The Reliability of Multidimensional Scales: A Comparison of Confidence Intervals and a Bayesian Alternative

Julius M. Pfadt, Don van den Bergh, Morten Moshagen

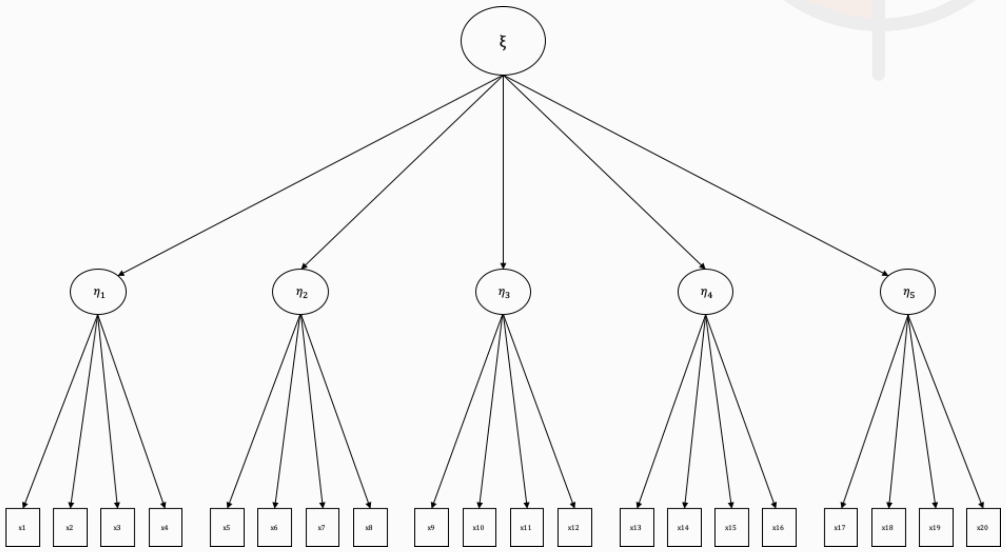
Correspondence to julius.pfadt@gmail.com

Outline

- Motivation
- Higher-Order Factor Model
- Confidence Intervals
- Bayesian Estimation
- Example
- Simulation
- Conclusion

Motivation – Reliability

- $\rho = \frac{\sigma_T^2}{\sigma_X^2}$
- For unidimensional tests: Coefficient α or coefficient ω_u
- For multidimensional tests:
    - Coefficient ω_t
    - Coefficient ω_h estimates different form of reliability: general factor saturation of a scale



Current Issues

Motivation – Current Issues

- Uncertainty estimation is neglected in reliability analysis
- Confidence intervals for ω_t and ω_h are rarely researched
- Credible intervals for ω_t and ω_h unavailable

Solution:

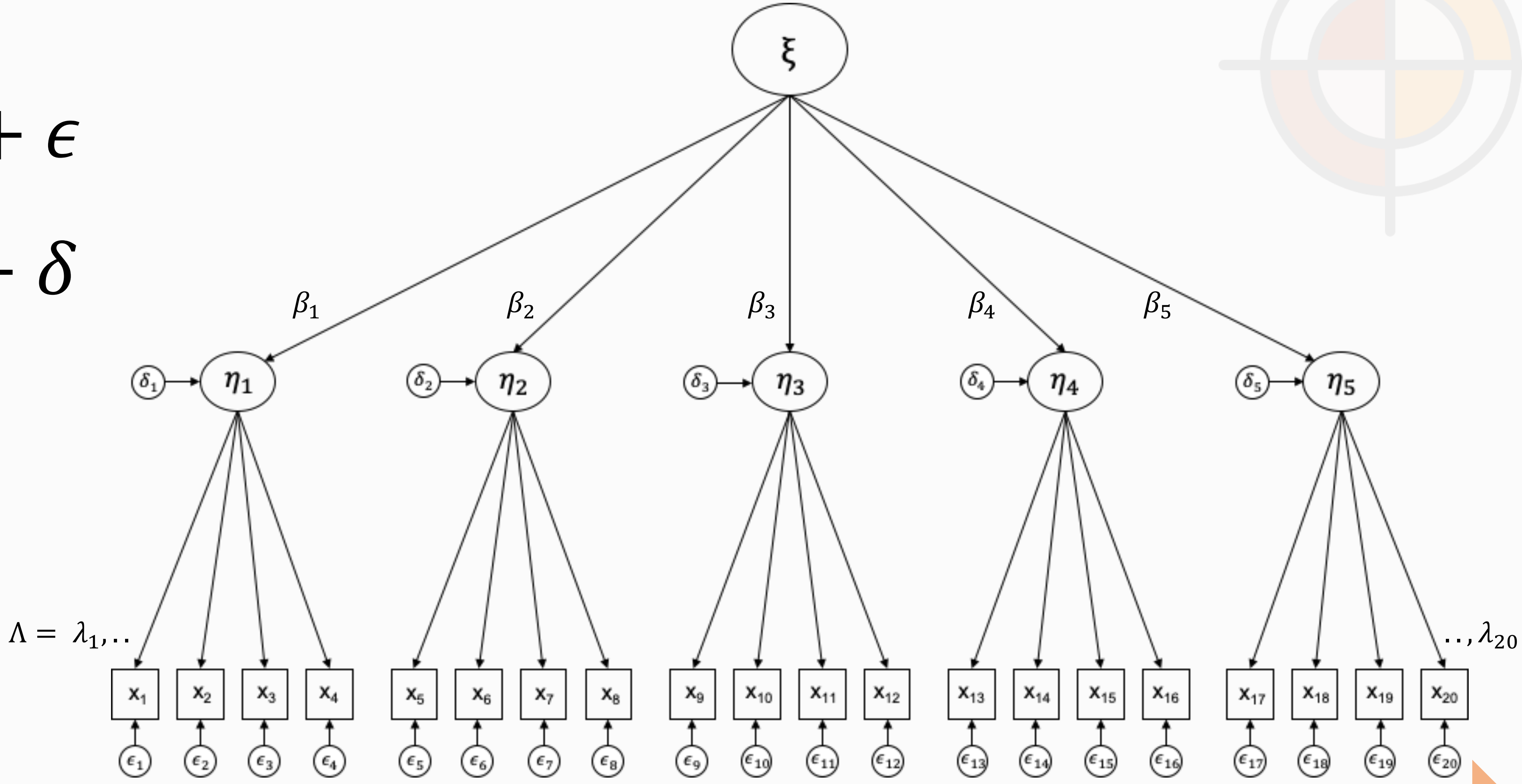
- Make all methods available through R and JASP
- Investigate six confidence intervals
- Fit the higher-order factor model in the Bayesian framework

Higher-order factor model

Higher-Order Factor Model (second-order factor model)

