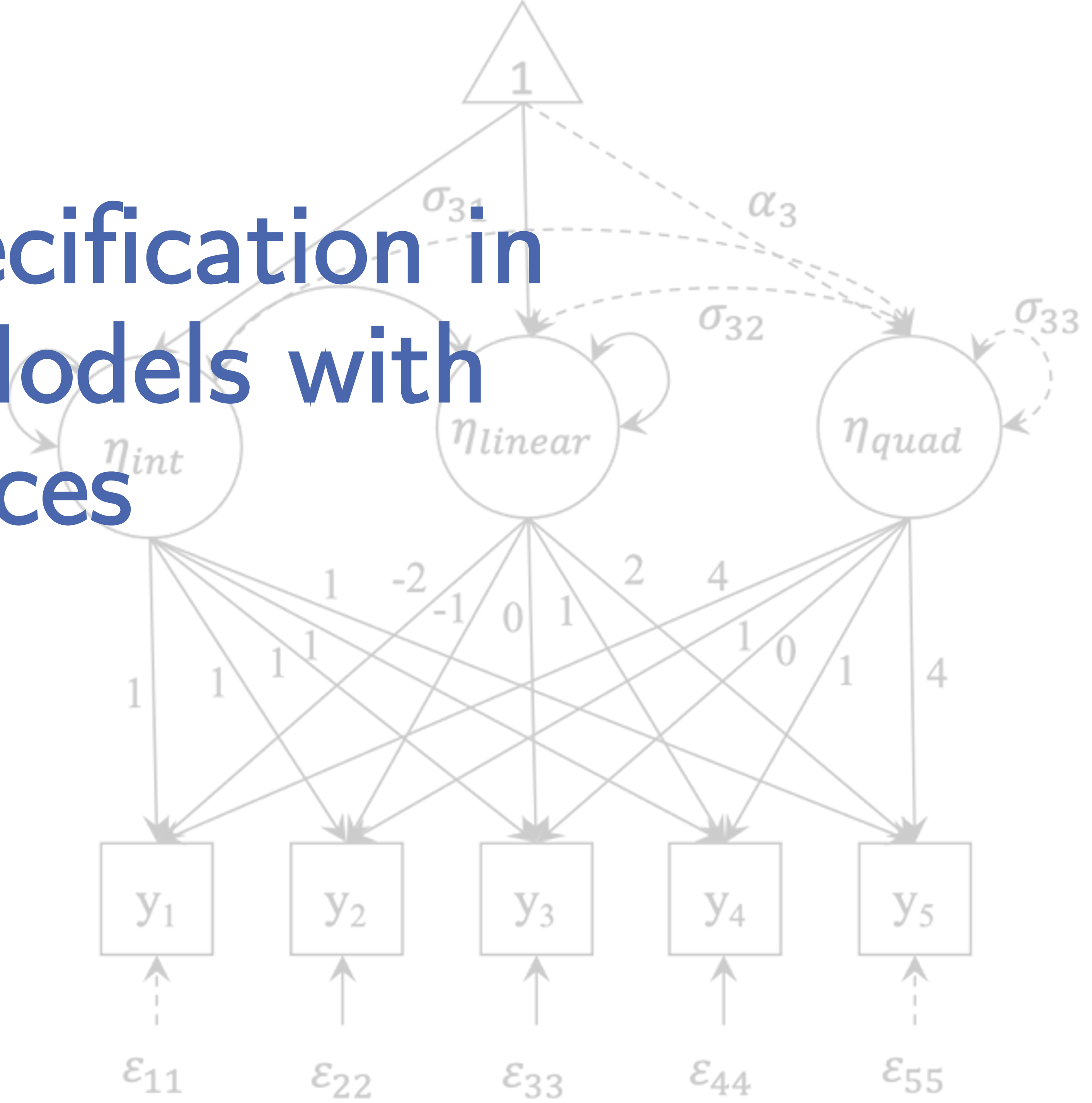
Detecting Misspecification in Latent Growth Models with Bayesian Fit Indices

Sonja D. Winter & Sarah Depaoli

Psychological Sciences, UC Merced

IMPS 2021



Outline

- Why do we need model fit and selection indices?
- Bayesian model fit and selection indices
- Simulation design
- Results
- Discussion

Why do we need model fit and selection indices with?

- **Model fit:** Does the hypothesized model provide a good fit to the observed data?
- **Model selection:** If multiple competing models exist, which of these models best represents the observed data?

- Need to use indices to quantify the misspecification

Bayesian Estimation of Structural Equation Models

- Advantages:
  - Explicitly update previous knowledge[1]
  - Prior distributions[2]
  - Posterior distribution[3]
- Disadvantages
  - Subjective[4]
  - Potentially reduced generalizability[5]
  - More effort required[6]
  - Limited options for model fit assessment (until recently)[7]

[1] van de Schoot et al., 2014; [2] Lee, 2007; Smid et al., 2019; [3] Kaplan & Depaoli, 2012; [4] Press, 2003; [5] Stromeyer et al., 2015; [6] MacCallum et al., 2012; [7] Lee, 2011

Bayesian model fit and selection indices

- New and improved model fit indices were recently introduced:[1]
  - Bayesian approximate model fit indices: Comparative Fit index (BCFI), Tucker-Lewis index (BTLI), and root mean square error of approximation (BRMSEA)
  - The posterior-predictive p-value (PPP) now handles missing data more appropriately
- Also available: information criteria such as the Bayesian or Deviance information criterion (BIC or DIC) for model selection

[1] Asparouhov & Muthén, 2020; Garnier-Villarreal & Jorgensen, 2019

Bayesian model fit and selection indices

- New and improved model fit indices were recently introduced:[1]
  - Bayesian approximate model fit indices: Comparative Fit index (BCFI), Tucker-Lewis index (BTLI), and root mean square error of approximation (BRMSEA)
  - The posterior-predictive *p*-value (PPP) now handles missing data more appropriately
- Also available: information criteria such as the Bayesian or Deviance information criterion (BIC or DIC) for model selection

- Exciting, but need more insight in how these indices function when
  - sample sizes are small[2]
  - missing data are wide-spread[3]
  - misspecification occurs in different parts of the model[4]
  - priors are informative[2]

[2] e.g., Cain & Zhang, 2019; [3] e.g., Asparouhov & Muthén, 2020; [4] e.g., Wu & West, 2010

