Bayesian estimation of the RMSEA

Abstract

The root mean square error of approximation (RMSEA) is a commonly used goodness of fit index in structural equation modeling. For this reason, the RMSEA has been extended to Bayesian SEMs. However, current methods of computing the RMSEA rely on the distribution of realized values. As an alternative, we present an approach to Bayesian estimation of the RMSEA that models the RMSEA as a parameter. By modeling the RMSEA as a parameter, uncertainty about structural parameters reflects the degree of model misspecification, yielding more reliable inference. These features of the proposed approach are demonstrated using a simulation study.

Keywords: model misspecification, Bayesian SEM, Bayesian RMSEA

Bayesian estimation of the RMSEA

Structural equation modeling (SEM) is a multivariate technique for modeling covariance structures. In SEMs, the modeler constrains the covariance structure based on hypotheses about the relations underlying the data. Given the constrained structure or structured covariance matrix, examining the degree to which the structured covariance matrix matches the true covariance matrix underlying the data is an important aspect of SEM – hence the focus on model misspecification. The standard way to assess model misspecification is to estimate the SEM, then compute the discrepancy between the structured covariance matrix and the sample covariance matrix. The discrepancy statistic is hypothesized to follow a particular distribution under the null hypothesis that the structured covariance matrix is not different from the true covariance matrix underlying the data. In practice, it is often a question of statistical power whether this null will be rejected, as the hypothesized structure will rarely ever be the true structure generating the data. For this reason, a variety of fit indices have been developed for quantifying the amount of misspecification. These indices function as effect sizes, and are interpreted to determine the degree to which misspecification is non-trivial.

In the Bayesian context, posterior predictive *p*-values (PPP, Levy, 2011) computed using a likelihood ratio discrepancy measure are used to assess the presence of model misspecification. As with frequentist estimation of SEMs, it is rare to obtain acceptable PPP values especially when the data are highly informative e.g. with larger samples. For this reason, Bayesian fit indices have also been developed to quantify the degree of model misspecification. Levy (2011) proposed a Bayesian standardized root mean squared residual (SRMR) based on the distribution of *realized values* of the gap between the model-implied covariance matrix and the sample covariance matrix. By distribution of realized values, we mean a distribution of quantities computed based on the posterior distribution of parameters – these quantities play no direct role in Bayesian model estimation. Similarly, Hoofs, van de Schoot, Jansen, and Kant (2018) and Garnier-Villarreal and Jorgensen

(2020) have developed different Bayesian root mean square error of approximation (RMSEA) statistics both based on the distribution of realized values. Moreover, the work of Garnier-Villarreal and Jorgensen (2020) introduced additional goodness of fit indices such as the comparative fit index (CFI) and Tucker-Lewis index (TLI). Differently from these approaches, Uanhoro (2023b) estimated Bayesian SEMs while simultaneously modeling a fit index akin to the correlation root mean square residual (CRMR).

In this paper, we present methods for *estimating* the RMSEA in Bayesian SEMs with saturated mean structure. This is different from other Bayesian SEM approaches that *return* the RMSEA post-model estimation. Our goal in this paper has already been accomplished in the frequentist literature (Wu & Browne, 2015), and our work here is an extension. The approach we present accomplishes two things: (i) return parameter estimates with uncertainty intervals that reflect the degree of model misspecification; (ii) returns the RMSEA with uncertainty intervals that contribute to examining the degree of model misspecification. Our work shares these similarities with the approach implemented by Uanhoro (2023b), the key difference being the goodness of fit index focused on. Notably, the RMSEA is not a *standardized* effect size measure as the size of misspecification implied by the RMSEA depends on several contextual factors (e.g. F. Chen, Curran, Bollen, Kirby, & Paxton, 2008; Savalei, 2012). The CRMR-type index returned by the approach in Uanhoro (2023b) is more easily interpretable as it is standardized. However, we agree with the consensus in the literature that is informative to consider multiple metrics in evaluating a given model (e.g. Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005; West, Taylor, & Wu, 2012). Hence, we find it reasonable to present an approach for estimating the RMSEA in Bayesian models.