$$X = \Lambda \eta + \epsilon$$

$$\eta = \beta \xi + \delta$$



ω_t and ω_h

Higher-Order Factor Model – Coefficients ω_t and ω_h

Higher-order model → Bi-factor model

Schmid-Leiman

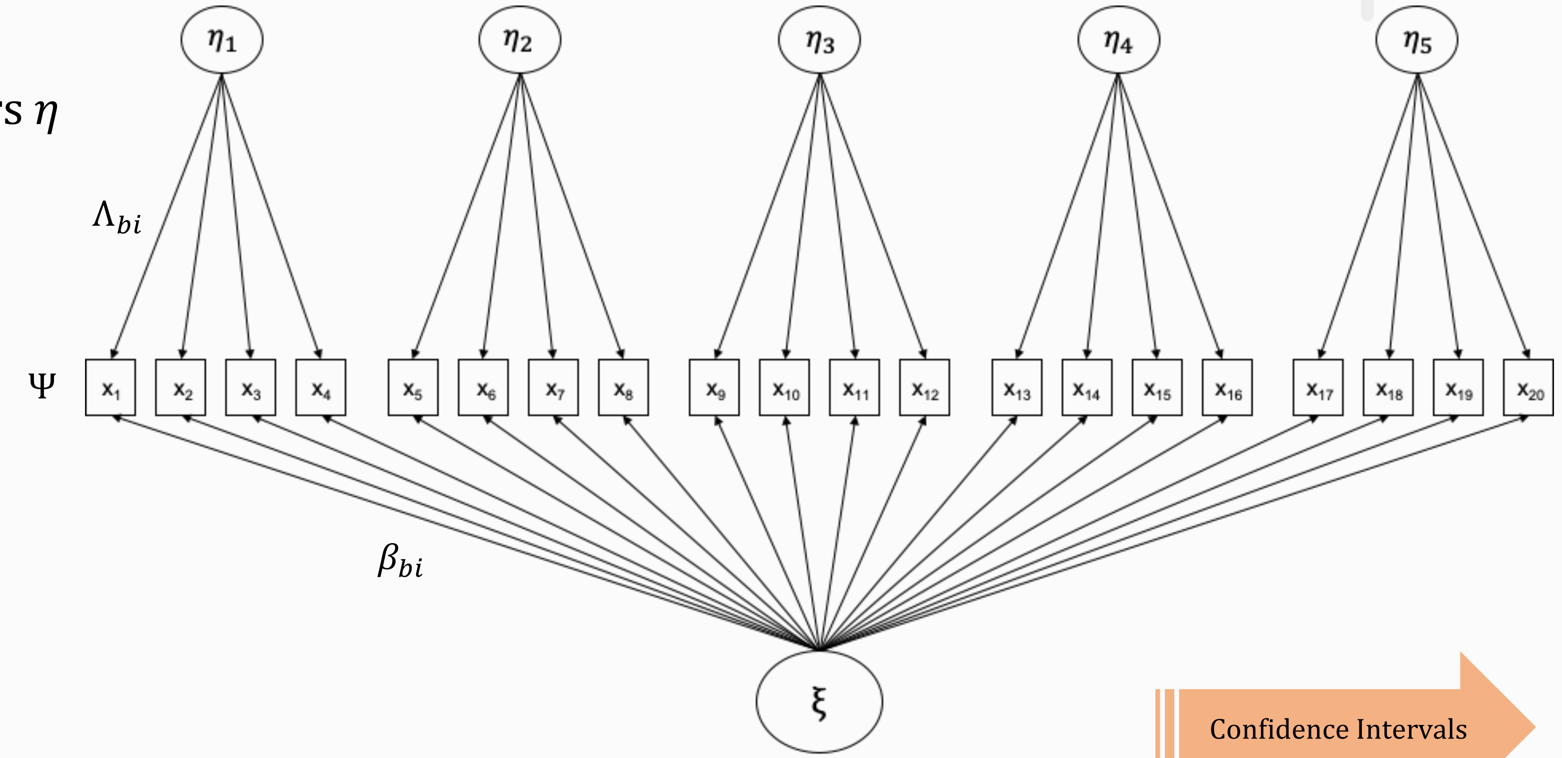
β_{bi} := loadings on the common factor ξ

Λ_{bi} := factor loadings on the group factors η

Ψ := residual covariance matrix

$$\omega_t = \frac{\sum \beta_{bi}\, \beta_{bi}^T + \sum \Lambda_{bi} \Lambda_{bi}^T}{\sum \beta_{bi}\, \beta_{bi}^T + \sum \Lambda_{bi} \Lambda_{bi}^T + \sum \Psi}$$

$$\omega_h = \frac{\sum \beta_{bi}\, \beta_{bi}^T}{\sum \beta_{bi}\, \beta_{bi}^T + \sum \Lambda_{bi} \Lambda_{bi}^T + \sum \Psi}$$



Confidence Intervals

Confidence Intervals

Frequentist procedures to fit the higher-order model:

- Exploratory Factor Analysis (EFA)

  → Bootstrap intervals

- Confirmatory Factor Analysis (CFA)

  → Wald-type interval

An X% *confidence interval for a parameter* θ *is an interval* (L, U) *generated by a procedure that in repeated sampling has an* X% *probability of containing the true value of* θ, *for all possible values of* θ (Morey et al., 2016; Neyman, 1937)

Bayesian Estimation

Bayesian Estimation

Prior distribution × Data → Posterior Distribution

Benefits

- 95% credible interval contains ω_t with 95% probability
- Probability that ω_t is larger than, e.g., .90