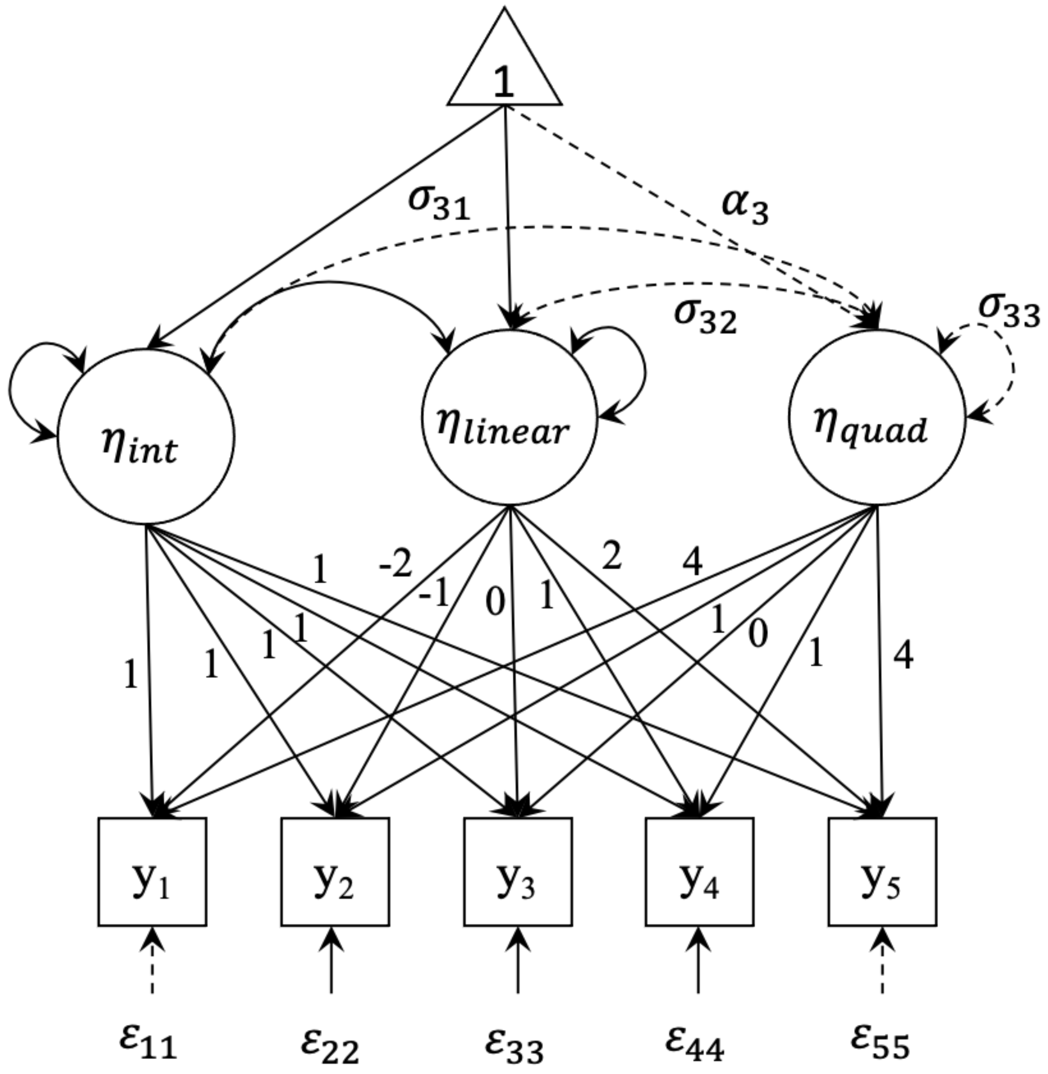
Goal and approach

- **Goal:** Examine the ability of Bayesian model fit and selection indices (i.e., PPP, BCFI, BTLI, BRMSEA, BIC, and DIC) to detect model misspecification in a latent growth model

- **Approach:** Using a simulation study, examine for each model fit or selection index:
  1. How often model misspecification was detected using common cutoff values (not for BIC/DIC)
  2. How often each model specification was selected as the "best" model

Simulation Design

Simulation design: Population Model



Substantively irrelevant

Level 1 misspecified model: $\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33} = \varepsilon_{44} = \varepsilon_{55}$

Level 2 misspecified model: $\sigma_{33} = 0, \sigma_{32} = 0$, and $\sigma_{31} = 0$

Level 3 misspecified model: $\sigma_{33} = 0, \sigma_{32} = 0, \sigma_{31} = 0$, and $\alpha_3 = 0$

Substantively relevant

$$\Phi = \begin{bmatrix} 3.00 & & \\ 0.24 & 0.30 & \\ 0.09 & 0.03 & 0.04 \end{bmatrix}$$

$$\alpha = \begin{bmatrix} 1.00 \\ 0.30 \\ -0.08 \end{bmatrix}$$

$$diag(\Theta_\varepsilon) = [.20 \quad .50 \quad .50 \quad .50 \quad .80]$$

Simulation design: Conditions

- **Sample size:** 50, 100, 250, 500
- **Missingness mechanism:** missing at random, related to observed value at time point 1
- **Variables with missing data:** 0, 1, or 4
- **Percentage of missing values:** 15% or 50%
  - With 4 variables with missing data, the missingness followed a dropout pattern
- **Prior Specification:** diffuse, aligned, or diverging priors for α_{int} and α_{linear} parameters

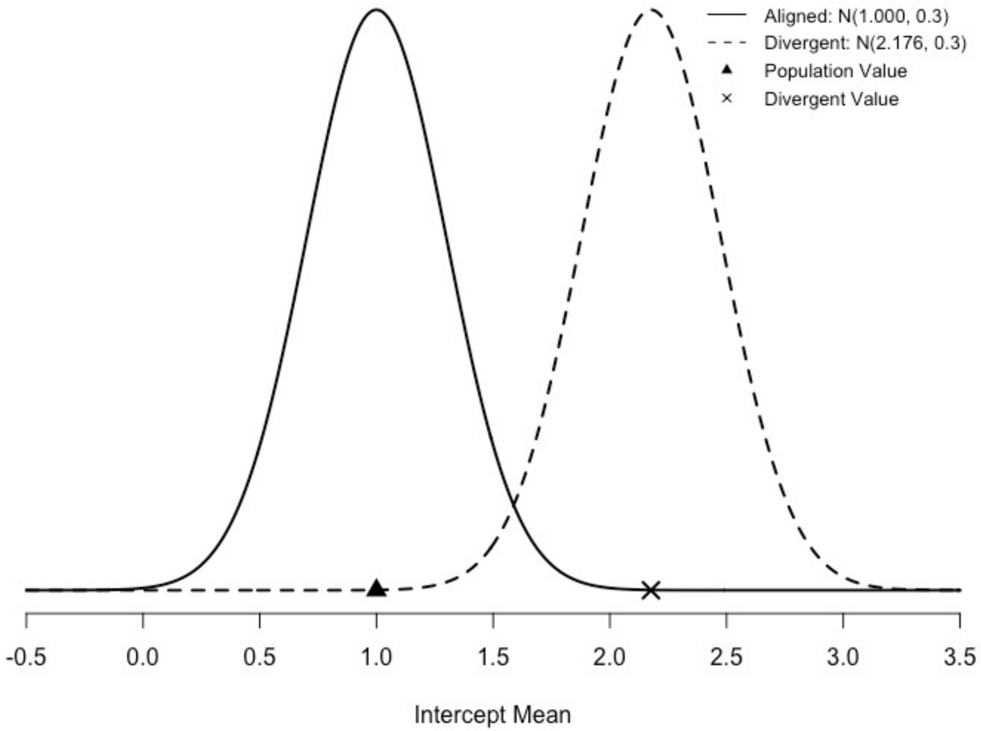
Simulation design: Estimation

- **Software:** M*plus*
- **Sampler:** Gibbs
- **Estimation:** 2 chains with 20,000 iterations (10,000 discarded as burn-in)
- Convergence assessed with $\hat{R}$-value (< 1.05) and effective sample size estimate ($> 1{,}000$)
- 1,000 replications per simulation cell

Simulation design: Priors

- Priors that were varied:
  - $\alpha_{int} \sim \mathrm{N}(\mu = 1.000, \sigma = 0.3)$ (aligned) or $\sim \mathrm{N}(\mu = 2.176, \sigma = 0.3)$ (diverging)
  - $\alpha_{linear} \sim \mathrm{N}(\mu = 0.300, \sigma = 0.1)$ (aligned) or $\sim \mathrm{N}(\mu = 0.692, \sigma = 0.1)$ (diverging)