In the next section of the paper, we elaborate the methods we present alongside their theoretical rationale. Then we conduct a simulation study to assess the adequacy of the methods we present for computing the RMSEA, alongside extant Bayesian RMSEAs.

Estimating a Bayesian RMSEA

Under the assumption that $n \times p$ data are multivariate normal, the $p \times p$ sample covariance matrix, $\mathbf{S}$, is a Wishart-variable: $\mathbf{S} \sim \mathcal{W}_p(n^*, \mathbf{\Sigma}/n^*)$, where $n^* = n - 1$ and $\mathbf{\Sigma}$ is the population covariance matrix underlying the data. Wu and Browne (2015) assume that $\mathbf{\Sigma}$ is specific to the conditions under which the data were collected. For example, assume the data are scale data collected from a random sample of college students at a university in the northeastern U.S. during cold winter months. However, the target of inference or the population for which the measurement model underlying the scale is hypothesized is unlikely to be this specific population in this specific time period. And this creates a difference between $\mathbf{\Sigma}$ and the true population covariance matrix for which the theory is hypothesized to hold. Wu and Browne (2015) characterize this difference as *adventitious error*. In this sense, adventitious error is random error that produces a population covariance matrix that is specific to a population and the measurement conditions under which the data were collected from the population. Any investigation analyzing the same instrument in a different context will similarly be subject to adventitious error, resulting in a new population covariance matrix for that context.

Wu and Browne (2015) assumed the following data generating mechanism for population covariance matrix: $\mathbf{\Sigma} \sim \mathcal{W}_p^{-1}(m, m \times \mathbf{\Omega}(\boldsymbol{\theta}))$, where $m(> p - 1)$ is the degrees of freedom of the inverse-Wishart distribution and $\mathbf{\Omega}(\boldsymbol{\theta})$ is a structured covariance matrix representing a hypothesized model. This model is assumed to hold in the true idealized population under general measurement conditions. As $m \to \infty$, $\mathbf{\Sigma} \to \mathbf{\Omega}(\boldsymbol{\theta})$ and the effect of adventitious error disappears. In this sense, m^{-1} is proportional to adventitious error. A key result in Wu and Browne (2015) is that the quantity, $\varepsilon = (m - p + 1)^{-\frac{1}{2}}$, approximates the RMSEA from assuming the hypothesized model holds for the studied population.

Finally, $\mathbf{\Sigma}$ is an artefact of adventitious error, i.e. it is not of substantive interest. Conveniently, the marginal distribution of the sample covariance with $\mathbf{\Sigma}$ integrated out is of known form. It is a generalized matrix beta type-II (GMB-II) distribution:

$$\mathbf{S} \sim \mathrm{GB}_p^{\mathrm{II}}\left(\frac{n^*}{2}, \frac{m}{2}, \frac{m}{n^*}\boldsymbol{\Omega}(\boldsymbol{\theta}), \mathbf{0}_{p\times p}\right), \tag{1}$$

with log-likelihood:

$$\begin{aligned}\ln \mathcal{L} = f(p, m + n^*) - f(p, m) - f(p, n^*) + \frac{1}{2}\Bigg(\\ (n^* - p - 1)\ln|\mathbf{S}| + m\ln|\boldsymbol{\Omega}(\boldsymbol{\theta})| - (n^* + m)\ln\left|\frac{m\boldsymbol{\Omega}(\boldsymbol{\theta}) + n^*\mathbf{S}}{m + n^*}\right|\Bigg),\end{aligned} \tag{2}$$

where $f(p, x) = \ln\Gamma_p(x/2) - \frac{1}{2}\left[xp\ln(x/2) - xp\right]$, and Γ_p is the multivariate gamma function (Gupta & Nagar, 1999, definition 1.4.2). The GMB-II distribution is the same as the matrix-F distribution presented by Mulder and Pericchi (2018), which has also been used in the context of Bayesian network psychometrics (Williams & Mulder, 2020).

