The Impact of Bayesian Priors on Specification Search of Structural Equation Modeling

Xinya Liang, University of Arkansas
Yanyun Yang, Florida State University

May 25, 2016
Modern Modeling Methods Conference
Storrs, CT

Outline

- Introduction to specification search methods
- Purposes of study
- Methods
- Results
- Discussion
- Future directions

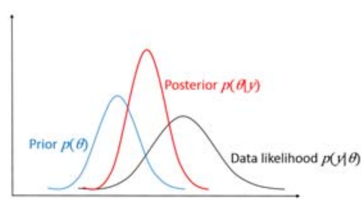
Specification Search

- Theoretical models are imperfect to varying extent (Box, 1979).
- Purposes of specification search:
  - improve model fit
  - reduce specification errors
  - elicit a model best representing the population model
  - provide meaningful interpretations for data
- A conventional search method is to use modification index (MI; Sörbom, 1989).

3

Bayesian Methods

- Bayesian Estimation
  - Posterior ∝ Prior x Likelihood
    $p(\theta|y) \propto p(\theta)p(y|\theta)$
  - Markov chain Monte Carlo (MCMC)



- Priors: non-informative, weakly informative, or informative.
- Model fit: posterior predictive checking
  - Posterior predictive *p*-value (*PPP*; if < .05, poor fit):
    $$PPP = P\left(f\left(x, \hat{\theta}_i\right) < f\left(x^{rep}, \hat{\theta}_i\right)\right)$$

4

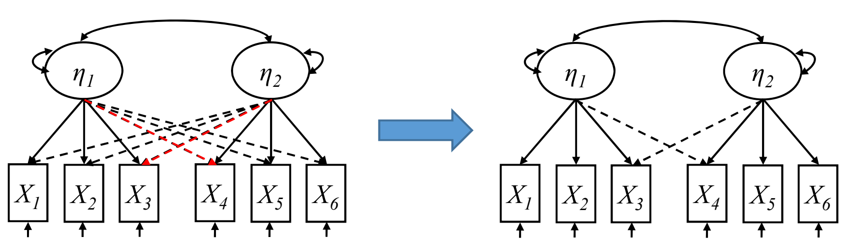
Bayesian Structural Equation Modeling (BSEM; Muthén & Asparouhov, 2012)

- BSEM starts with an over-parameterized model, and use a backward search method.
- BSEM specifies informative priors with small variances on parameters that are nearly 0 but believed not to be exactly 0.
  - i.e., cross-loadings, correlated errors, etc.
- These parameters are suggested to be freely estimated if the Bayesian credibility interval from the parameter posterior does not cover 0.
  - i.e., parameter estimates are significant.
- BSEM can be used as a search method.

5

BSEM – with Cross-Loadings

- Informative priors are typically applied to cross-loadings.
- For other parameters believed to be nonzero, such as primary factor loadings, factor and residual covariance matrices, non-informative or weakly informative priors can be imposed.



Initial model | Population model

6

Features of BSEM

- Advantages of BSEM
  - Incorporates prior knowledge into the posterior estimation
  - Does not dependent on the asymptotic theory or multivariate normality assumption
  - Handles under-identified models
  - Provides multiple suggestions in one analysis
- Disadvantage of BSEM
  - Potentially a long running time

7

Purpose of Research

- In practice, factor structures with cross-loadings are common, mainly due to:
  - Random errors in items
  - Items measuring more than one latent construct
- Constraining small cross-loadings to be 0 may result in inflated factor correlation estimates, and more.
- A critical step in BSEM analyses is to make a good selection of priors through sensitivity analyses (Asparouhov, Muthén, & Morin, 2015).

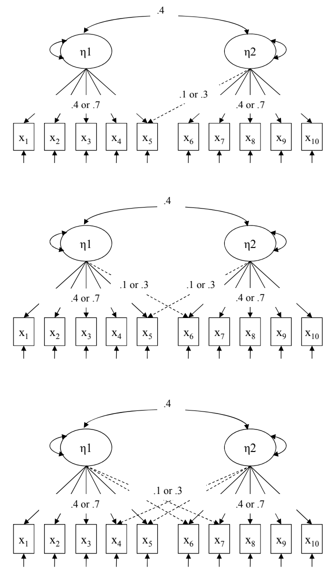
- This study aims to investigate the impact of prior distributions in specification search of small cross-loadings in confirmatory factor analysis (CFA).

8

Methods

- # conditions: 84
- # replications: 2000
- Design factors

3 *Factor structures*	Two-factor models with 1, 2, and 4 cross-loadings
4 *loading specifications*	Primary loading: .4 or .7 Cross-loading: .1 or .3
7 *Sample sizes*	50, 100, 200, 400, 600, 800, 1000



9

Data Analysis

- Each dataset was fit to a two-factor model with primary loadings and all possible cross-loadings.
- If cross-loadings were significant at the .05 level, they were retained in the model.
- The model was then re-specified and compared with the data-generating model.
- Estimation:
  - Gibbs sampler
  - Two MCMC chains
  - Medians as parameter estimates
  - predictive checking procedure for model fit