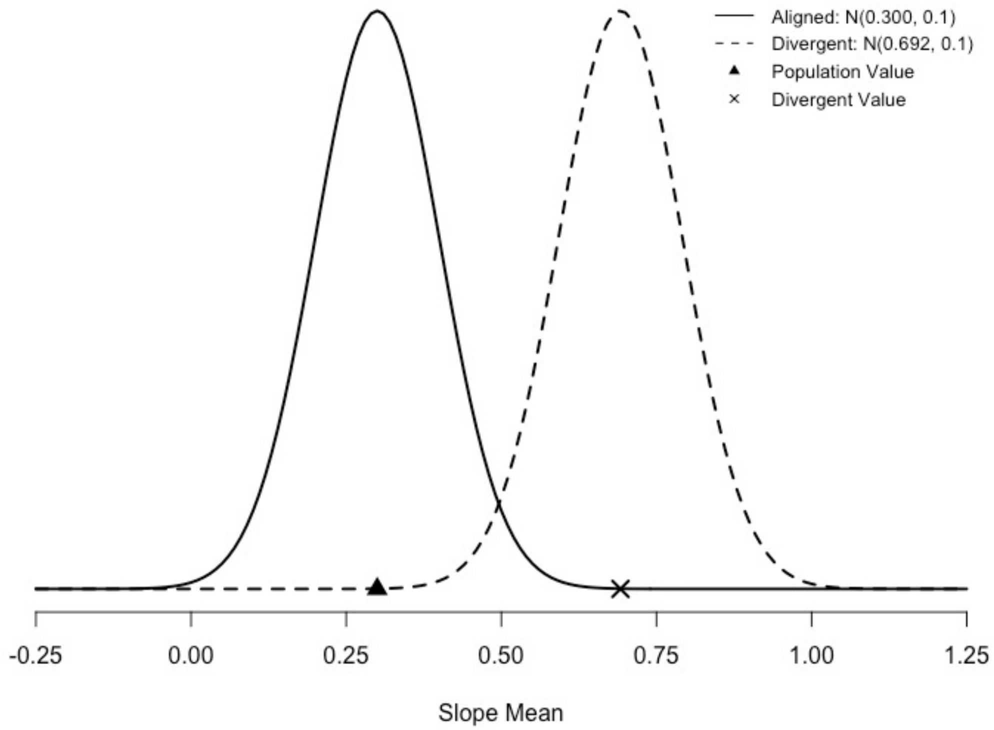
- Priors on remaining parameters:
  - $\alpha \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\nu \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\lambda \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\Phi \sim \mathrm{IW}(\Sigma = \mathbf{I},\ \nu = p + 1)$, where p equals the number of latent factors and $\mathbf{I}$ is an identity matrix of dimension p
  - $\epsilon \sim \mathrm{IG}(\alpha = -1, \beta = 0)$



Simulation design: Priors

- **Priors that were varied:**
  - $\alpha_{int} \sim \mathrm{N}(\mu = 1.000, \sigma = 0.3)$ (aligned) or $\sim \mathrm{N}(\mu = 2.176, \sigma = 0.3)$ (diverging)
  - $\alpha_{linear} \sim \mathrm{N}(\mu = 0.300, \sigma = 0.1)$ (aligned) or $\sim \mathrm{N}(\mu = 0.692, \sigma = 0.1)$ (diverging)

- **Priors on remaining parameters:**
  - $\alpha \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\nu \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\lambda \sim \mathrm{N}(\mu = 0, \sigma^2 = 10^{10})$
  - $\Phi \sim \mathrm{IW}(\Sigma = \mathbf{I},\ \nu = p + 1)$, where p equals the number of latent factors and $\mathbf{I}$ is an identity matrix of dimension p
  - $\epsilon \sim \mathrm{IG}(\alpha = -1, \beta = 0)$

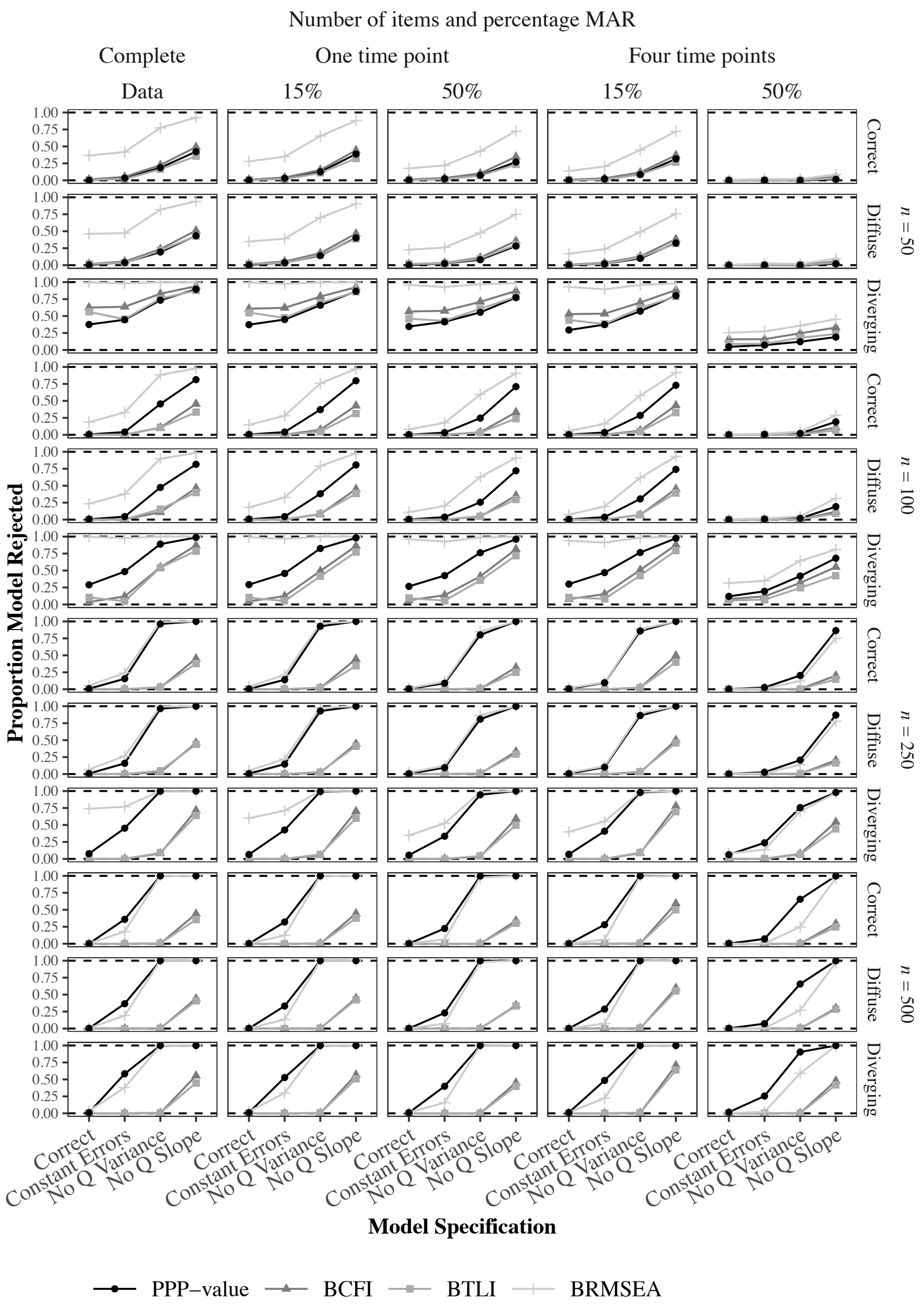


Results

Results: Model Rejection Rates

Cutoff values used:

