

Small Sample Bayesian Factor Analysis

PhUSE 2014

Paper SP03

Dirk Heerwegh

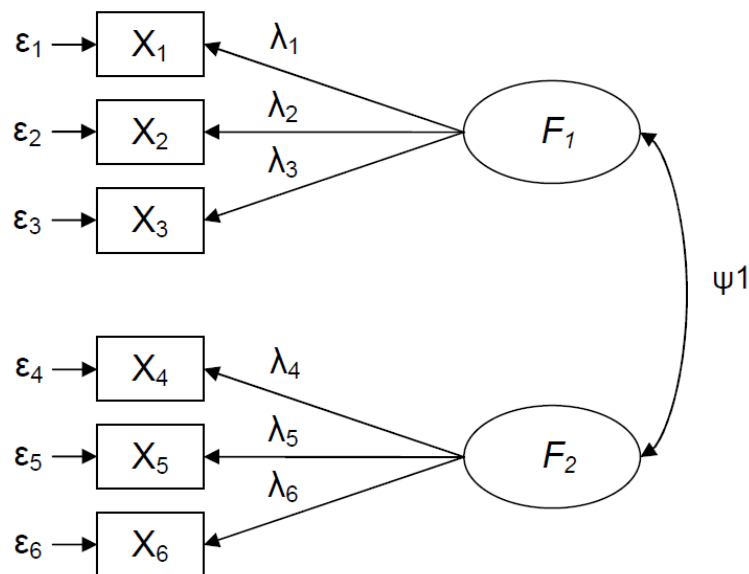


Overview

- Factor analysis
 - Maximum likelihood
 - Bayes
- Simulation Studies
 - Design
 - Results
- Conclusions

Factor Analysis (FA)

- Explain correlation between observed variables based on common causes (factors)
- Exploratory vs. Confirmatory FA



Factor Analysis (FA)

- Maximum likelihood estimation (MLE)
 - Large sample technique (e.g. $N > 200$)
 - If N is too low:
 - Model non-convergence
 - Negative residuals (Heywood cases)
 - Other problems? (statistical power, accuracy of estimates)

Bayesian Statistics

- Does not rely on large sample theory
 - Better performance in small samples expected
 - But prior distribution more influential in smaller samples
- Incorporation of prior knowledge
 - *Before* seeing the data, based on theory, previous studies
 - Captured in prior distribution

Simulation studies

- General research questions
 - How does Bayesian CFA compare to ML CFA at moderate to very small sample sizes?
 - What is the influence of priors in Bayesian CFA?
 - Non-informative versus informative priors
 - Correctly versus incorrectly specified priors
- Method
 - Monte Carlo simulations in Mplus 7.0

Simulation study 1

- 1-factor CFA model with 4 observed variables
 - 48 conditions in the simulation study:
 - Sample sizes: 200, 100, 50, 25
 - Standardized factor loadings: .80, .60, .40
 - ML or BAYES estimator
 - If BAYES: Prior distribution for factor loadings:
 - Non-informative (mean = 0, variance = infinity)
 - Informative: mean = 0.40 and variance = .05 or .01
 - Note that the prior is misspecified for the conditions in which the true factor loadings are .80 and .60

TITLE: BAYES CFA Monte Carlo Simulation study for Phuse

! SIMULATION PARAMETERS

! -----

! Condition 3

! Estimator: BAYES

! Sample size: 200

! Lambda: .8

! Residual variance: 0.36

! Prior for the lambdas:

! Normal distribution with mean = 0.4 and variance = 0.05

MONTECARLO:

NAMES ARE x1-x4;

NOBSERVATIONS = 200;

NREPS=1000;

SEED=12345;

MODEL POPULATION:

f1 BY x1-x4*.8;

x1-x4*0.36;

f1@1;

ANALYSIS:

estimator = bayes;

proc = 2;

fbiter = 10000;

MODEL:

f1 BY x1-x4*.8 (a1-a4);

x1-x4*0.36;

f1@1;

model priors:

a1-a4 ~ N(0.4,0.05);

