Bayesian regularized SEM

What, Why, When, and How?

Sara van Erp
s.j.vanerp@uu.nl

SEM Conference 11-03-2022
https://github.com/sara-vanerp/bayesregsem

Structural equation modeling

Aim: finding a good enough description of the phenomenon under investigation that is as parsimonious as possible.

Independence ⟷ **Saturated**

Restrictive:
- all parameters fixed to 0
- high bias/low variance

Free:
- all parameters estimated
- low bias/high variance

The problem

Traditional confirmatory approach relies on theory to restrict parameters.

If fit is not acceptable, modifications can be made that shift the model from restrictive to free.

Model modifications

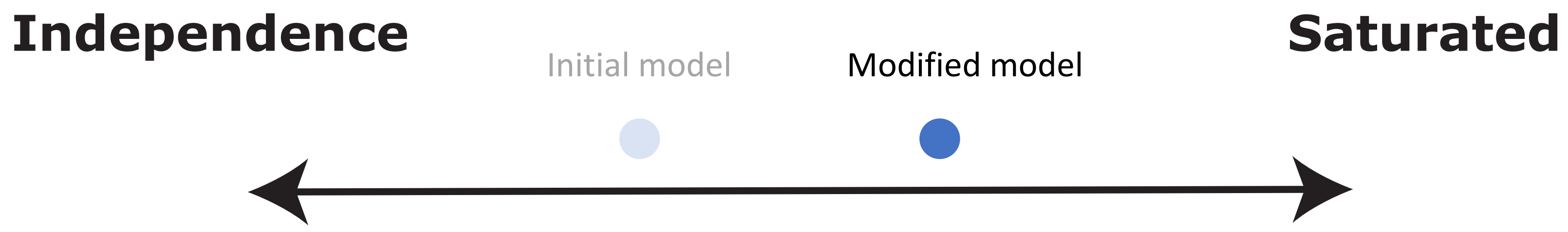
Independence

Saturated

Restrictive:
- all parameters fixed to 0
- high bias/low variance

Free:
- all parameters estimated
- low bias/high variance

Model modifications



Restrictive:
- all parameters fixed to 0
- high bias/low variance

Free:
- all parameters estimated
- low bias/high variance

The problem

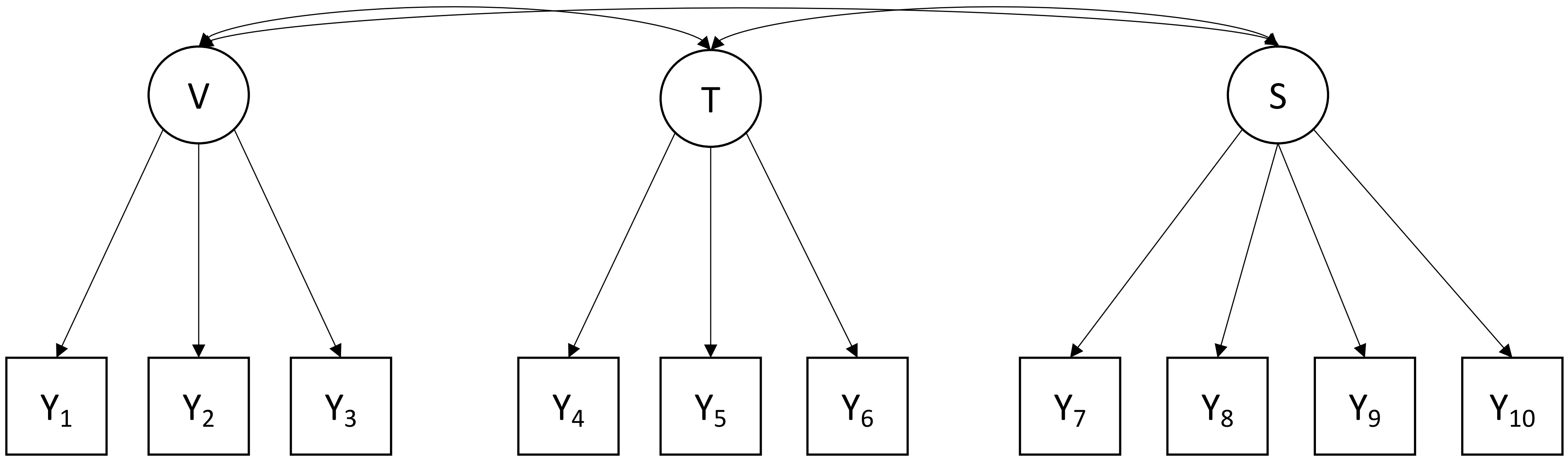
Traditional confirmatory approach relies on theory to restrict parameters.

If fit is not acceptable, modifications can be made that shift the model from restrictive to free.

Problem: reduces bias, at the cost of increased variance.

