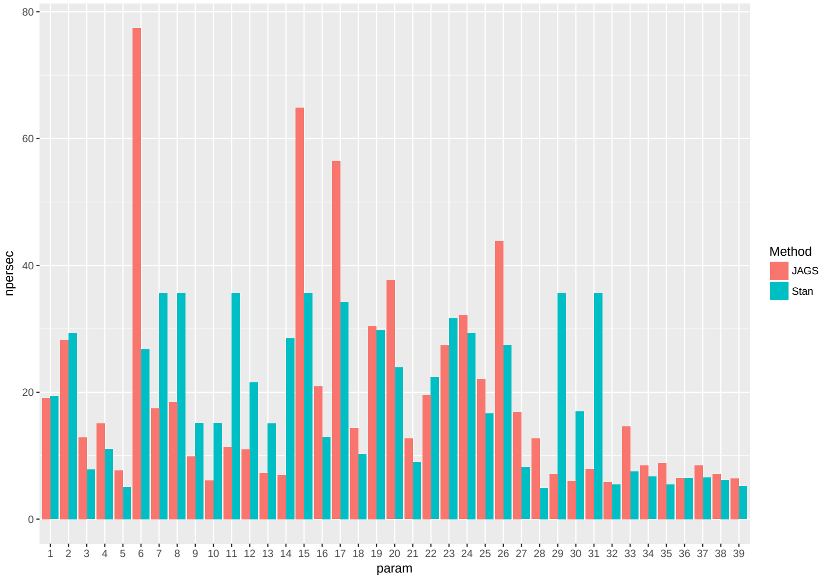
- Run 3 parallel chains in JAGS and in Stan for 1000 sample iterations (1000 burnin in JAGS, 100 burnin in Stan).
  - Posterior means/SDs are equal enough
  - JAGS: 15sec
  - Stan: 115sec with compilation, 84sec after compilation

Comparison 1

- But raw timing isn't really want we want. Stan produces better samples (less autocorrelation), so we can get by with fewer Stan samples.
- We should instead look at effective sample size, translating 1000 correlated samples to some smaller number of independent samples.

Comparison 1

- Effective sample size per second:



Comparison 2

- What about something more complex? (from Kievit et al, in press)

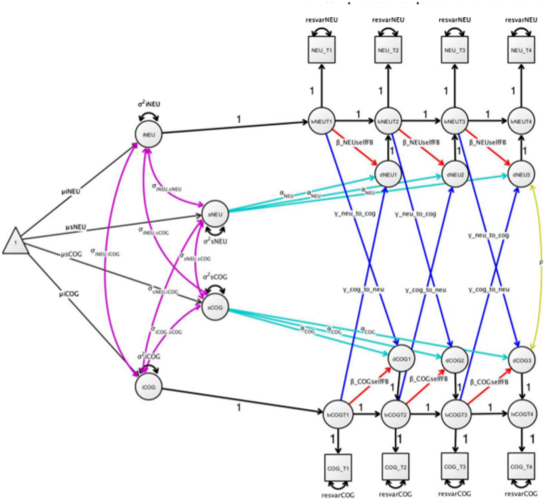


Fig. 5. Bivariate Dual Change Score Model. This more complex latent change score model captures both the stable change over time in the form of slopes (sCOG and sNEU), as well as more fine-grained residual changes. Note this model incorporates latent variables at each timepoint – See Newsom (2015, p. 135) for more detail.

Comparison 2

- Run 3 parallel chains in JAGS and in Stan for 1000 sample iterations (4000 burnin in JAGS, 300 burnin in Stan).
  - JAGS: 20min, but fails to converge due to high autocorrelation. Requires longer runs with thinning.
  - Stan: 1hr25min with compilation, 1hr23min after compilation

Comparison 2

- Effective sample size per *minute*:

Comparison 2

- For parameters of main interest (feedback, slope/intercept parameters, "coupling" parameters), Stan has a small edge on JAGS. The reverse is true of other parameters.
- JAGS chains have large autocorrelation, so thinning and long chains are required there.
- If you dislike waiting, neither is optimal.