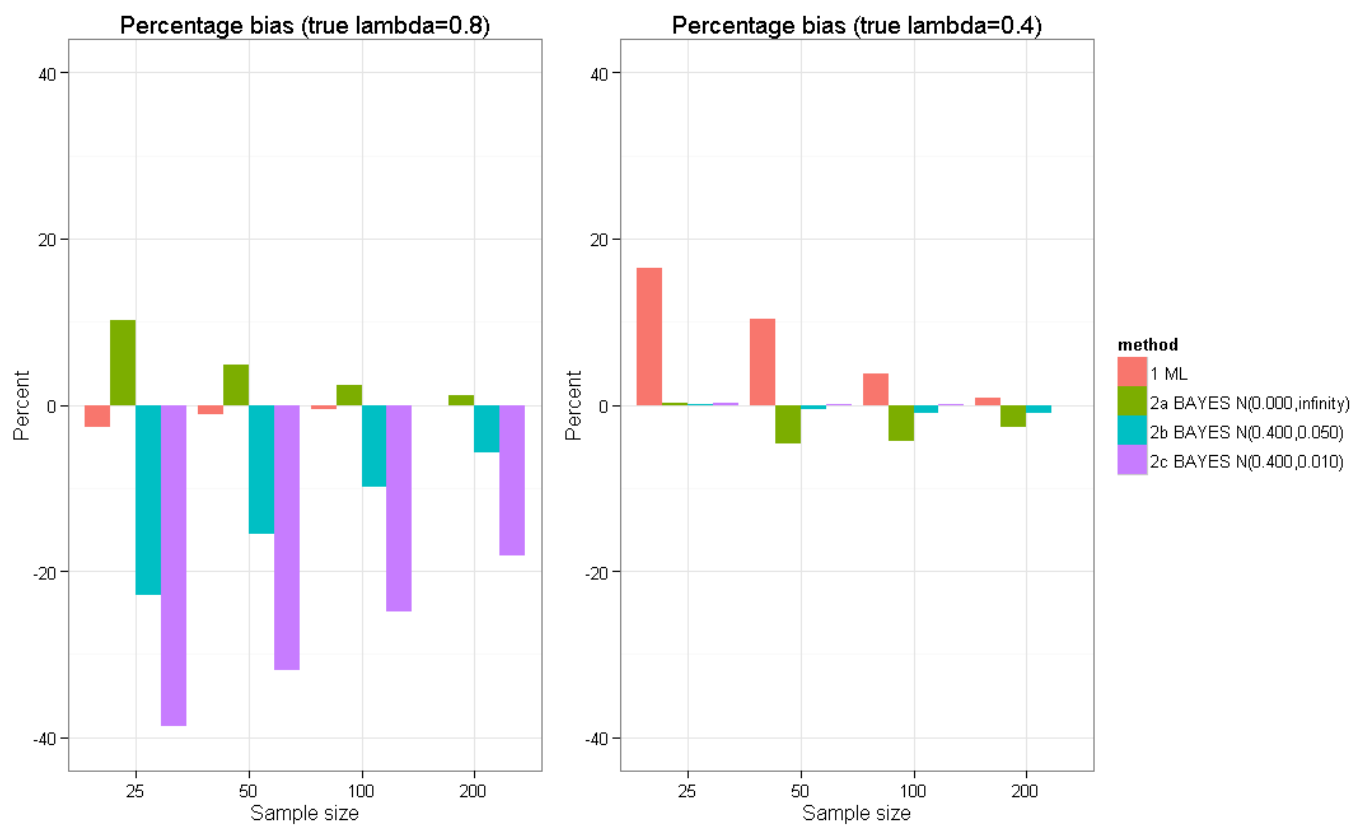
OUTPUT:

TECH9;