See also MacCallum et al. (1992)

An illustration



An illustration

Generated data (*N* = 5000) based on Holzinger & Swineford (1939)

Added one item uncorrelated to the others

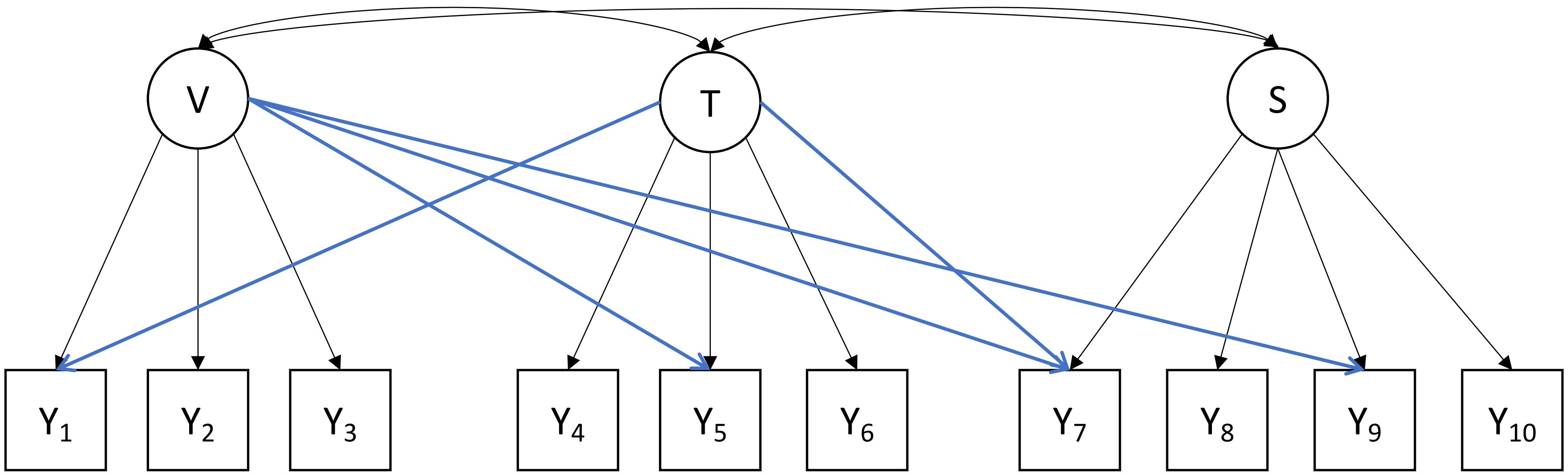
Sampled a training (*N* = 300) and test (*N* = 300) set

Fit the model and free cross-loadings until fit is “good”

See also https://github.com/sara-vanerp/bayesregsem

An illustration

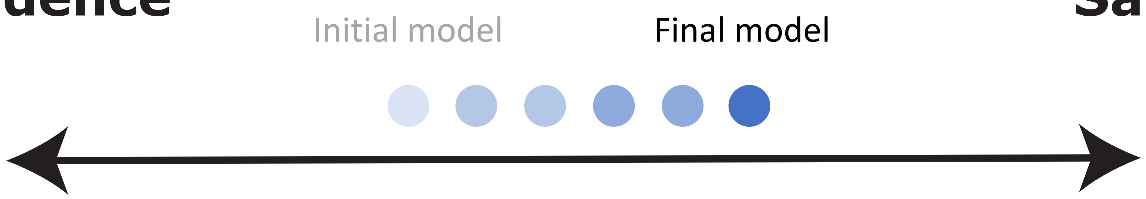
Model	*p*	*CFI*	*TLI*	*RMSEA*	*SRMR*
Original (training set)	< .001	0.88	0.83	0.11	0.074
Adapted (training set)	< .001	0.95	0.92	0.075	0.045
Adapted (test set)	< .001	0.93	0.89	0.086	0.044



Model modifications

Independence

Saturated



Restrictive:
- all parameters fixed to 0
- high bias/low variance

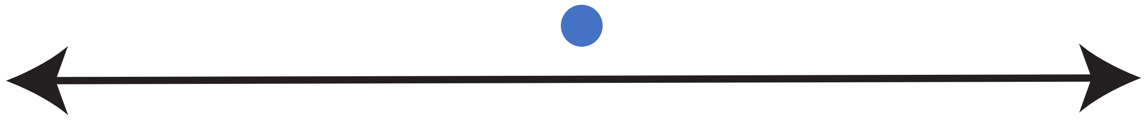
Free:
- all parameters estimated
- low bias/high variance