One can think of the model in equation 1 as a two-level hierarchical model, where m (or ε) functions similarly to the level-2 precision (or dispersion) term, and n functions similarly to level-1 precision term. Accordingly, the model relaxes the local independence assumption, by permitting all indicators to have residual correlations. To see this, note that the posterior mode of $\boldsymbol{\Sigma}$ is $(m\boldsymbol{\Omega}(\boldsymbol{\theta}) + n^*\mathbf{S})\big/(m + n^*)$, i.e. the population covariance matrix is a weighted mean of the structured covariance matrix and the sample covariance matrix. Unlike standard SEM models, the population covariance matrix already contains deviations from the hypothesized covariance structure. These deviations are shrunken towards zero, and the degree of shrinkage is higher when adventitious error is low relative to sampling error ($m \gg n$).

Hence, the approach here shares connections with other Bayesian approaches that relax local independence (e.g. J. Chen, 2022; Muthén & Asparouhov, 2012; Uanhoro, 2023b).

Simulation study

We conduct a simulation study to assess parameter recovery with the proposed approach. The data generation process (DGP) was:

$$\begin{aligned}
&\mathbf{S} \sim \mathrm{GB}_p^{\mathrm{II}}\left(\frac{n^*}{2}, \frac{m}{2}, \frac{m}{n^*}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\intercal + \boldsymbol{\Theta}), \mathbf{0}_{p\times p}\right), \\
&\boldsymbol{\Lambda}^\intercal = \underbrace{\begin{bmatrix}\lambda & \lambda & \cdots & \lambda\end{bmatrix}}_{p \text{ elements}},\ \boldsymbol{\Theta} = (1-\lambda^2)\mathbf{I}_p,\ m = \varepsilon^{-2} + p - 1 \\
&\lambda \in \{0.6, 0.8\},\ \varepsilon \in \{0.04, 0.08, 0.16\},\ n \in \{200, 500, 2000\},\ p \in \{5, 10, 20\}
\end{aligned} \tag{3}$$

such that the factor influence was either weak ($R^2 = 36\%$) or strong ($R^2 = 64\%$), the RMSEA, sample size and the number of indicators were ‘small’, ‘medium’ and ‘large’ resulting in 54 design conditions. We simulated 1000 datasets per condition, and analyzed the data using standard frequentist and Bayesian models as well as the correct model. Hence, we assess recovery of structural parameters and the recovery across all three models. Precisely, we assess both the relative bias and empirical coverage of the 90% interval of the loadings on average, error variance on average and the RMSEA.[1] For the standard Bayesian model, we used the recommended RMSEA in Garnier-Villarreal and Jorgensen (2020). Bayesian models were fit with Stan (Carpenter et al., 2017) via the minorbsem package (Uanhoro, 2023a) using default priors. 500 iterations were retained across 3 chains for inference. All scripts are available at `https://osf.io/dzxh7/?view_only=eb3cdf6c82fa4205a6b3ff40937ab687`.

Simulation results

Structural parameters estimates were largely unbiased across methods, Figure 1. Estimation of the RMSEA was often unbiased for 10 and 20 indicators, while the proposed

---

[1] With the DGP in equation 3, the expected values of loadings and variance parameters are scaled up by $\sqrt{\frac{m}{m-p-1}}$ and $\frac{m}{m-p-1}$ respectively. This follows from the mean of the inverse-Wishart distribution. The standard frequentist and Bayesian models attempt to the estimate expected values of parameters. Hence, we adjusted the population parameters accordingly for the standard frequentist and Bayesian models.

approach produced upwardly biased RMSEAs when the number of indicators was small, see Figure 2. With regard to coverage, the proposed approach always had adequate coverage for both structural parameters and the RMSEA. And the standard frequentist and Bayesian models produced intervals that had lower coverage than nominal, see Figure 3. Additionally, the under-coverage of the standard approaches got worse as the sample size increased (scan any row of results left-to-right in Figure 3), matching results in Wu and Browne (2015). In summary and as expected, ignoring uncertainty due to misspecification results in overly confident inference (too narrow uncertainty intervals) about structural parameters and misfit indices.