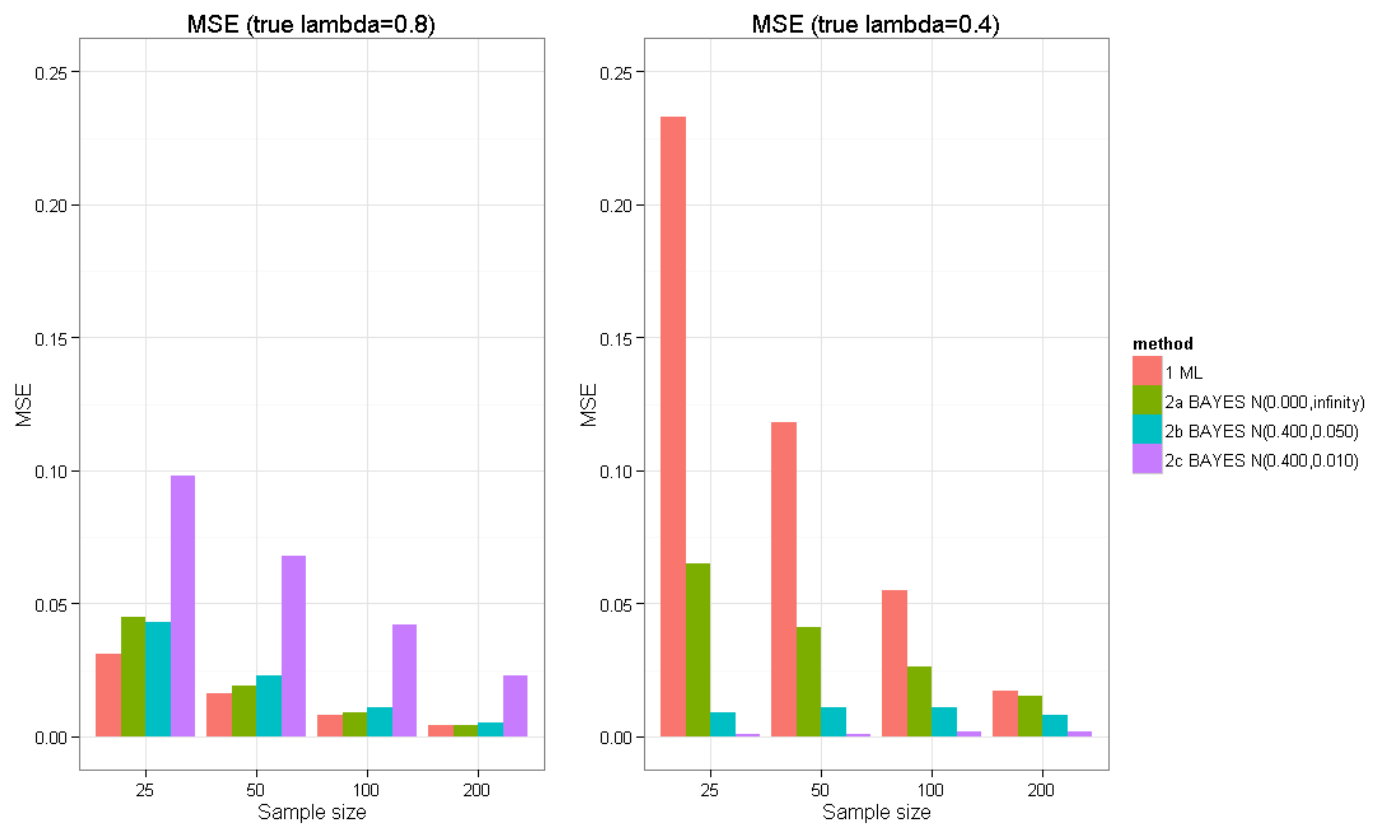
Simulation study 1

- Criteria:
 - Percentage parameter bias:
 - $100 \times (\text{mean estimated value} - \text{popul. value}) / \text{popul. value}$
 - Averaged over the 4 observed variables in the simulation study.
 - MSE (Mean Squared Error)
 - MSE is equal to the variance of the estimates across the replications plus the square of the bias.
 - Averaged over the 4 observed variables in the simulation study.
 - Statistical power to detect significant factor loading
 - Non-convergence of the models or errors in the model estimation (residual covariance matrix not positive definite, untrustworthy standard errors).