Priors

Bayesian Estimation – Priors

Factor model:

$X = \Lambda \eta + \epsilon$

$\eta = \beta \xi + \delta$

We assume X is multivariate normally distributed:

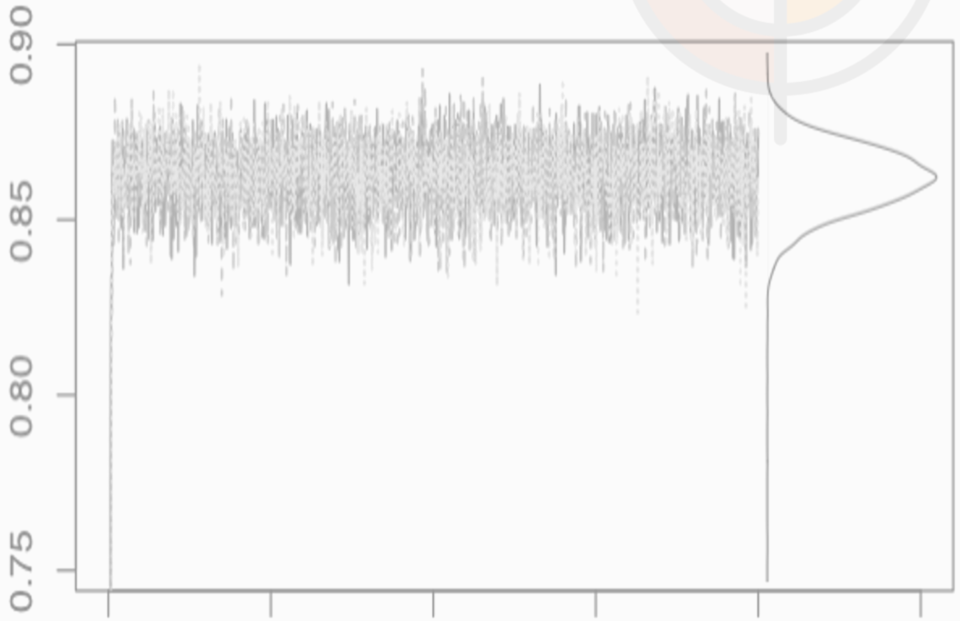
Parameter	$\boldsymbol{\Lambda}$ Group factor loadings	$\boldsymbol{\beta}$ Common factor loadings	$\boldsymbol{\psi_{\epsilon}^2}$ Variances of manifest residuals	$\boldsymbol{\psi_{\delta}^2}$ Variances of the latent residuals	$\boldsymbol{\Omega = (\xi, \eta)}$ Factor scores of all latent variables	$\boldsymbol{\Phi}$ Covariance matrix of latent variables
Prior	$N(0, \Sigma_{\Lambda})$	$N(0, \sigma_{\beta}^2)$	$\Gamma^{-1}(\alpha_{\epsilon}, \beta_{\epsilon})$	$\Gamma^{-1}(\alpha_{\delta}, \beta_{\delta})$	$N(0, \Sigma_{\Omega})$	$W^{-1}(\nu, \Psi)$