Conclusion

We have presented a Bayesian method for estimating the RMSEA by extending the model of Wu and Browne (2015) to the Bayesian context. The method behaves similarly to the original approach, i.e. the uncertainty about structural parameters reflects uncertainty due to model misspecification. One flaw in the method is the tendency to over-estimate the RMSEA when the number of indicators is low, even though structural parameters are correctly estimated in this condition. This limitation may be due to either the low number of indicators or the small nominal model degrees of freedom. We intend to examine this question in the future.

More broadly, our paper demonstrates a need to account for model misspecification during model estimation. Uncertainty about model correctness should be reflected in uncertainty about structural parameters. Broadly, the larger the sample size, the more power there is to detect misspecification. In the context of the simulation results, the failure of standard models to account for uncertainty is worsened with larger samples.

Finally, we intend to explore the size of misfit in the context of the proposed Bayesian RMSEA. The standard RMSEA is known to be influenced by several factors. Given that the proposed approach allows for identifying a population RMSEA, we intend to explore how a given RMSEA corresponds to varying degrees of model misfit given changes in the structure of the data, such as number of indicators, and the quality of measurement.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). doi: 10.18637/jss.v076.i01

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008, May). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. doi: 10.1177/0049124108314720

Chen, J. (2022, November). A generalized partially confirmatory factor analysis framework with mixed bayesian lasso methods. *Multivariate Behavioral Research*, *57*(6), 879–894. doi: 10.1080/00273171.2021.1925520

Garnier-Villarreal, M., & Jorgensen, T. D. (2020, February). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihood. *Psychological Methods*, *25*(1), 46–70. doi: 10.1037/met0000224

Gupta, A. K., & Nagar, D. K. (1999). *Matrix variate distributions.* CRC Press.

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, I. (2018, August). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement*, *78*(4), 537–568. doi: 10.1177/0013164417709314

Hu, L.-t., & Bentler, P. M. (1999, January). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi: 10.1080/10705519909540118

Levy, R. (2011, October). Bayesian data-model fit assessment for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 663–685. doi: 10.1080/10705511.2011.607723

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary* *psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mulder, J., & Pericchi, L. R. (2018, December). The matrix-F prior for estimating and testing covariance matrices. *Bayesian Analysis*, *13*(4), 1193–1214. doi: 10.1214/17-BA1092

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. doi: 10.1037/a0026802

Savalei, V. (2012, December). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, *72*(6), 910–932. doi: 10.1177/0013164412452564

Uanhoro, J. O. (2023a, June). minorbsem: An R package for structural equation models that account for the influence of minor factors. *Journal of Open Source Software*, *8*(86), 5292. doi: 10.21105/joss.05292

Uanhoro, J. O. (2023b, April). Modeling misspecification as a parameter in Bayesian structural equation models. *Educational and Psychological Measurement*, *0*(0). doi: 10.1177/00131644231165306