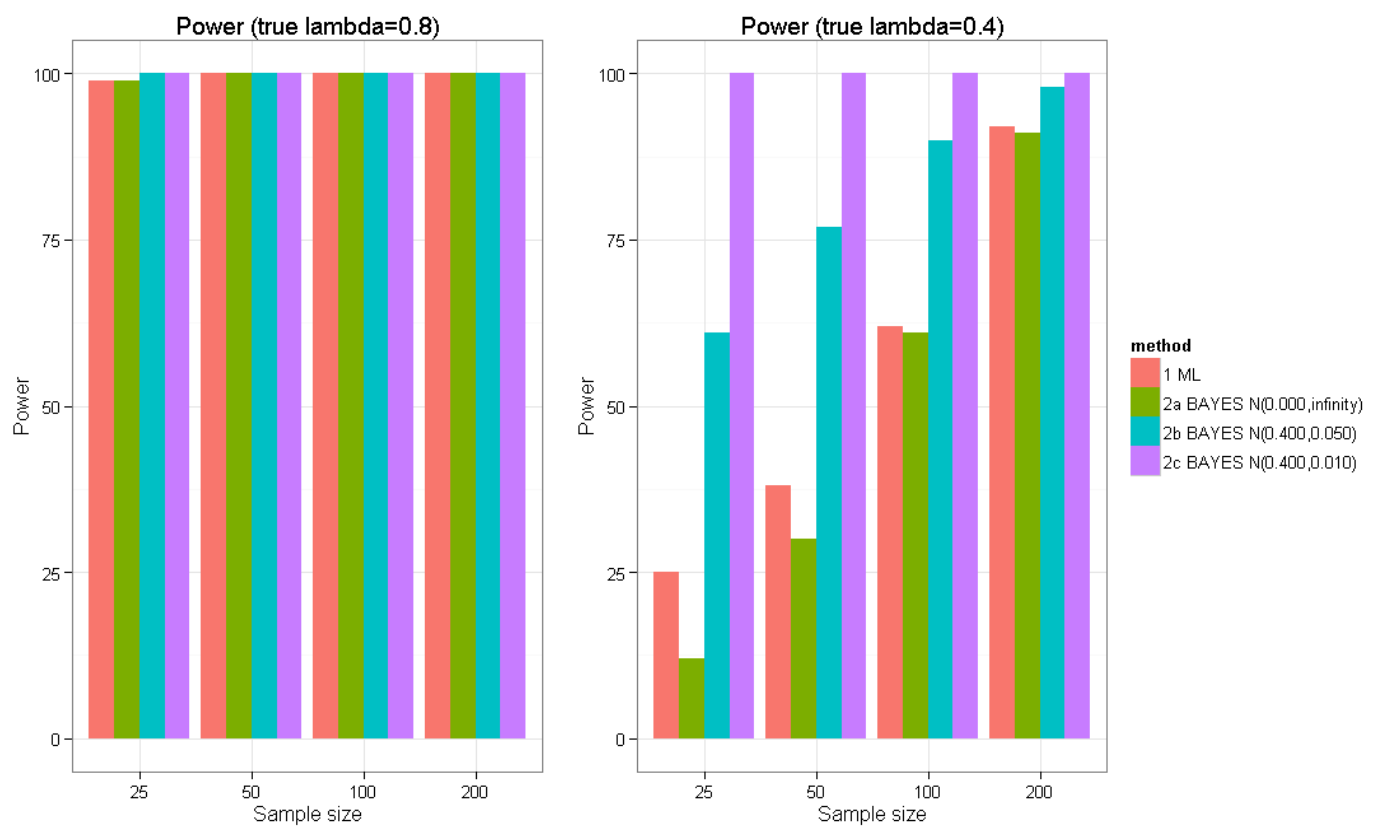
Results: % parameter bias



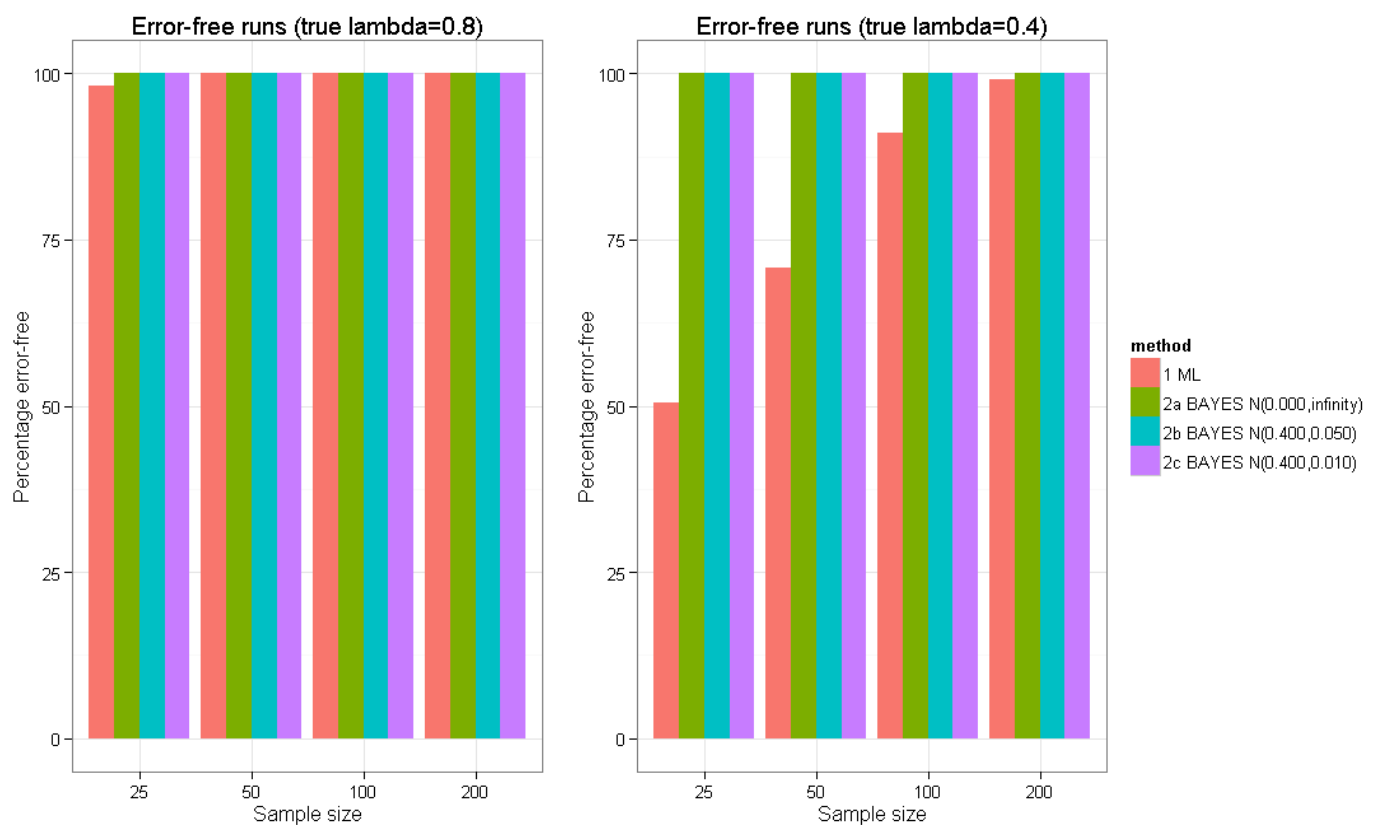
Results: MSE



Results: Power



Results on error-free runs



Conclusions study 1

- Main driving force: factor loadings
 - ML CFA performs well at high factor loadings
 - Few model convergence errors, high power, low bias
 - As factor loadings become smaller, Bayesian CFA becomes more appealing
 - Fewer model convergence problems, less biased estimates (non-informative or correctly specified informative priors)
 - But power is slightly lower than with MLE if using non-informative priors.
- Miss-specified priors bias the results as we would expect