An alternative

Independence

Final model

Saturated



Restrictive:
- all parameters fixed to 0
- high bias/low variance

Free:
- all parameters estimated
- low bias/high variance

Regularized SEM

Idea: add a penalty to the parameters to regularize

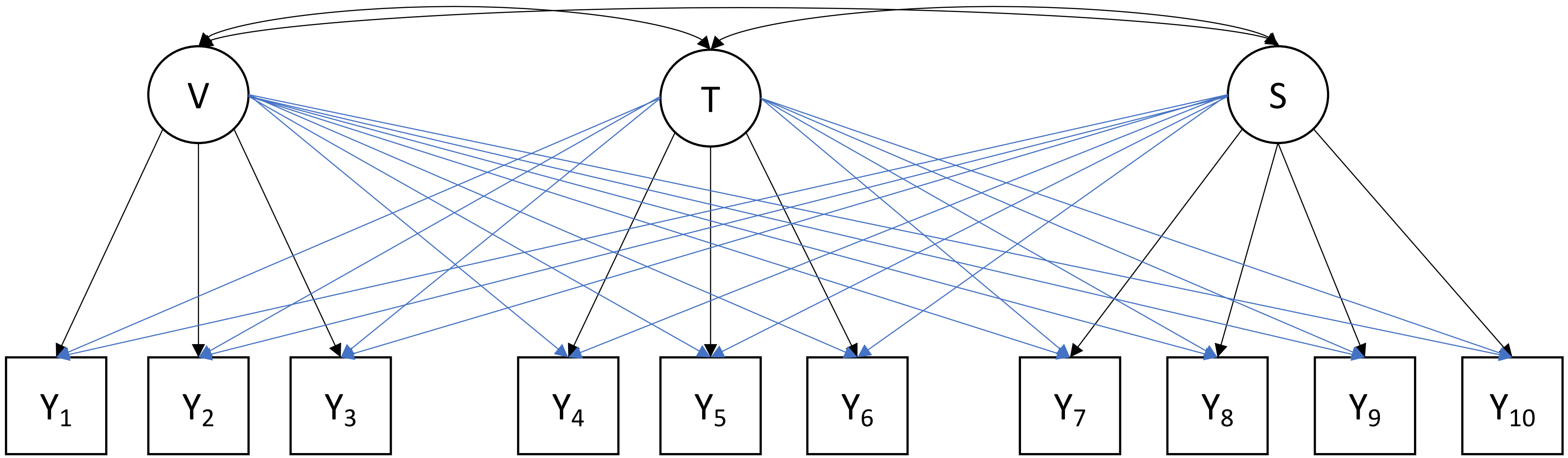
$$F_{regsem}\big(S, \Sigma(\theta)\big) = F\big(S, \Sigma(\theta)\big) + \lambda P\big(\theta_{reg}\big)$$

e.g. lasso, ridge, or elastic net.

See also Jacobucci et al. (2016)

Regularized SEM

All blue cross-loadings are penalized



Regularized SEM: possible applications

- Cross-loadings in CFA (Jacobucci et al., 2016)
- Loadings in EFA (Jin et al., 2018; Trendafilov et al., 2017; Hirose and Yamamoto, 2014)
- Selection of covariates in MIMIC models (Jacobucci et al., 2016, 2019)
- Selection of mediators in mediation analysis (Serang et al., 2017; Zhao and Luo, 2016; van Kesteren and Oberski, 2019)
- More stable estimation of latent class models in small samples (Chen et al., 2017; Robitzsch, 2020)
- Detection of violations of MI (Magis et al., 2015; Tutz and Schauberger, 2013)
- ...

Any model with (too) many parameters where some are assumed to be zero.
(Let me know if you have any other examples!)

Bayesian regularized SEM

Idea: use a prior instead of a penalty

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Prior shrinks small coefficients towards zero, some priors in the same way as classical penalties.

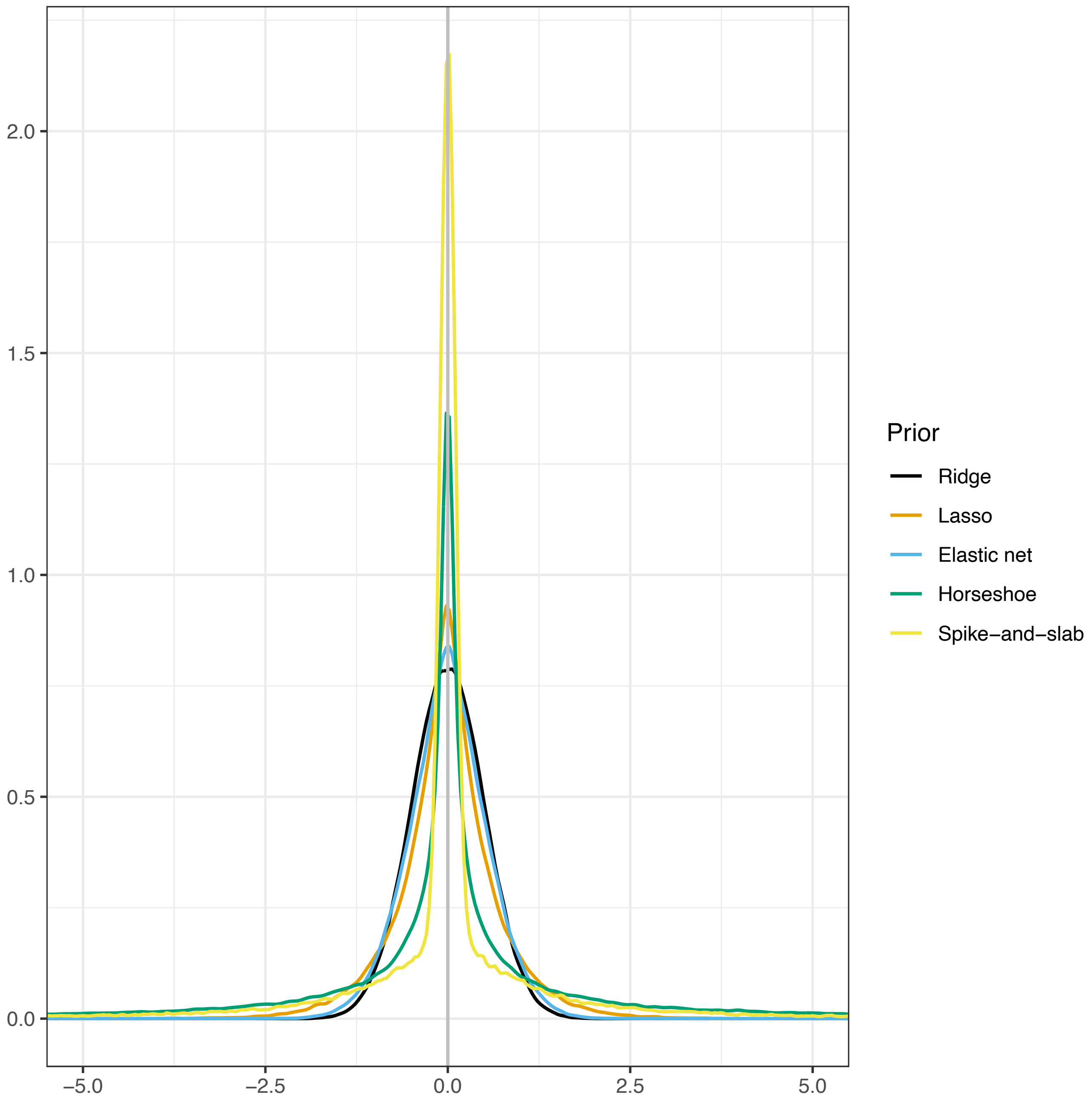
But: many priors do not correspond to a classical penalty