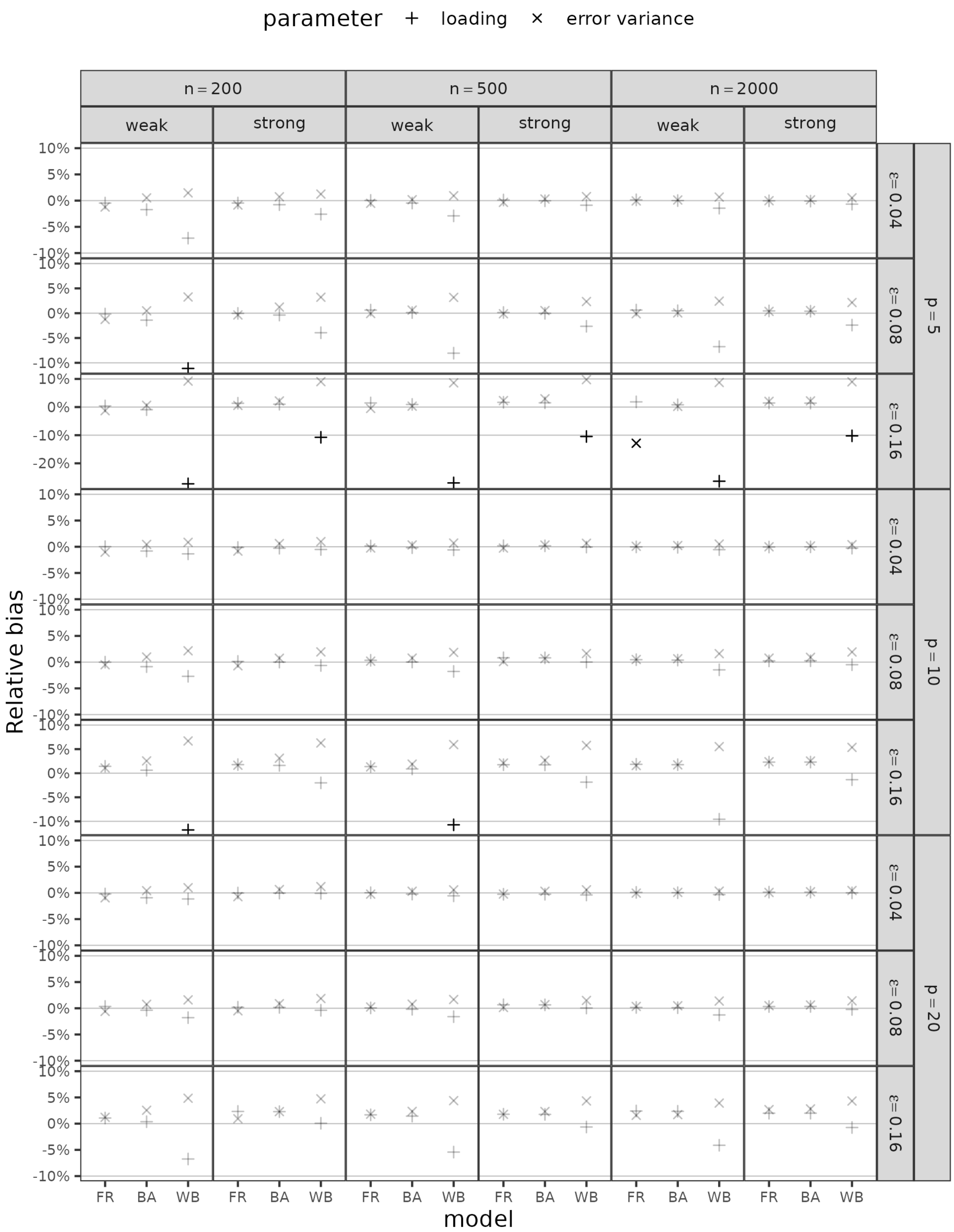
West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (Vol. 1, pp. 209–231). Guilford Press. (Publisher: New York)

Williams, D. R., & Mulder, J. (2020, December). Bayesian hypothesis testing for Gaussian graphical models: Conditional independence and order constraints. *Journal of Mathematical Psychology*, *99*, 102441. doi: 10.1016/j.jmp.2020.102441

Wu, H., & Browne, M. W. (2015, September). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*(3), 571–600. doi: 10.1007/s11336-015-9451-3