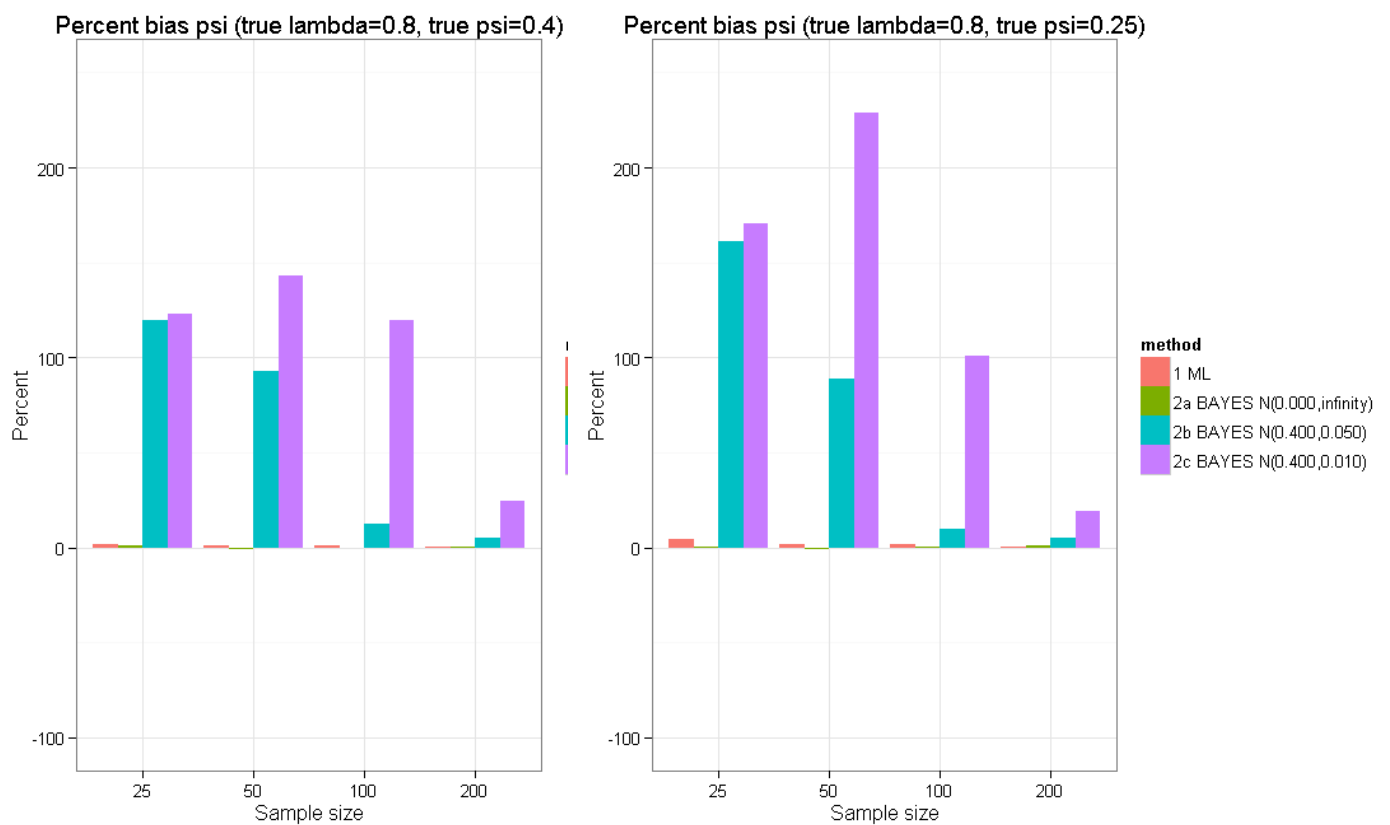
Simulation study 2

- 2-factor CFA model with 3 observed variables each and no cross-loadings (“simple structure”)
 - 96 conditions in the simulation study:
 - Sample sizes: 200, 100, 50, 25
 - Standardized factor loadings: .8, .6, .4
 - Factor correlations of .25 and .40
 - ML or BAYES estimator
 - If BAYES: Prior distribution for factor loadings:
 - Non-informative (mean = 0, variance = infinity)
 - Informative: mean = 0.4 and variance = .05 or .01
 - Note that the prior is miss specified for the conditions in which the true factor loadings are .8 and .6
 - If BAYES: non-informative prior on factor covariance
 - Method: Monte Carlo simulations in Mplus 7