See also van Erp et al. (2019)

Shrinkage priors

Properties:

1. Very peaked around zero
2. Heavy tails

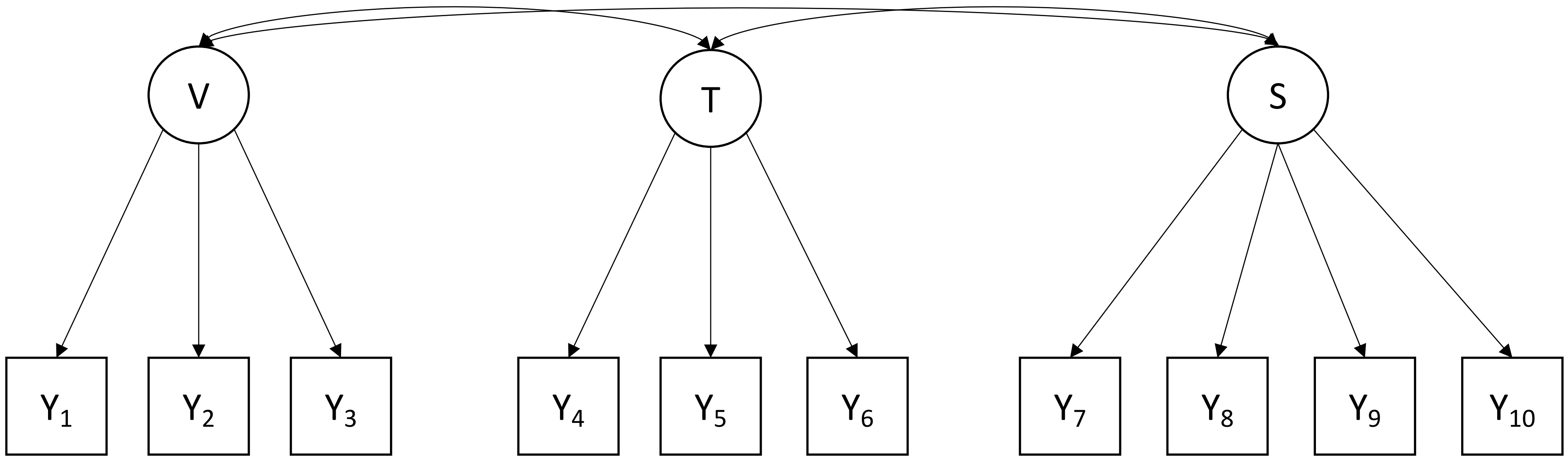


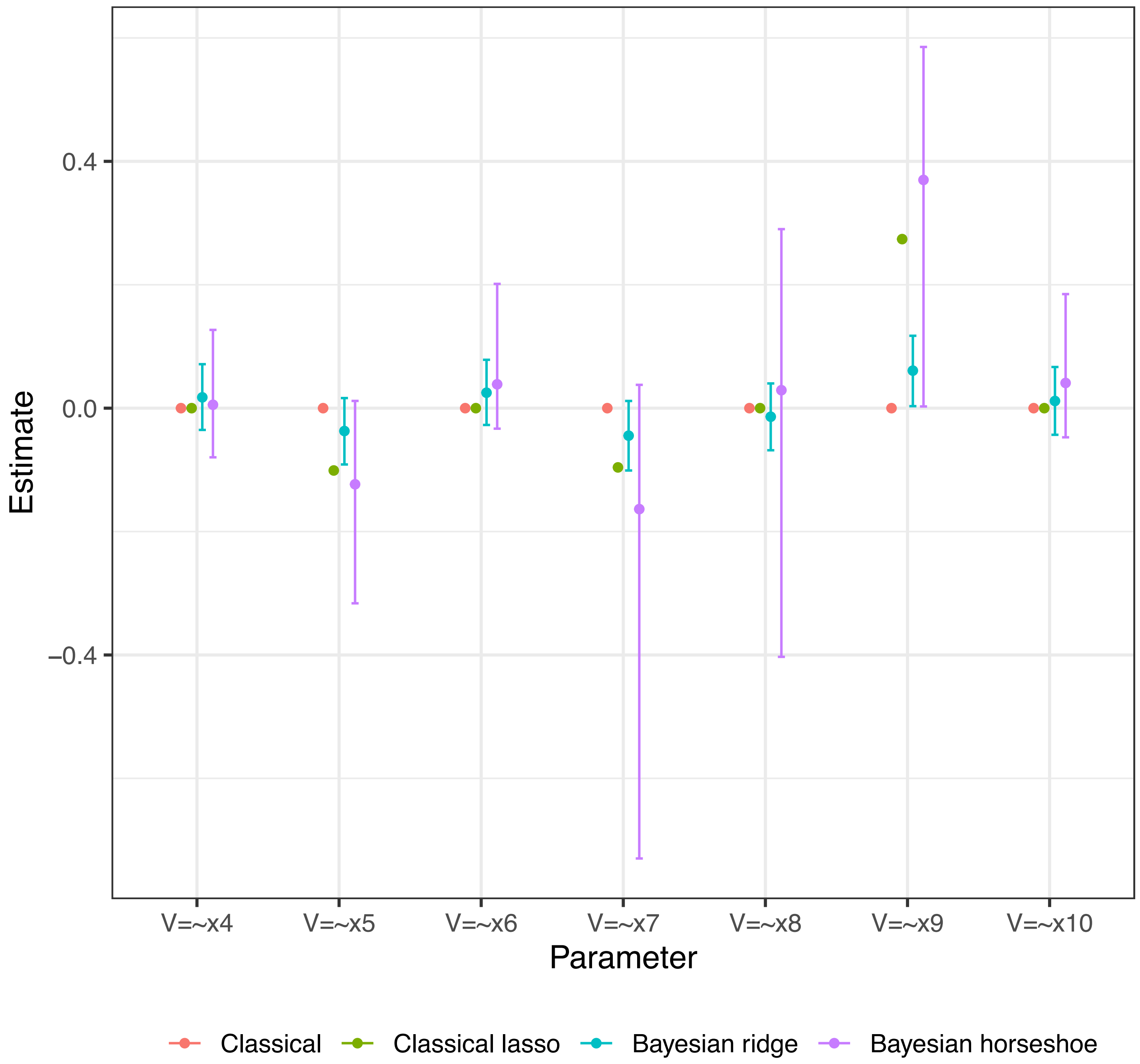
Advantages Bayesian regularized SEM