Figure 1

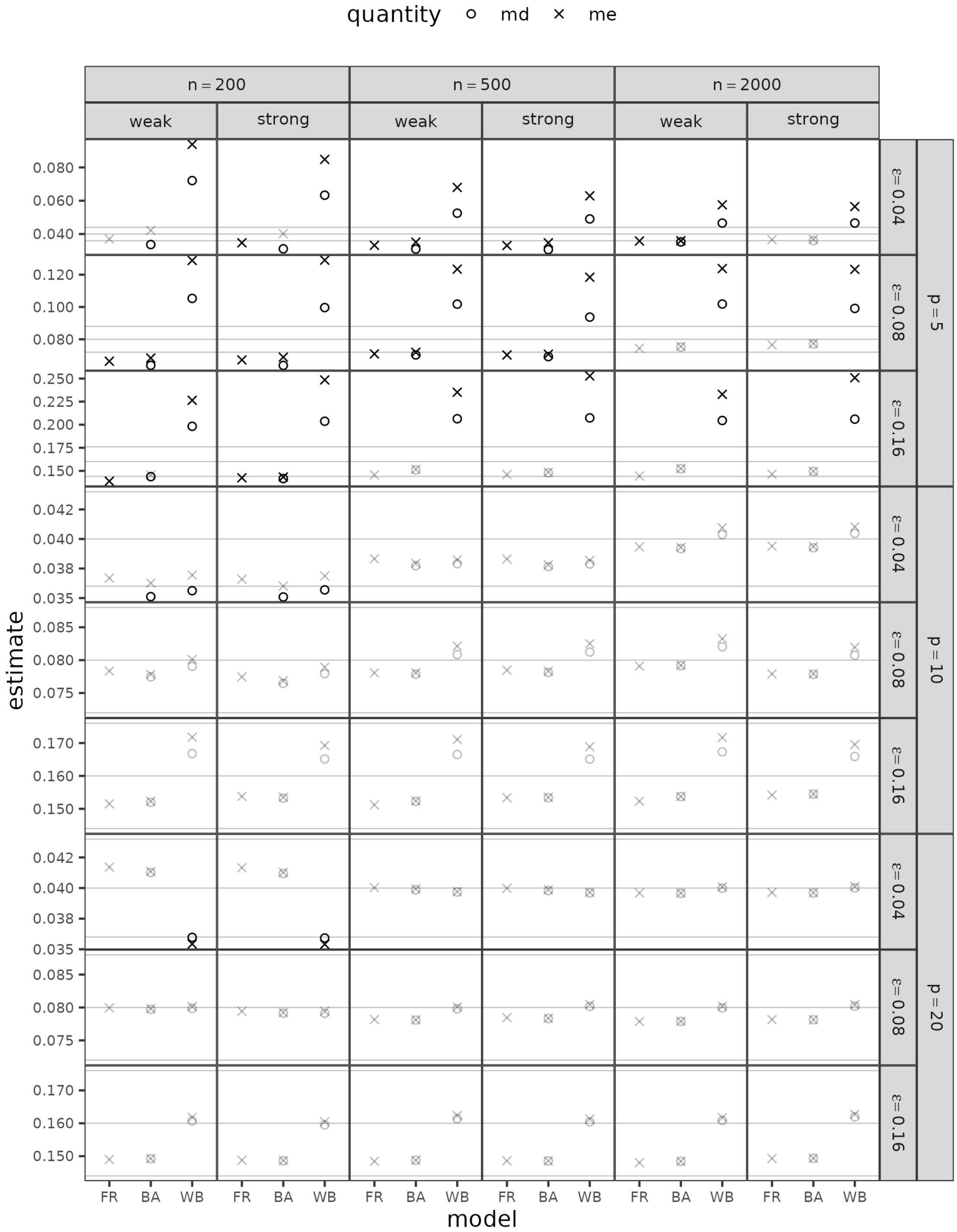
Relative bias of structural parameters



Note. FR = Standard frequentist, BA: Standard Bayesian, WB: Bayesian translation of Wu and Browne (2015). ‘weak’ and ‘strong’ refer to strength of factor loadings. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal expectation. Horizontal dashed lines represent the liberal bounds.