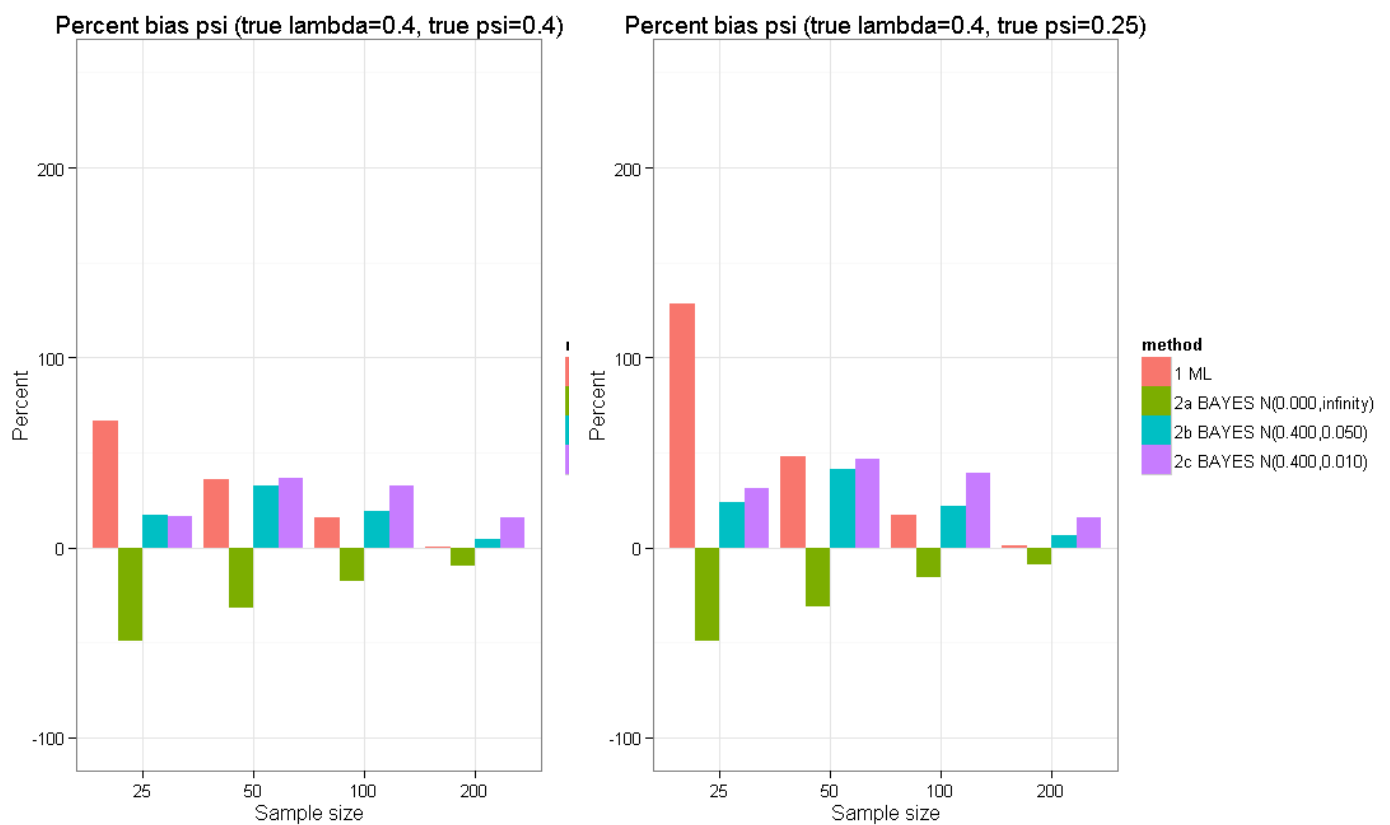
Results

- See paper for full description
 - Focus here on bias on factor covariance
 - Recall: non-informative prior used for factor covariance

Results: % bias on factor covariance



Results: % bias on factor covariance



Conclusions study 2 (selection)

- Miss specified priors on factor loadings lead to bias in estimated factor covariance
 - In a CFA model, all parameters work together in a system
 - Introducing an error in one part of the model will influence other parts of the model
 - Important if you're working with complex models!

In conclusion

- Bayesian CFA certainly delivers on some aspects:
 - It can be used in (very) small samples / measurement models with low factor loadings, where ML CFA either cannot estimate model parameters or does so with bias
 - But it is important to judiciously select the prior. This should be done based on theory, but keep in mind the possibility of spill-over effects!
- Further (simulation) studies are needed to gain a fuller understanding of Bayesian CFA models as well as Bayesian SEM models

In conclusion

- Additional appeals of Bayesian CFA:
 - Near-zero cross-loadings and/or error covariances
 - Not possible in ML CFA because of model non-identification
 - Benefits:
 - May be more in line with theory, improves model fit, can aid in recovering a theorized simple structure

See reference in paper to a study on the Hospital Anxiety and Depression Scale.