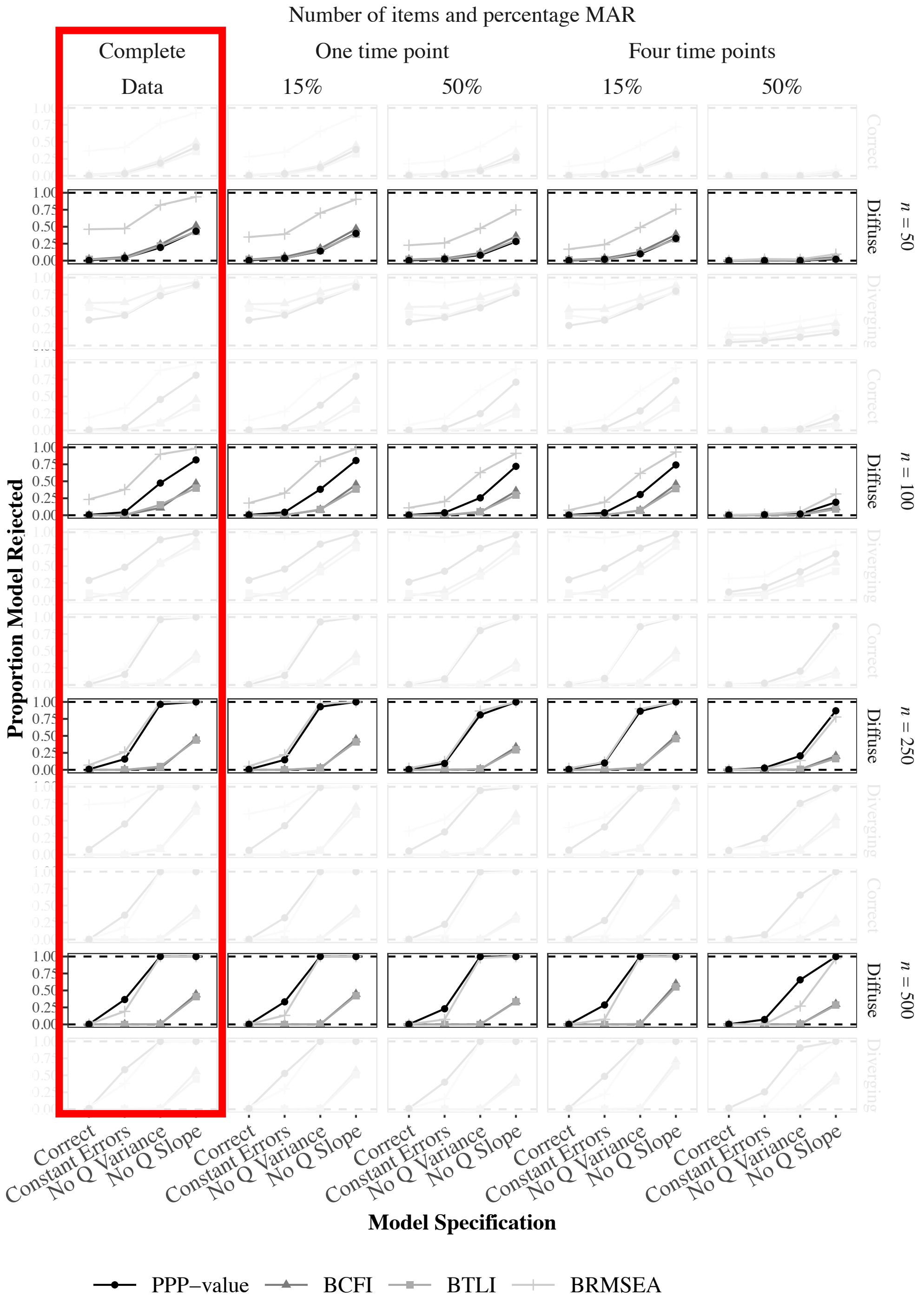
- PPP > .05
- BCFI/BTLI > .95
- BRMSEA < .06



Results: Model Rejection Rates

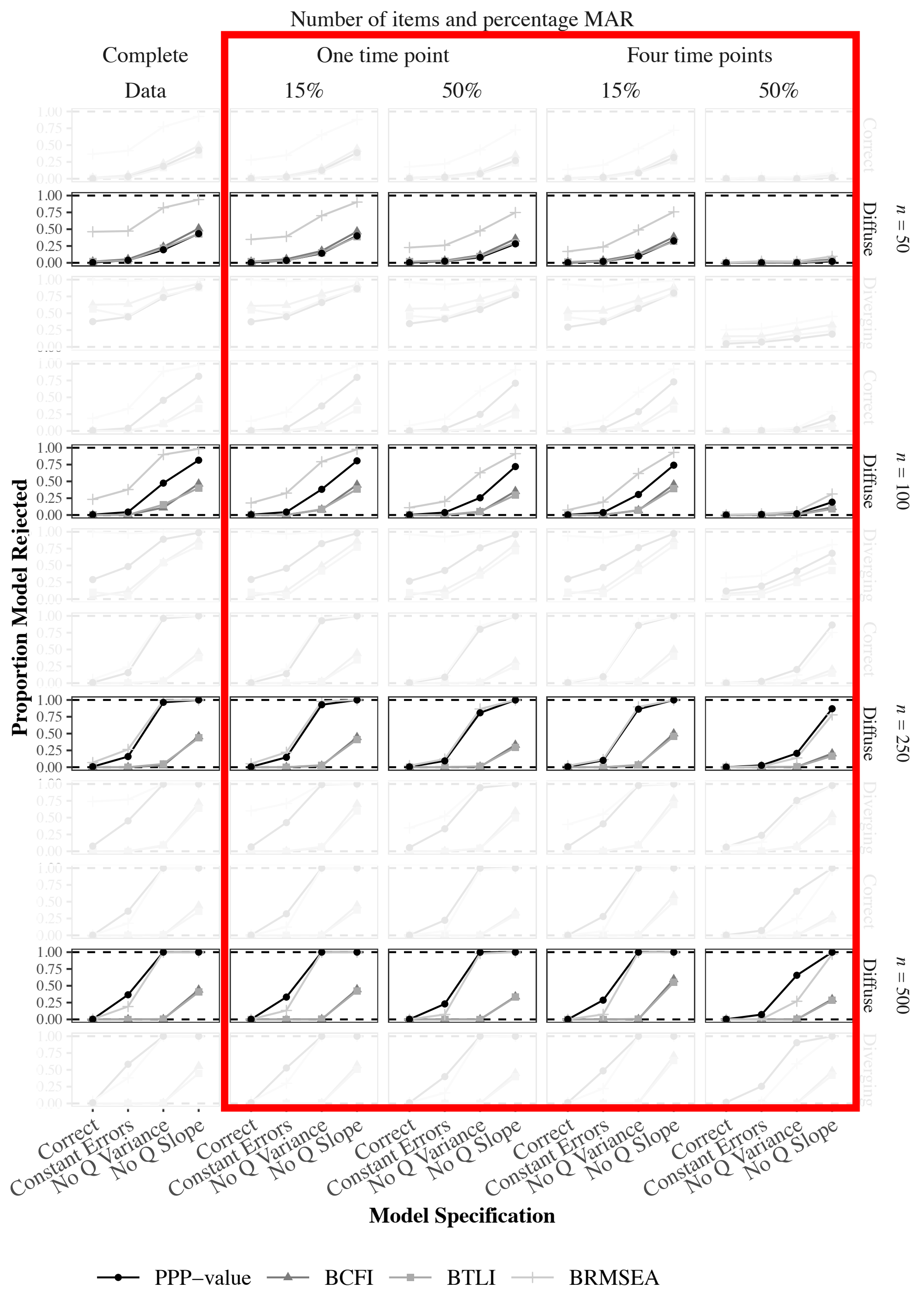
- Complete data: PPP-value and BRMSEA likely to reject substantively relevant misspecification if $n \geq 250$

- Missing data: All fit indices (esp. BCFI and BTLI) less likely to reject misspecified models