Figure 2

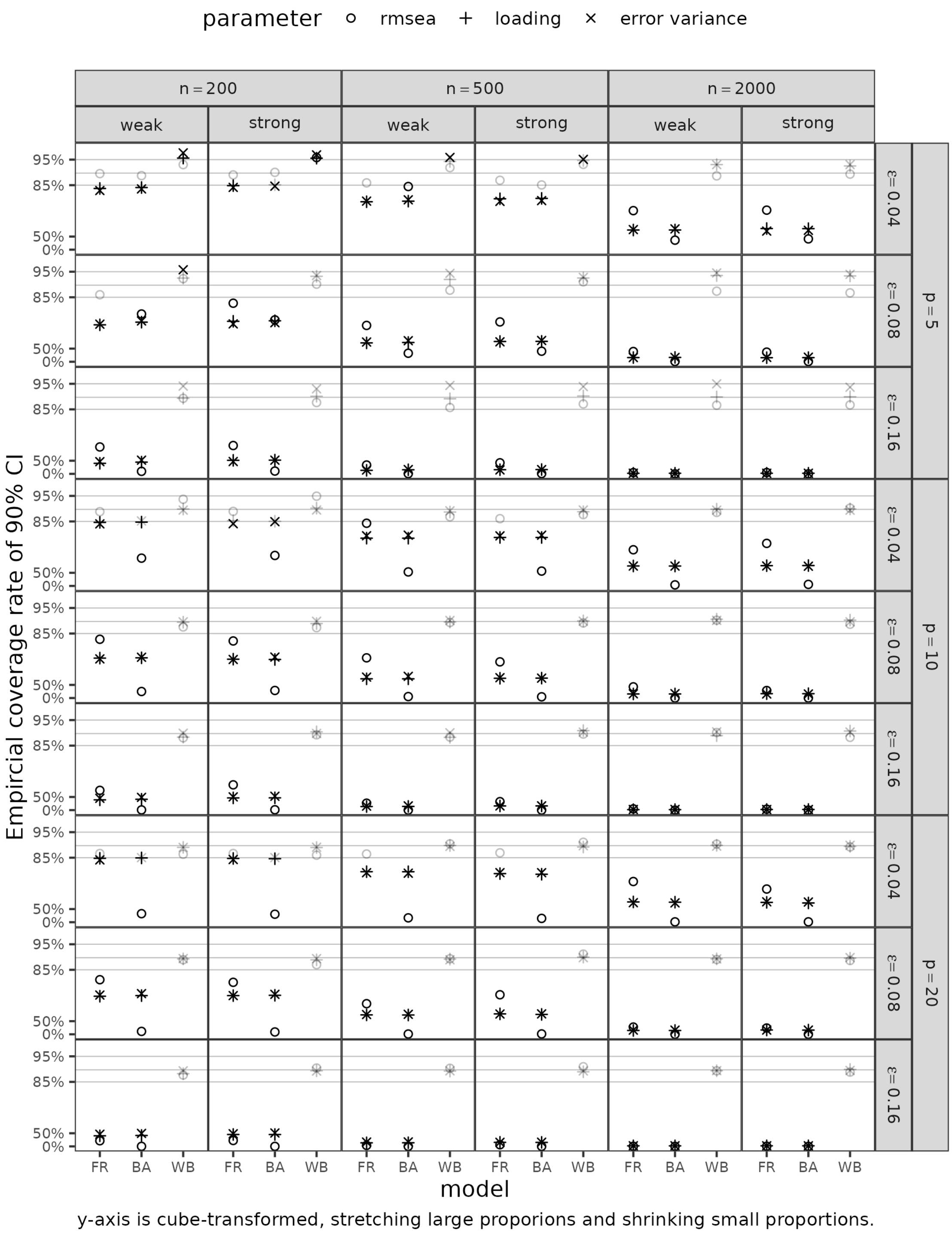
Relative bias of RMSEA



Note. FR = Standard frequentist, BA: Standard Bayesian, WB: Bayesian translation of Wu and Browne (2015). 'weak' and 'strong' refer to strength of factor loadings. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal expectation. Horizontal dashed lines represent the liberal bounds.

Figure 3

Empirical coverage of 90% interval



Note. FR = Standard frequentist, BA: Standard Bayesian, WB: Bayesian translation of Wu and Browne (2015). ‘weak’ and ‘strong’ refer to strength of factor loadings. Ideally, coverage estimates fall within ±2.5% of 90%, ±5% is a more liberal expectation. Horizontal dashed lines represent the liberal bounds.