Lee, 2007

Procedure

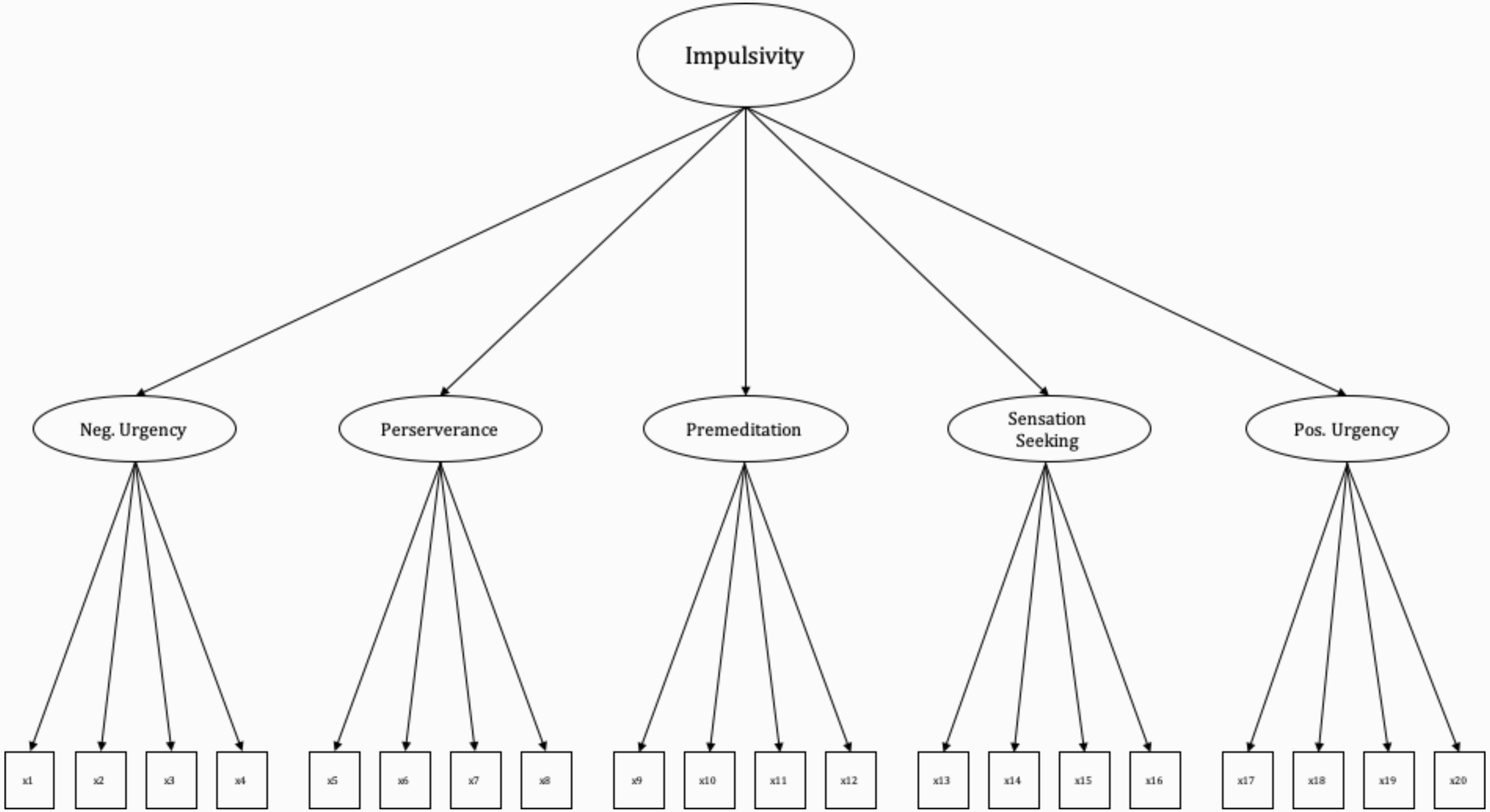
Bayesian Estimation – Procedure

1. MCMC: Draw consecutively from conditional posterior distributions of the parameters of the higher-order factor model (Lee, 2007)
2. Compute ω_t and ω_h for each posterior sample of factor model parameters (loadings and residuals)