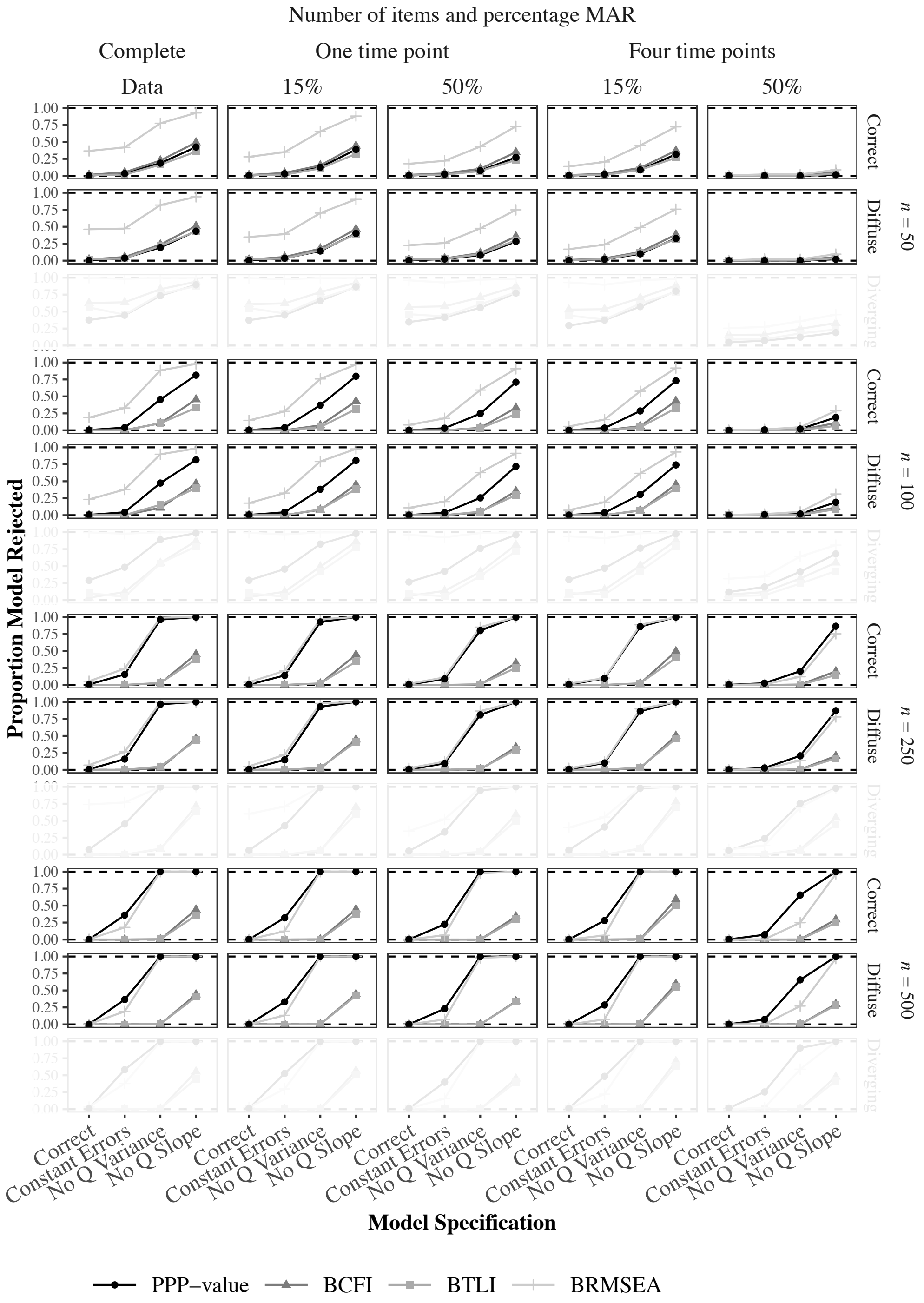
Results: Model Rejection Rates

- Complete data: PPP-value and BRMSEA likely to reject substantively relevant misspecification if $n \geq 250$
- Missing data: All fit indices (esp. BCFI and BTLI) less likely to reject misspecified models



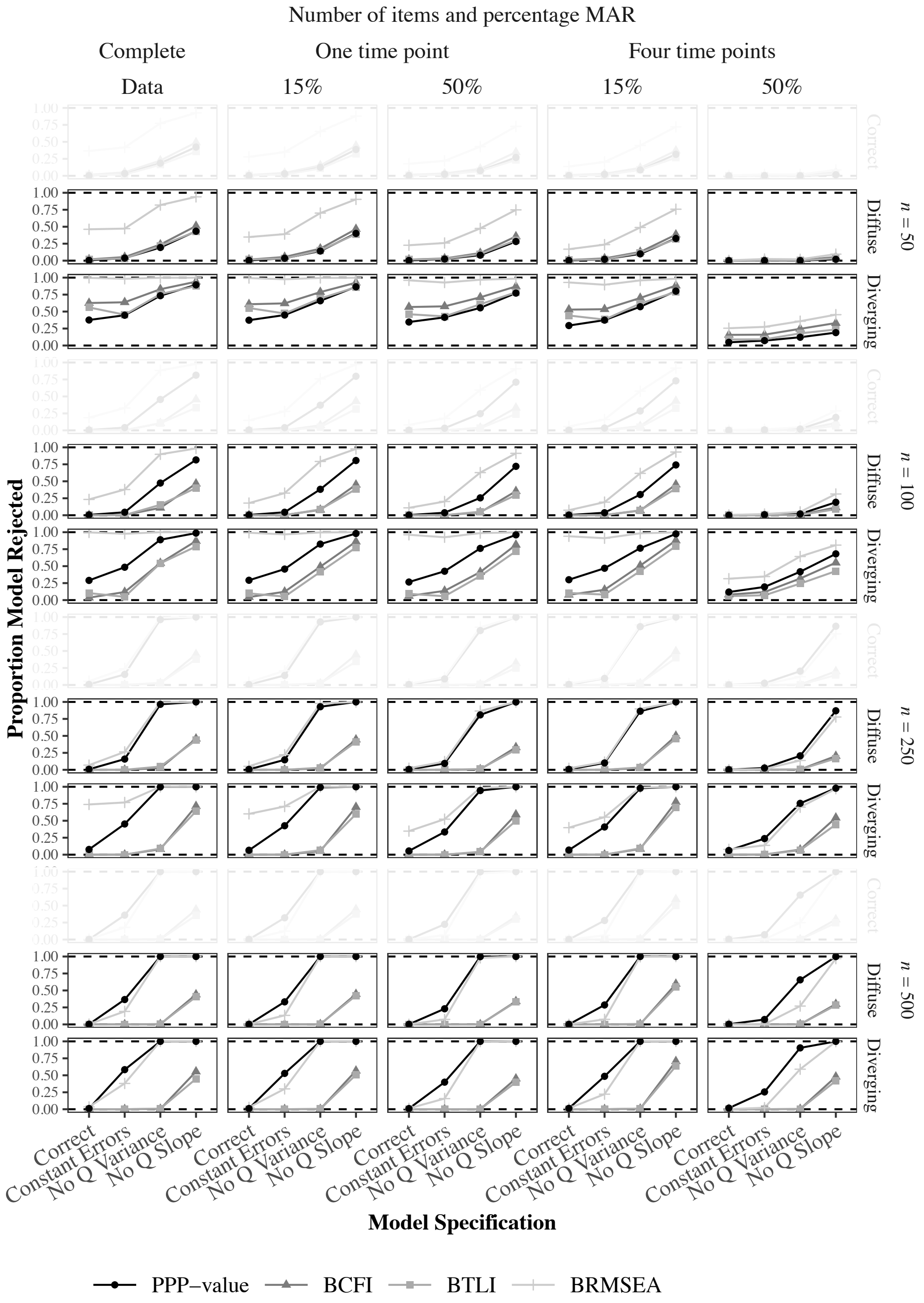
Results: Model Rejection Rates

- Aligned priors do not increase model rejection rates for misspecified models
- Aligned priors provided equal or less information compared to data