- Intuitive interpretation
- Automatic uncertainty estimates
- Incorporation of prior information

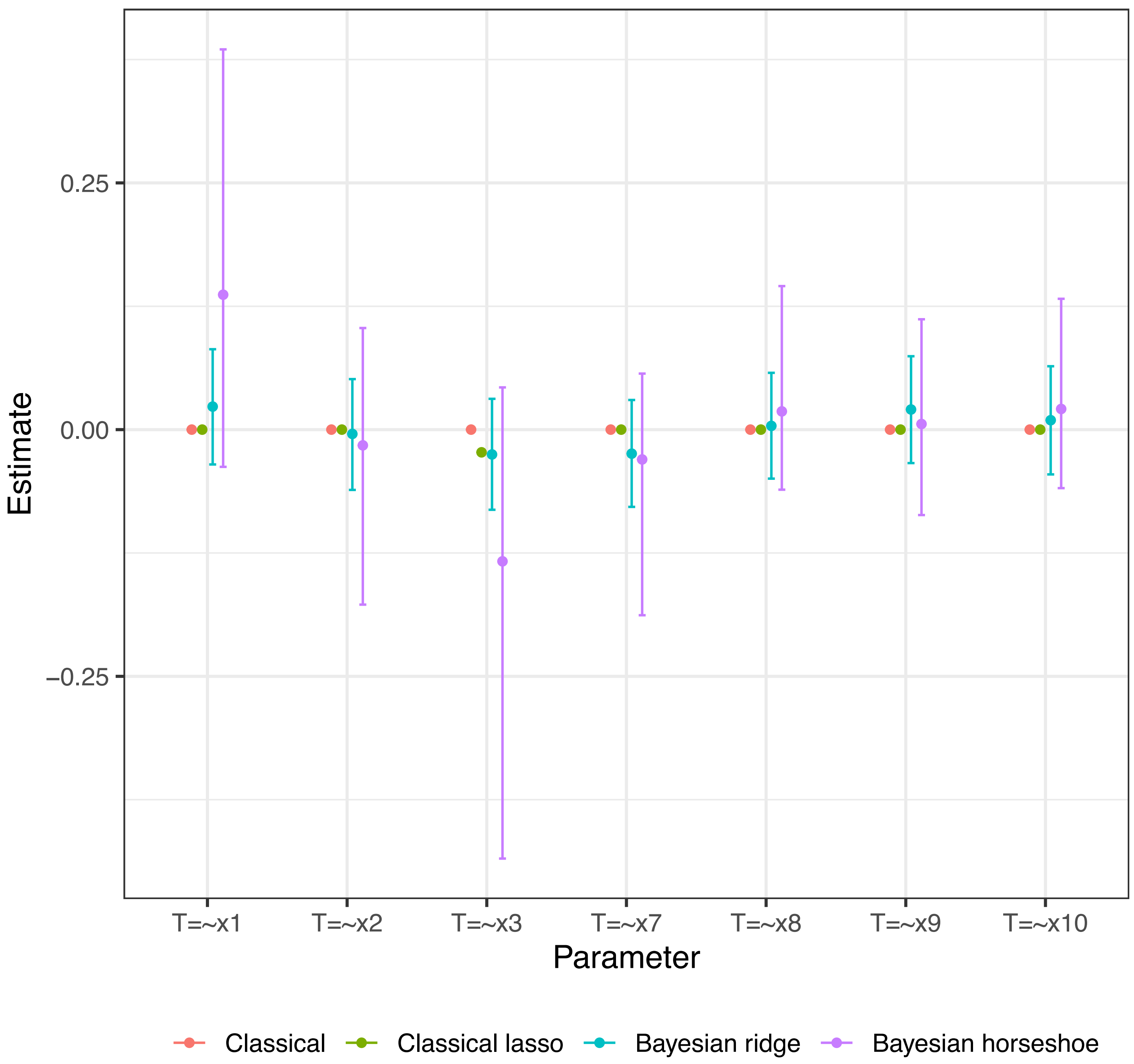
- Flexibility in terms of priors considered
- Automatic estimation penalty parameter