Example

Example – Impulsivity-Scale

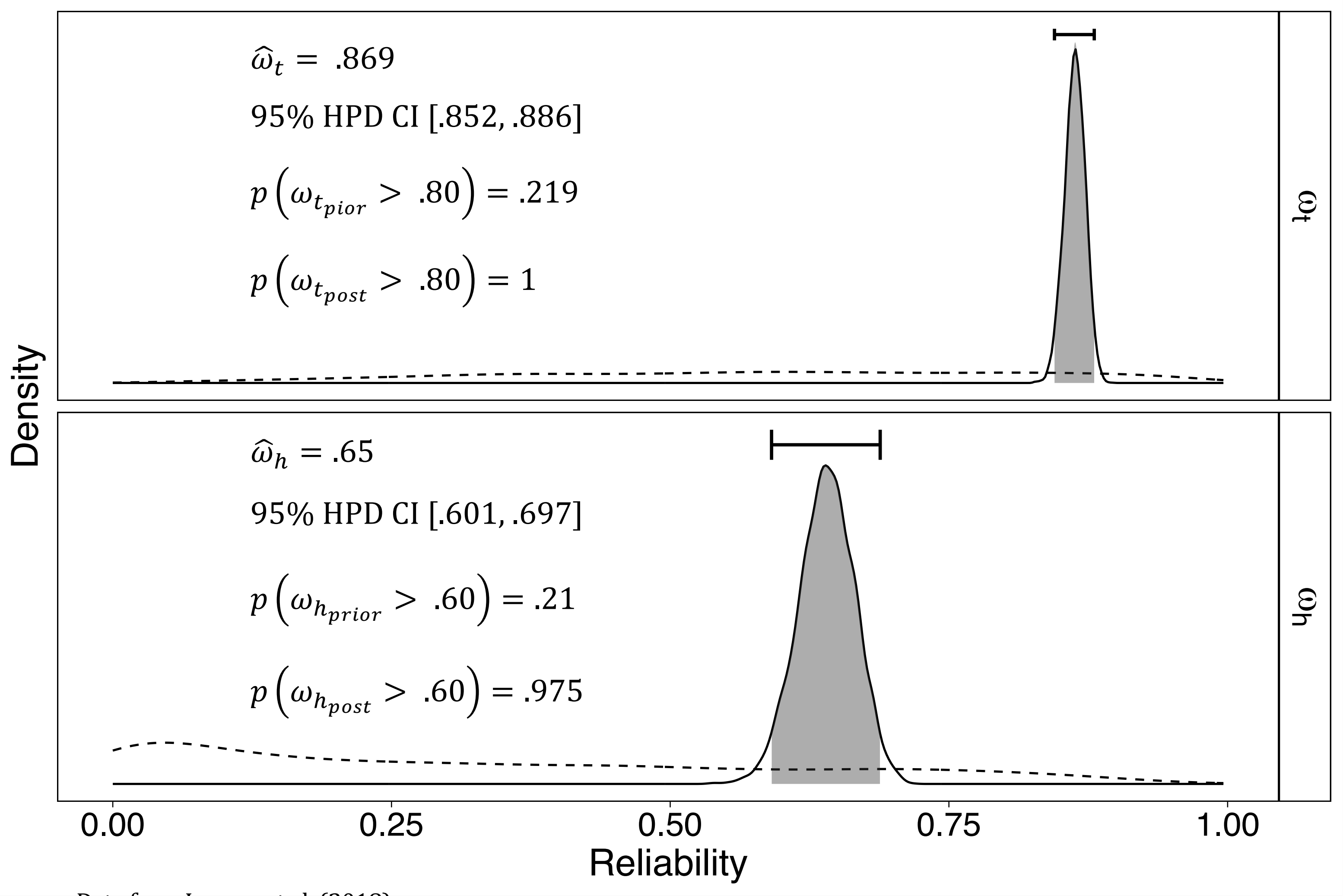


UPPS-P questionnaire (Cyders et al., 2014):

- Impulsivity measured by 20 Likert-scaled items
- Items: “I finish what I start”, “I quite enjoy taking risks”

Results

Example – Impulsivity-Scale – Results

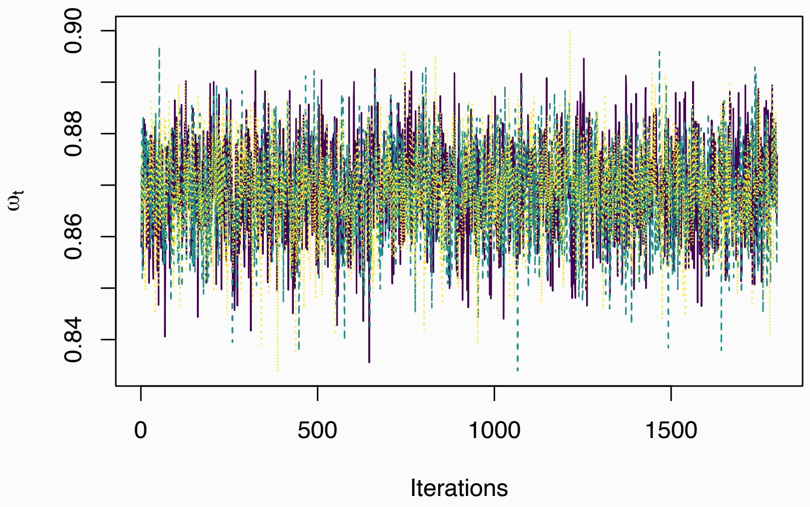


Data from Lozano et al. (2018)