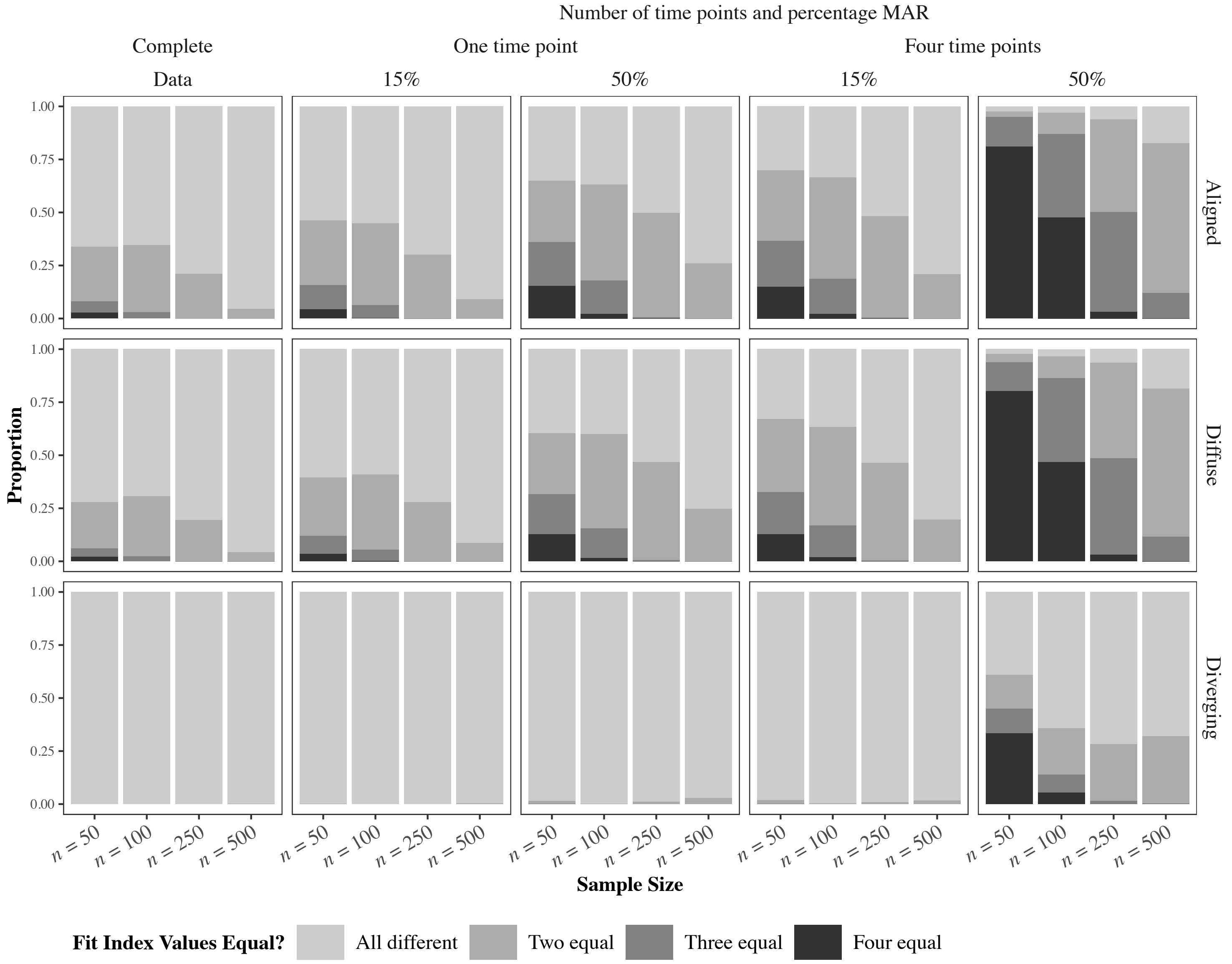


Results: Model Rejection Rates

- Diverging priors resulted in inflated model rejection rates for the correctly specified model based on BCFI/BTLI ($n = 50$), PPP-value ($n \leq 100$), and BRMSEA ($n \leq 250$)