Information criteria

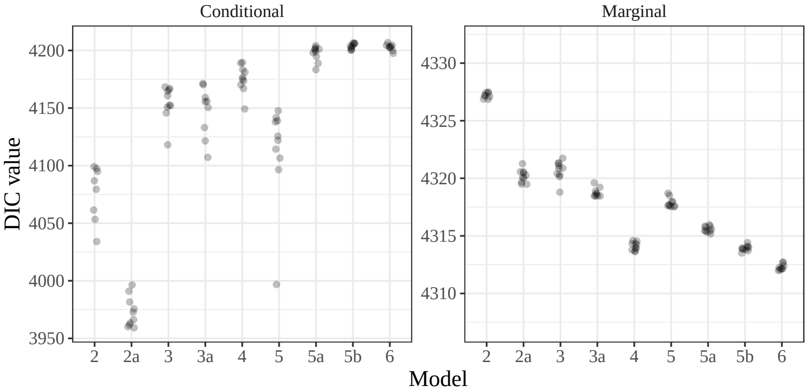
- Another issue: many (often overlooked) ways to compute information criteria in Bayesian SEM. (Merkle, Furr, Rabe-Hesketh, under review)
  - MCMC algorithms typically sample the latent variables, which implicitly counts them as parameters in a "conditional" likelihood.
  - But traditional applications of SEM integrate out latent variables, yielding "marginal" metrics that focus on generalization to new people.
  - Further, JAGS and BUGS use different equations for DIC, so they will seldom agree exactly.

Information criteria

- Interpretations of Bayesian criteria
  - Conditional: Ability of model to generalize to new data from the same individuals/cases. ("Leave one unit out" cross-validation)
  - Marginal: Ability of model to generalize to new data from new individuals/cases. ("Leave one cluster out" cross-validation)
  - In most SEM applications, marginal is preferable. But this is typically not what we would automatically obtain from BUGS/JAGS/Stan.

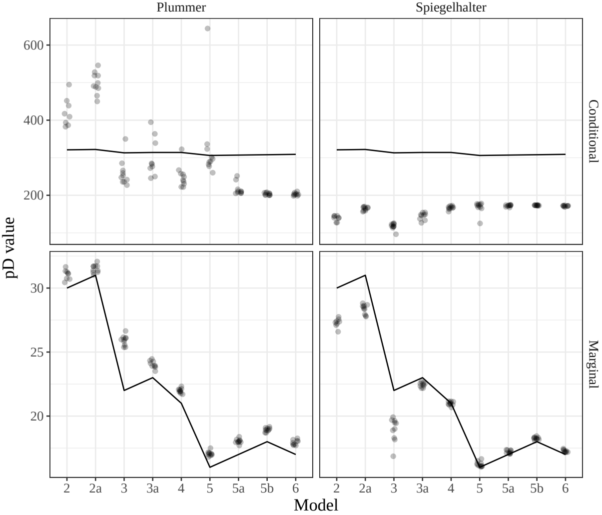
Information criteria

- DIC computations for nine CFA models (10 replications each; models from Wicherts et al., 2005)



Information criteria

- Effective number of parameters for nine CFA models



Conclusions

- DIC (also WAIC, LOO-CV) values/conclusions depend on conditional vs marginal likelihood, with marginal being preferred.
- The metrics have large Monte Carlo error (larger than individual parameters), so long chains are required to obtain stable values.
- DIC values/conclusions differ from BUGS to JAGS.

Conclusions

- For traditional SEMs, JAGS and Stan often perform similarly.
- For complex models (of primary interest for Bayesian SEM?), Stan is more likely to converge in fewer iterations and without thinning. But it is also slow.
- All computations described here are implemented in *blavaan*.

Resources

- Merkle, E. C. & Rosseel, Y. (in press). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*.

- Merkle, E. C., Furr, D., & Rabe-Hesketh, S. (under review). Bayesian model assessment: Use of conditional vs marginal likelihoods. `https://arxiv.org/abs/1802.04452`

- `http://faculty.missouri.edu/~merklee/blavaan/`

- `install.packages("blavaan")`

Thanks

Questions