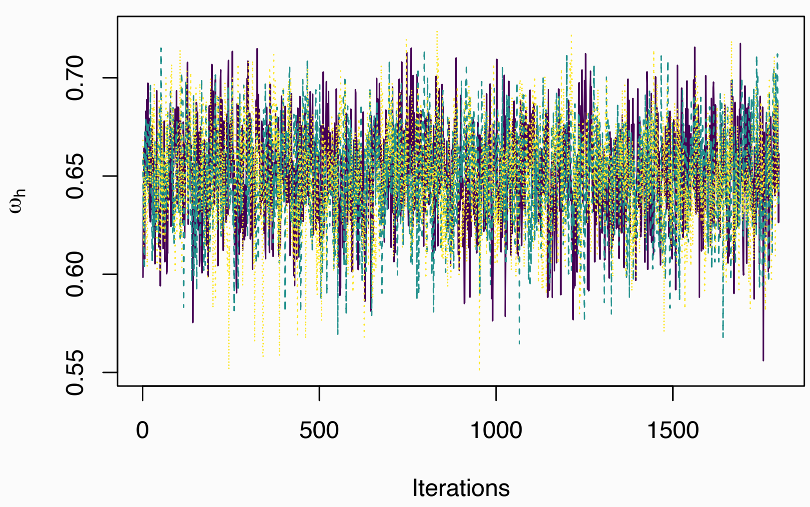
Convergence

Example – Impulsivity-Scale – Convergence

ω_t: $\hat{R} = 1.0005$

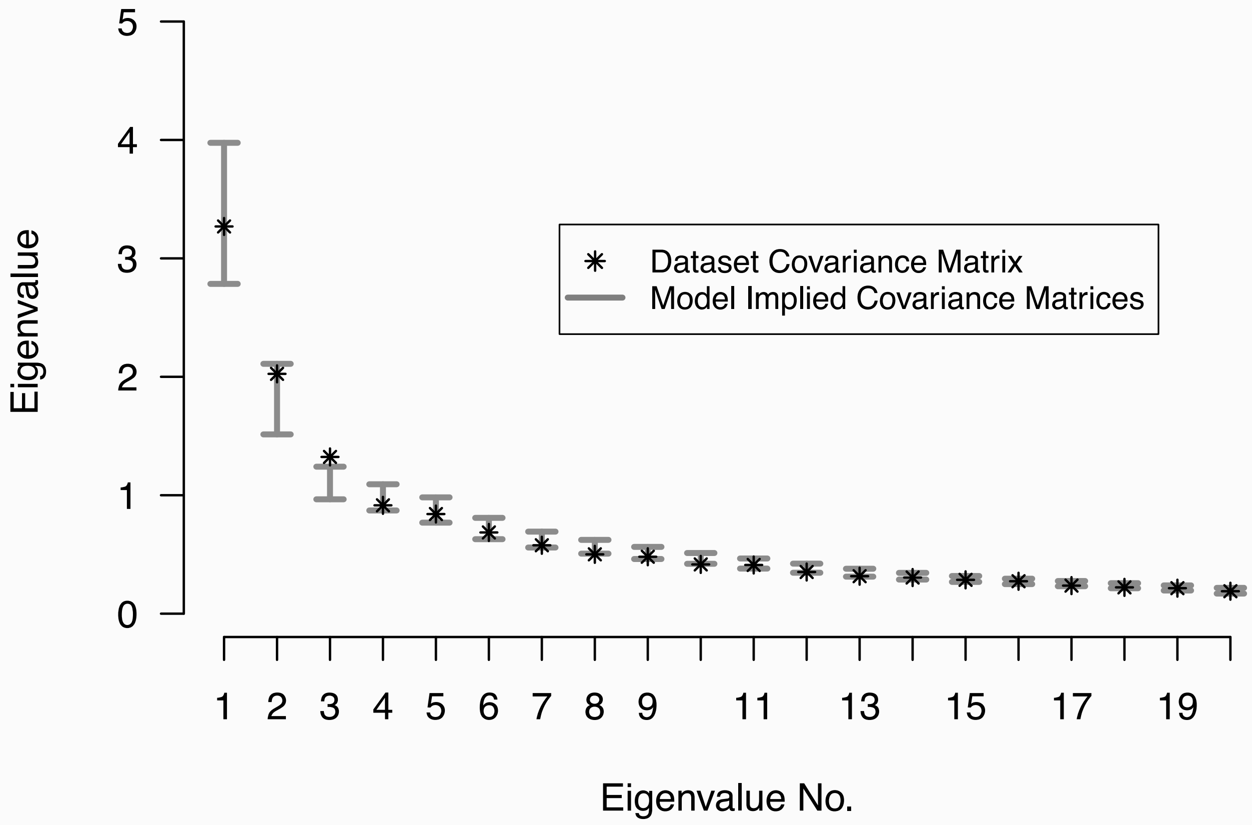


ω_h: $\hat{R} = 1.0002$



PPC

Example – Impulsivity-Scale - Posterior predictive check



!

Always check model fit

Simulation

Simulation – Setup

Comparison of six confidence intervals…

EFA	CFA
Bootstrap confidence intervals	Wald-type confidence interval
Standard error interval (SE) Standard error bias corrected interval (SE_{Bias}) Standard error log-transformed interval (SE_{Log}) Percentile interval (Perc) Bias corrected and accelerated interval (BCA)	

… and credible intervals for ω_t and ω_h

Results

Simulation – Results

95% Coverage results (excerpt):

Interval	ω_t	ω_h
SE	.927	.946
SE_{Bias}	.930	.940
SE_{Log}	.934	.947
Perc	.925	.954
BCA	.935	.944
Wald	.943	.941
Credible interval (HPD)	.942	.942

9 items, 3 group factors, n = 500, $\omega_t = .8$, $\omega_h = .6$

- generally confidence intervals and credible intervals agreed
- the SE, SE_{Log}, and Wald intervals performed well
- the Bayesian credible intervals performed well

Conclusion

Conclusion

- ! Uncertainty estimation is important
- Well-performing confidence intervals available for ω_t and ω_h
- Posterior distributions for ω_t and ω_h offer simple inferences and interpretation

Recommendation