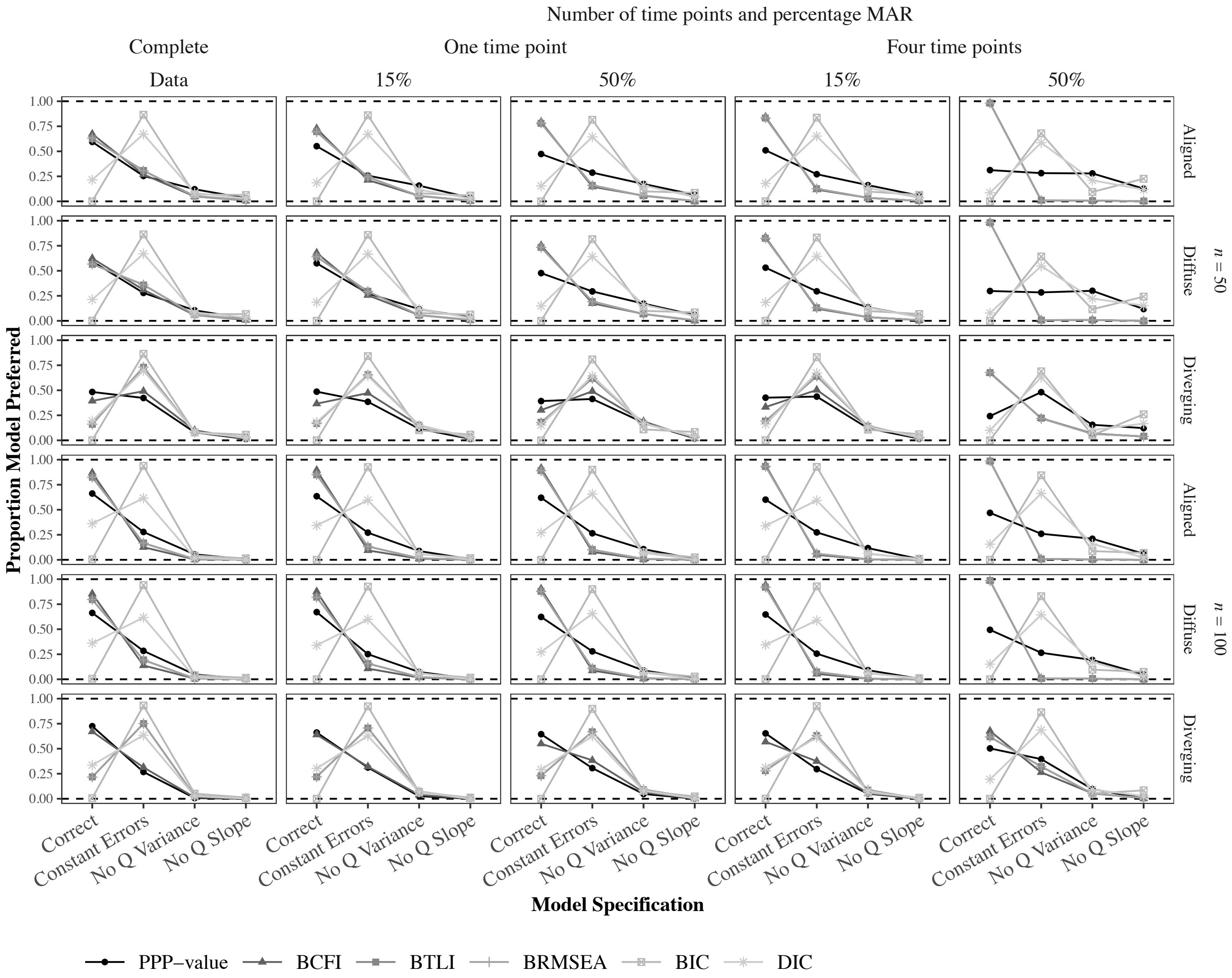


Results: Model Selection Rates

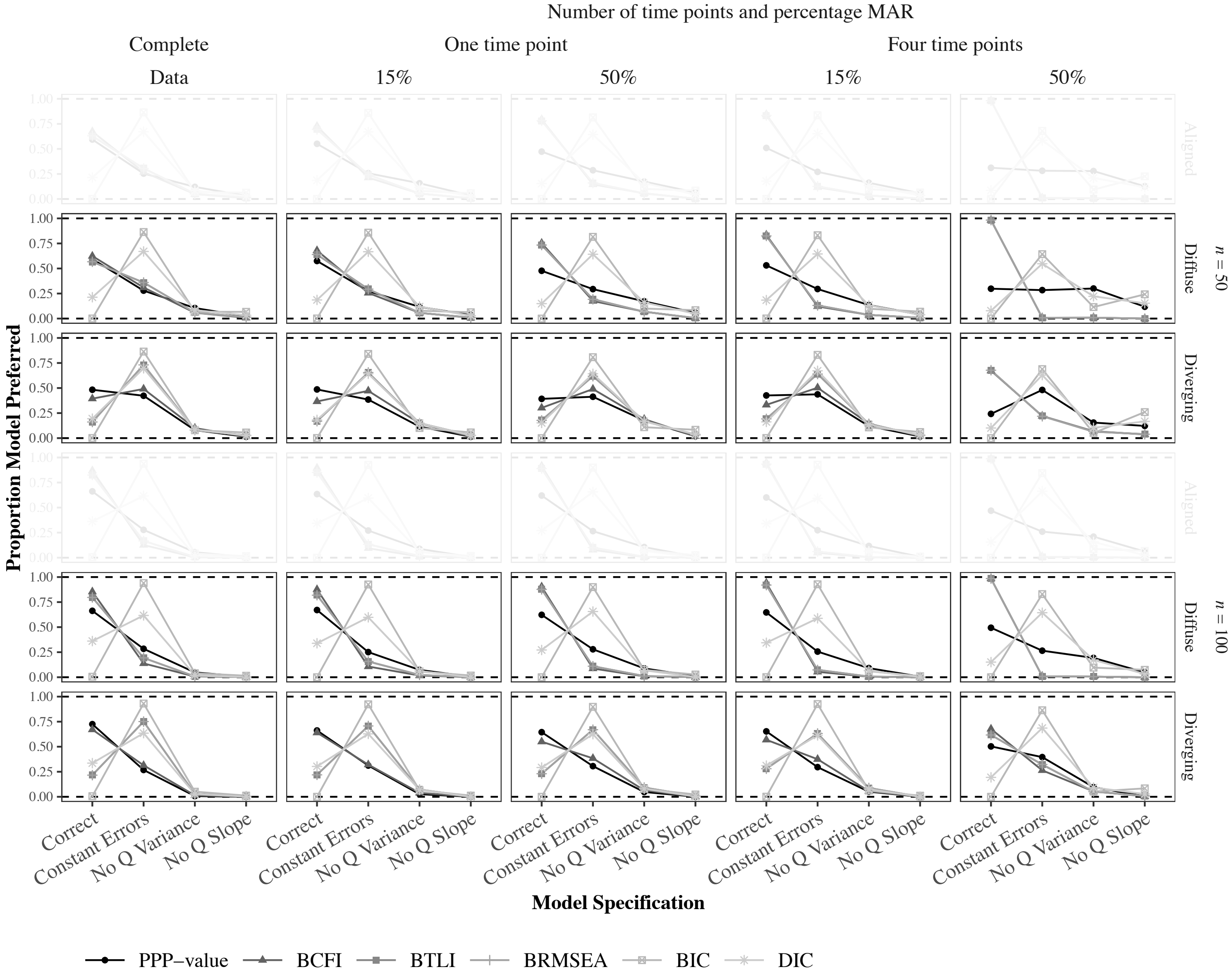


- Computational difficulties with approximate fit indices:
- Looks like "perfect" fit
- Equivalent across models