An illustration

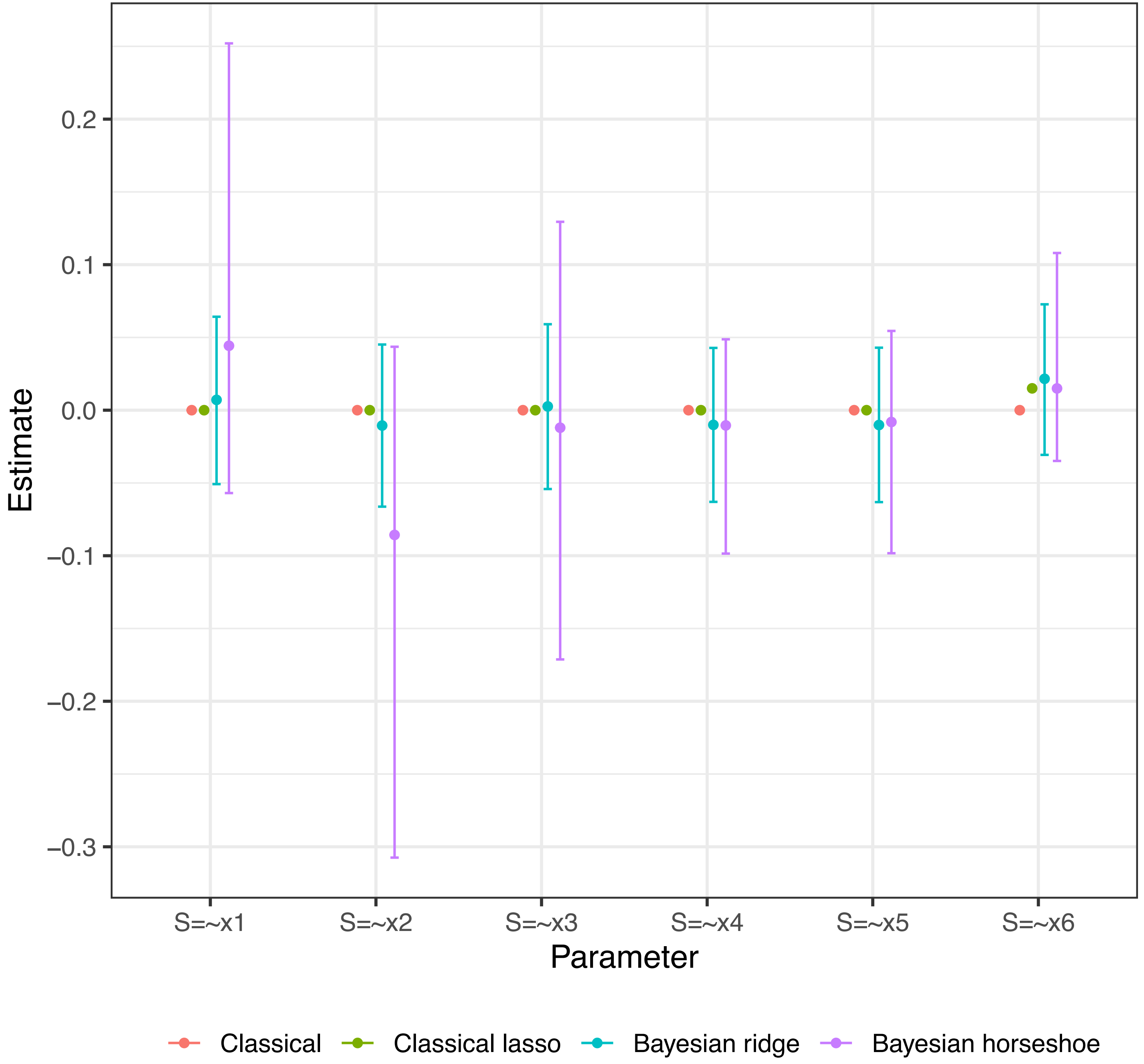




Estimate
0.4
0.0
−0.4
V=~x4
V=~x5
V=~x6
V=~x7
V=~x8
V=~x9
V=~x10
Parameter
Classical
Classical lasso
Bayesian ridge
Bayesian horseshoe



0.25
0.00
−0.25
Estimate
T=~x1
T=~x2
T=~x3
T=~x7
T=~x8
T=~x9
T=~x10
Parameter
Classical
Classical lasso
Bayesian ridge
Bayesian horseshoe



Estimate
0.2
0.1
0.0
−0.1
−0.2
−0.3
S=~x1
S=~x2
S=~x3
S=~x4
S=~x5
S=~x6
Parameter
Classical
Classical lasso
Bayesian ridge