Conclusion – Recommendation

How to obtain the intervals for ω_t and ω_h:

- In R :
  - bootstrap confidence intervals: `psych`-package
  - Wald intervals: `lavaan`-package (tedious), or `Bayesrel`-package (easier)
  - Credible intervals and posterior probabilities through the `Bayesrel`-package
- In JASP : coming soon…

Bayesrel

Bayesian reliability estimates for unidimensional and mutlidimensional tests

By J.M. Pfadt, D. van den Bergh, and J. Goosens

JASP

Statistics program with GUI, offers standard analysis procedures in both their classical and Bayesian form

By EJ Wagenmakers and Team, University of Amsterdam

References

Cyders, M. A., Littlefield, A. K., Coffey, S., & Karyadi, K. A. (2014). Examination of a short English version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors, 39(9*), 1372–1376. https://doi.org/10.1016/j.addbeh.2014.02.013

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons Ltd. https://doi.org/10.1002/9780470024737

Lozano, Ó. M., Díaz-Batanero, C., Rojas, A. J., Pilatti, A., & Fernández-Calderón, F. (2018). Concordance between the original and short version of the Impulsive Behaviour Scale UPPS-P using an IRT model. *PLOS ONE, 13*(3), 1–15. https://doi.org/10.1371/journal.pone.0194390

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London, Series A, 236*(767), 333–380. https://doi.org/10.1098/rsta.1937.0005