Results: Model Selection Rates

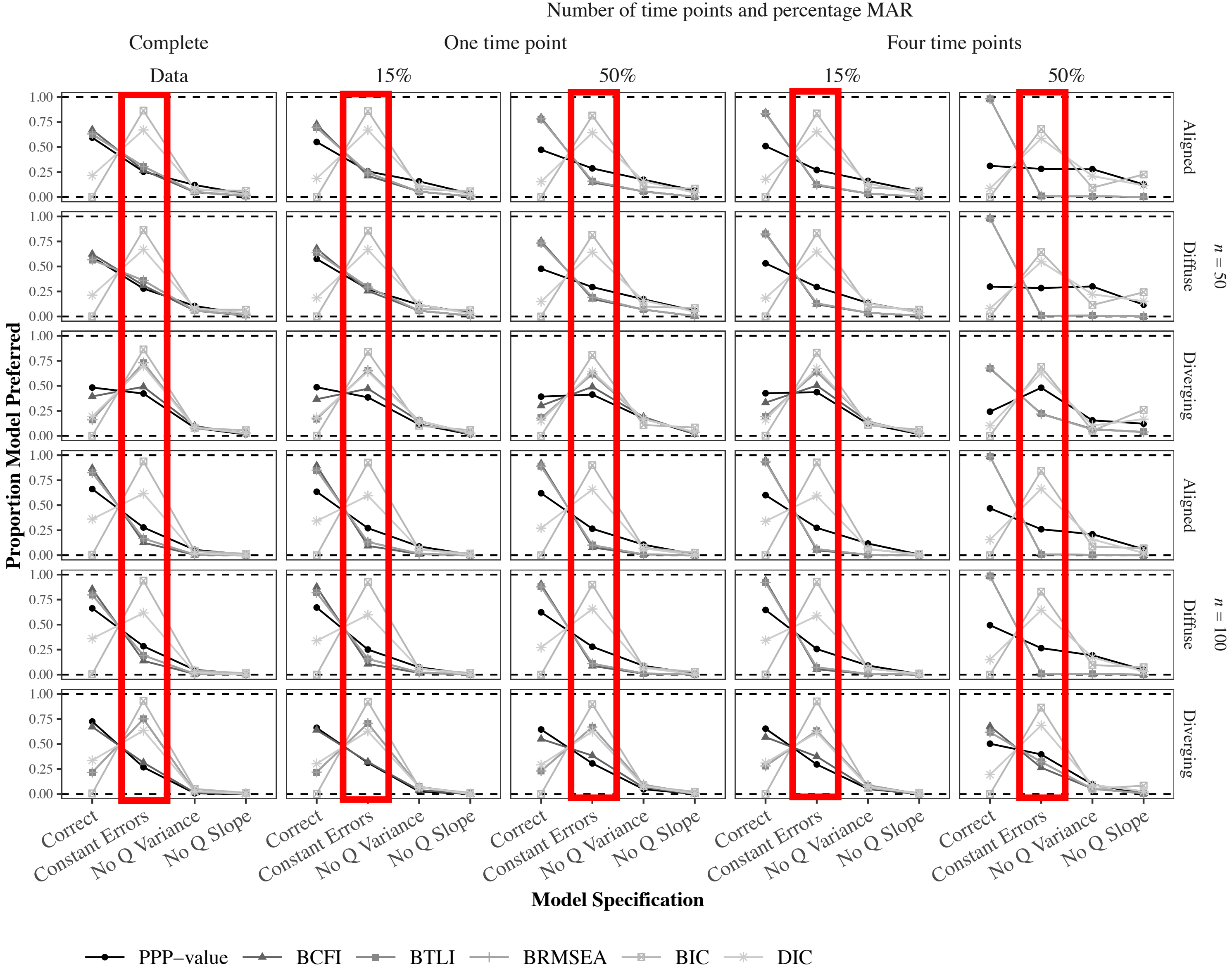


Results: Model Selection Rates



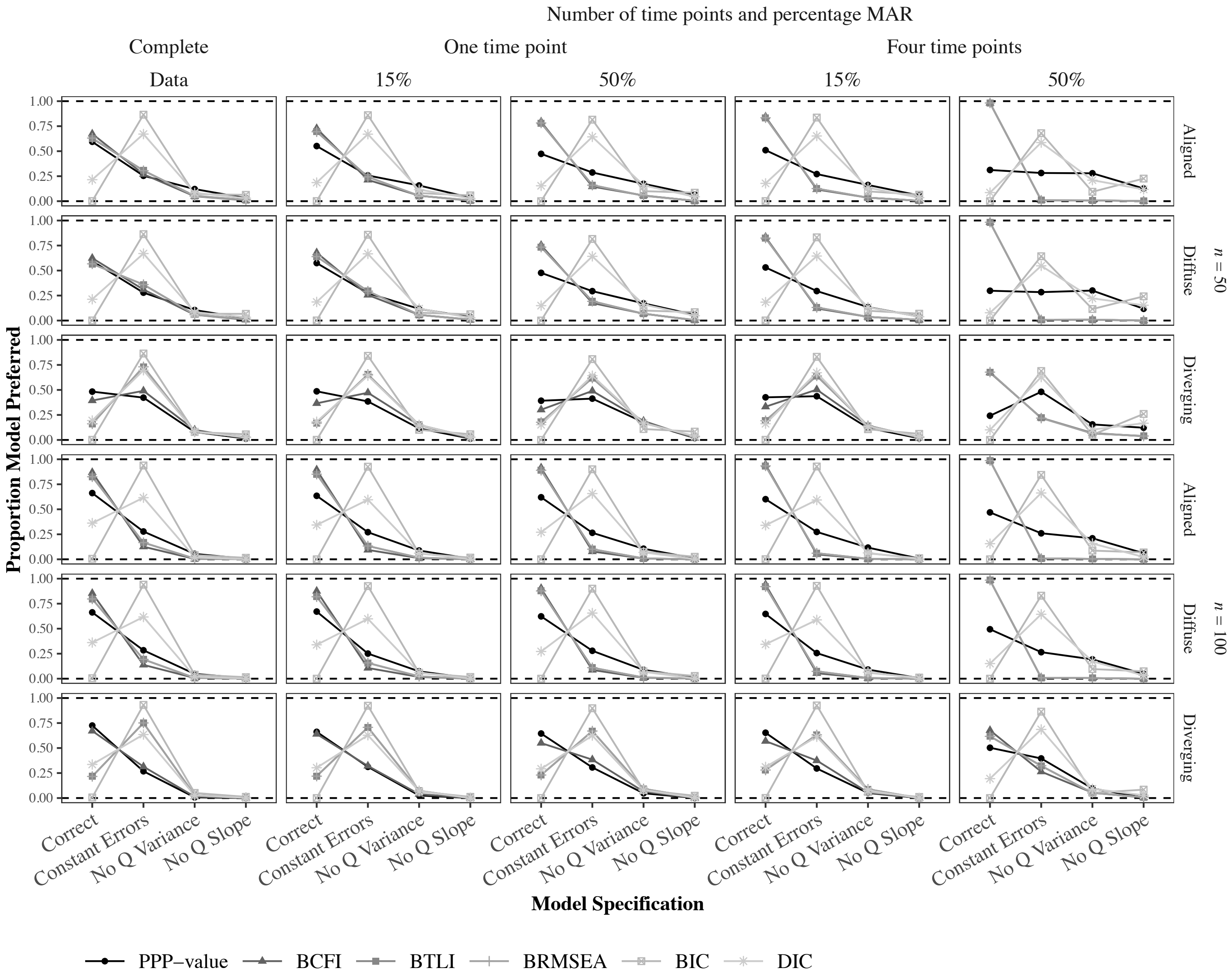
- Diverging priors affect model selection most when $n = 50$

Results: Model Selection Rates



- Diverging priors affect model selection most when $n = 50$
- BIC/DIC prefer more parsimonious model over correct model

Results: Model Selection Rates



- Diverging priors affect model selection most when $n = 50$
- BIC/DIC prefer more parsimonious model over correct model
- Even at $n = 50$, correct model is likely to be selected
- Large numbers of missing values reduce correct model selection rates

Discussion

Discussion: diffuse priors

- BCFI and BTLI cutoffs (.95) unlikely to reject substantively relevant misspecification
- BRMSEA and PPP-value more likely to reject substantively relevant misspecification
- Approximate fit indices difficult to compute with small samples or large numbers of missing values
  - Fit of baseline and estimated model become too similar

Discussion: diffuse priors

- PPP-value is likely to select the correct model
- If approximate fit indices can be computed and are different across models, they are likely to select the correct model
- Missing values reduce ability to select correct model across indices
- BIC/DIC may select parsimonious model over correct model

Discussion: aligned or diverging priors

- Including informative priors may interfere with model fit and model selection if priors diverge from data
- Researchers who include informative priors *need* to examine the impact of the prior through a prior sensitivity analysis, or by examining prior-data disagreement

Thank you!

- If you have any questions or comments, email me at swinter@ucmerced.edu!

References:

Asparouhov, T., & Muthén, B. (2020). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*. https://doi.org/10.1080/10705511.2020.1764360

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An Evaluation of PPP and DIC for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(1), 39–50. https://doi.org/10.1080/10705511.2018.1490648

Garnier-Villarreal, M., & Jorgensen, T. D. (2019). Adapting Fit Indices for Bayesian Structural Equation Modeling: Comparison to Maximum Likelihood. *Psychological Methods*. https://doi.org/10.1037/met0000224

Kaplan, D., & Depaoli, S. (2012). Bayesian Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). Guilford Press. https://doi.org/10.1016/B978-044452044-9/50011-2

Levy, R. (2011). Bayesian Data-Model Fit Assessment for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 663–685. https://doi.org/10.1080/10705511.2011.607723

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Wiley.

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and Cautions in Implementing Bayesian Structural Equation Modeling. *Psychological Methods*, *17*(3), 340–345. https://doi.org/10.1037/a0027131

Press, S. J. (2003). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications* (Second edi). John Wiley & Sons, Inc.

Smid, S. C., McNeish, D. M., Miočević, M., & van de Schoot, R. (2019). Bayesian Versus Frequentist Estimation for Structural Equation Models in Small Sample Contexts: A Systematic Review. *Structural Equation Modeling: A Multidisciplinary Journal*. https://doi.org/10.1080/10705511.2019.1577140

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The Prowess and Pitfalls of Bayesian Structural Equation Modeling: Important Considerations for Management Research. *Journal of Management*, *41*(2), 491–520. https://doi.org/10.1177/0149206314551962

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*, *85*(3), 842–860. https://doi.org/10.1111/cdev.12169

Wu, W., & West, S. G. (2010). Sensitivity of Fit Indices to Misspecification in Growth Curve Models. *Multivariate Behavioral Research, 45*(3), 420–452. https://doi.org/10.1080/00273171.2010.483378