10

Prior Distributions

- All possible cross-loadings

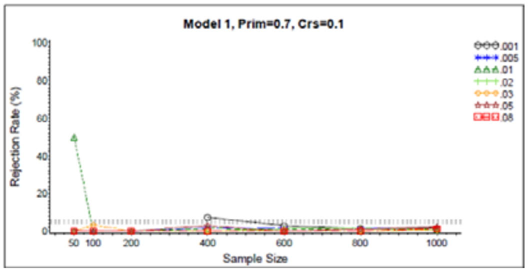
Prior	95% CI
N(0, .001)	±.06
N(0, .005)	±.14
N(0, .01)	±.20
N(0, .02)	±.28
N(0, .03)	±.34
N(0, .05)	±.44
N(0, .08)	±.55

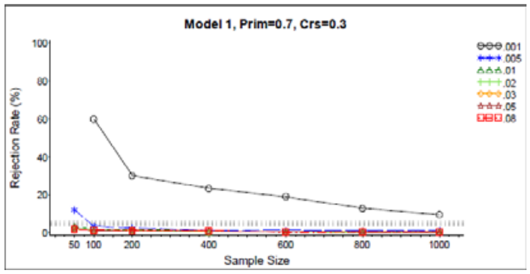
- primary loadings: $N(0, 10^{10})$
- factor covariance: $IW(0, -3)$
- residual variances: $IG(-1, 0)$

Evaluation

- Model fit
  - *PPP* rejection rates at the .05 level
- Model recovery
  - Model recovery rate: proportion of replications successfully recovered the population models.
  - Solution positive rate: proportion of replications that recovered the population model and extra parameters.
  - 95% coverage rate: proportion of replications where the 95% credibility interval covers the population value.

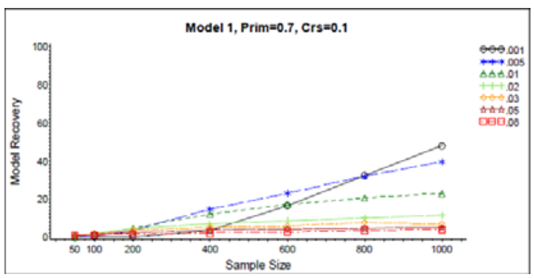
Rejection Rates

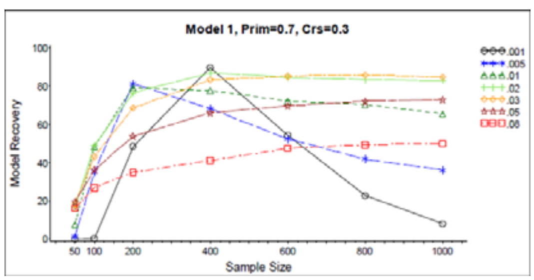




13

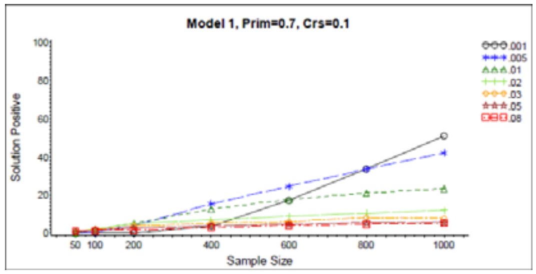
Model Recovery Rates

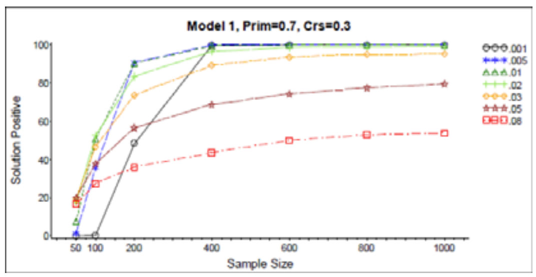




14

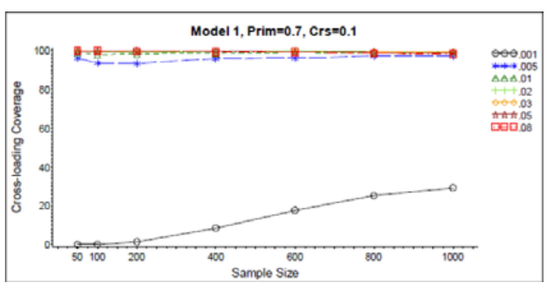
Solution Positive Rates

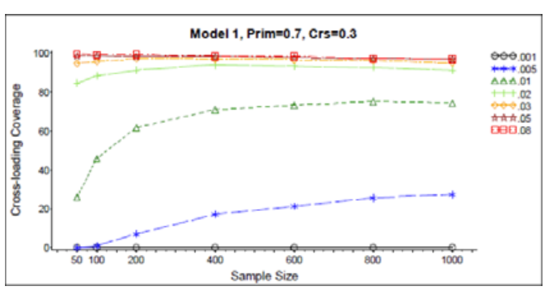




15

Coverage for Cross-Loadings





16

Discussion

- Better model recovery was associated with a simpler model, larger cross-loading value, and greater sample size.
- Model recovery was the best, when the prior CI was roughly around the population cross-loading value.
  - > 80% for Models 1 and 2, and > 60% for Model 3.
- Impact of priors:
  - Too small variance -> more false positive recovery
  - Too large variance -> low model recovery
- Recommendation:
  - Consider an informative prior with the CI boundaries close to the true value.
  - Consider BSEM search results along with MI suggestions.
  - Examine sensitivity of search results to prior distributions.

17

Future Directions

- Weakly informative priors on primary loadings and covariance matrices.
- Categorical and non-normally distributed data.
- Subsequent Bayesian searches.
- Model comparison.

18

Thank you!
Questions or Comments?

Xinya Liang (xl014@uark.edu)
Yanyun Yang (yyang3@fsu.edu)