Bayesian horseshoe

Fit to the test set

Classical lasso: free loadings with an estimate $\neq 0$ (5)

Shrinkage priors: free loadings with a 95% CI not including 0 (1)

Model	*p*	*CFI*	*TLI*	*RMSEA*	*SRMR*
Adapted (test set)	< .001	0.93	0.89	0.086	0.044
Classical lasso (test set)	< .001	0.91	0.87	0.094	0.057
Shrinkage priors (test set)	< .001	0.92	0.87	0.092	0.051

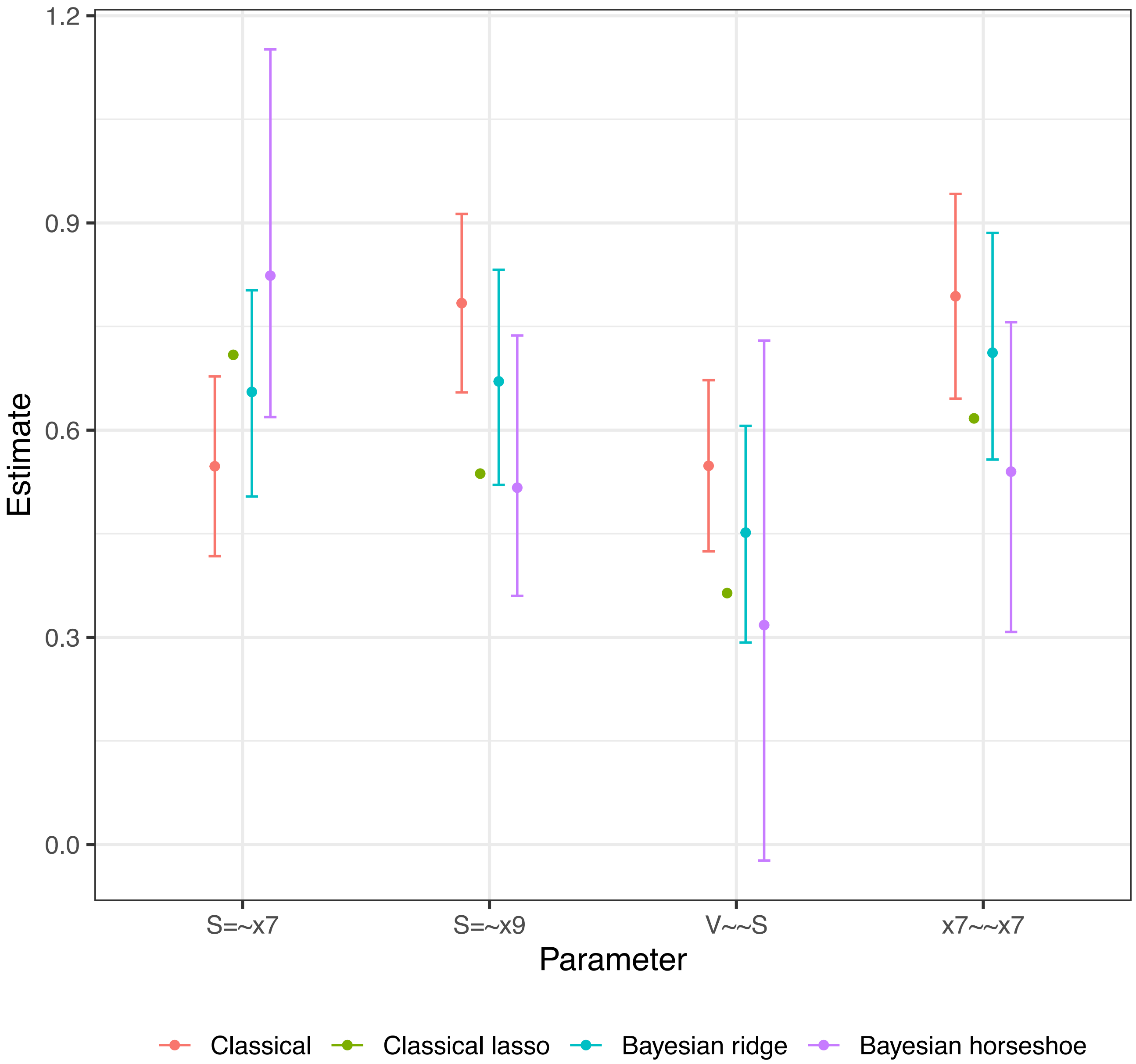
Open question: How to decide which parameters to free?

Parameter selection

- Classical penalization methods lead to estimates equal to 0 or not
- Bayesian methods require additional selection step, e.g. based on CI or threshold
- Existing methods arbitrary and rely on marginal posteriors
- More advanced methods needed (projpred?)

See also Zhang, Pan & Ip (2021)

Other parameters can be influenced as well



When to use Bayesian regularized SEM?

- Same applications as classical regularized SEM
- In theory: any non-identified model
- In practice: depends on software availability (and additional research)

Implementations of Bayesian regsem

User-friendly

- Mplus: only ridge available
- R-package LAWLB: adaptive and covariance lasso for CFA (Chen, 2021)
- R-package blcfa: covariance lasso for CFA (Zhang et al., 2021)

Not so user-friendly:

- Description MCMC algorithm or (R) implementation online
- Code it yourself (Stan or JAGS)

Added difficulty: prior should be informative enough to identify the model

Current and future projects

Goal: make Bayesian regsem available to the applied user

- **User-friendly software**
- Projection predictive inference to select parameters (Piironen & Vehtari, 2016)
- Regularized horseshoe for SEM
- How to choose the prior and its hyperparameters?
- Moving to high-dimensional settings

Contact me if you would like to collaborate!

References (See https://github.com/sara-vanerp/bayesregsem)

Chen, J. (2021). A Generalized Partially Confirmatory Factor Analysis Framework with Mixed Bayesian Lasso Methods. Multivariate Behavioral Research, 1-16.

Chen, Y., Li, X., Liu, J., and Ying, Z. (2017). Regularized latent class analysis with application in cognitive diagnosis. Psychometrika, 82(3):660–692.

Hirose, K. and Yamamoto, M. (2014). Sparse estimation via nonconcave penalized likelihood in factor analysis model. Statistics and Computing, 25(5):863–875.

Jacobucci, R., Brandmaier, A. M., and Kievit, R. A. (2019). A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. Advances in Methodsand Practices in Psychological Science, 2(1):55–76.

Jacobucci, R., Grimm, K. J., and McArdle, J. J. (2016). Regularized structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 23(4):555–566.

Jin, S., Moustaki, I., and Yang-Wallentin, F. (2018). Approximated penalized maximum likelihood for exploratory factor analysis: An orthogonal case. Psychometrika, 83(3):628–649.

MacCallum, R. C., Roznowski, M., and Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. Psychological Bulletin, 111(3):490–504.

Magis, D., Tuerlinckx, F., and Boeck, P. D. (2015). Detection of differential item functioning using the lasso approach. Journal of Educational and Behavioral Statistics, 40(2):111–135.

Piironen, J., & Vehtari, A. (2016). Comparison of Bayesian predictive methods for model selection. Statistics and Computing, 27(3), 711–735.

Robitzsch, A. (2020). Regularized latent class analysis for polytomous item responses: An application to SPM-LS data. Journal of Intelligence, 8(3):30.

Serang, S., Jacobucci, R., Brimhall, K. C., and Grimm, K. J. (2017). Exploratory mediation analysis via regularization. Structural Equation Modeling: A Multidisciplinary Journal, 24(5):733–744.

Trendafilov, N. T., Fontanella, S., and Adachi, K. (2017). Sparse exploratory factor analysis. Psychometrika, 82(3):778–794.

Tutz, G. and Schauberger, G. (2013). A penalty approach to differential item functioning in rasch models. Psychometrika, 80(1):21–43.

van Erp, S., Oberski, D. L., and Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. Journal of Mathematical Psychology, 89:31–50.

van Kesteren, E.-J. and Oberski, D. L. (2019). Exploratory mediation analysis with many potential mediators. Structural Equation Modeling: A Multidisciplinary Journal, 26(5):710–723.

Zhang, L., Pan, J., Dubé, L., & Ip, E. H. (2021). blcfa: An R package for Bayesian model modification in confirmatory factor analysis. Structural Equation Modeling: A Multidisciplinary Journal, 28(4), 649-658.

Zhang, L., Pan, J., & Ip, E. H. (2021). Criteria for Parameter Identification in Bayesian Lasso Methods for Covariance Analysis: Comparing Rules for Thresholding, p-value, and Credible Interval. Structural Equation Modeling: A Multidisciplinary Journal, 28(6), 941-950.

Zhao, Y. and Luo, X. (2016). Pathway lasso: estimate and select sparse mediation pathways with high dimensional mediators. arXiv preprint arXiv:1603.07749.