Modeling misspecification as a parameter in Bayesian structural equation models

James Ohisei Uanhoro
Research, Measurement & Statistics, Department of Educational Psychology
University of North Texas

Abstract

Accounting for model misspecification in Bayesian structural equation models is an active area of research. We present a uniquely Bayesian approach to misspecification that models the degree of misspecification as a parameter – a parameter akin to the correlation root mean squared residual. The misspecification parameter can be interpreted on its own terms as a measure of absolute model fit, and allows for comparing different models fit to the same data. By estimating the degree of misspecification simultaneously with structural parameters, the uncertainty about structural parameters reflect the degree of model misspecification. This results in a model that produces more reliable inference than extant Bayesian SEMs. Additionally, the approach estimates the residual covariance matrix which can be the basis for diagnosing misspecifications and updating a hypothesized model. These features are confirmed using simulation studies. Demonstrations with a variety of real world examples show additional properties of the approach.

Keywords: model misspecification, Bayesian SEM, CRMR

Structural equation modeling (SEM) is a popular statistical method for modeling covariance matrices. SEMs usually return a set of structural parameters ($\boldsymbol{\theta}$) that are configured in a way that is substantively interesting to the data analyst, e.g. the relation between observed indicators and unobserved factors in confirmatory factor models or directional relations between latent factors in latent regression models. Theoretically, this configuration is simply a structured covariance matrix ($\boldsymbol{\Sigma}(\boldsymbol{\theta})$ – $\boldsymbol{\theta}$ being a substantively interesting parameter vector) which is often different from the true covariance matrix, $\boldsymbol{\Sigma}$. Restated, the information in the true covariance matrix can be summarized by $\boldsymbol{\theta}$ configured in a way as to have substantively interesting interpretations – SEMs are an attempt at data reduction. If the data reduction is to be believed, then $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ should not be too distinct from

The Version of Record of this manuscript has been accepted at Educational and Psychological Measurement (2023) https://dx.doi.org/10.1177/00131644231165306. All code for simulation studies, data analyses, Stan scripts and supplementary materials are available at https://osf.io/29ydb/

$\boldsymbol{\Sigma}$. It is for this reason that model misspecification plays a big role in SEMs, and the gap between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ often forms the basis for SEM goodness of fit indices, of which there are many.

Bayesian SEMs (BSEM, Levy & Mislevy, 2016; Song & Lee, 2012) are an approach to SEMs formulated on the basis of Bayesian probability that often rely on Markov chain Monte Carlo (MCMC) methods to approximate the posterior distribution of $\boldsymbol{\theta}$. BSEMs are increasingly popular for a wide variety of reasons: a general increase in the accessibility and use of Bayesian methods, and for a number of advantages BSEMs have over traditional or frequentist SEMs – the most commonly cited being the ability to estimate $\boldsymbol{\theta}$ even when there is limited information in the data to estimate $\boldsymbol{\theta}$ (e.g. Dunson, 2000; Lee, Song, & Tang, 2007; Scheines, Hoijtink, & Boomsma, 1999).

However, given the relative recency of BSEMs, model misspecification is an area of active research, and a number of traditional fit indices have been adapted to the Bayesian context. Posterior predictive p-values (Levy, 2011) are akin to the frequentist χ^2 test of absolute fit in a Bayesian context. Levy (2011) also proposed a Bayesian standardized root mean squared residual (SRMR) – an *effect-size* style fit index. Hoofs, van de Schoot, Jansen, and Kant (2018) noted that posterior predictive p-values may be sensitive to trivial forms of misspecification in large samples and developed a Bayesian variant of the root mean square error of approximation (RMSEA) for use in large samples. And the work of Garnier-Villarreal and Jorgensen (2020) is the most extensive adaption of several frequentist fit indices for BSEM. Finally, generic information-criteria based fit indices are also used for comparing different BSEMs fit to the same data (Cain & Zhang, 2019; Merkle & Rosseel, 2018).

In this paper, we present an approach to BSEMs with saturated mean structure that models the degree of model misspecification as a parameter. Although modeling misspecification as a parameter has been accomplished in the frequentist literature (via modeling the RMSEA, Wu & Browne, 2015), our approach is uniquely Bayesian in that it leverages Bayesian techniques and approaches to quantifying uncertainty. The approach permits: (i) investigation of model fit in absolute terms; (ii) model comparisons; and (iii) investigation of potential modifications to the model. The approach accomplishes (i) by returning a misspecification parameter that is similar to the correlation root mean squared residual (CRMR, Ogasawara, 2001), which can be substantively interpreted on its own terms (Maydeu-Olivares, 2017). Model comparisons may then be performed

using the posterior distribution of the returned fit index. Finally, one can investigate potential model modifications by focusing on the residual covariance matrix estimated by the approach.

Similarly to Wu and Browne (2015), we are not simply intent on presenting another approach for evaluating the fit of SEMs. Our approach models misspecification simultaneously with structural parameters, such that uncertainty due to model misspecification is then reflected in uncertainty about structural parameters. This aspect of our work is an advantage over other BSEMs which ignore this source of uncertainty. Practically, our approach will often result in structural parameters with wider intervals than a more typical BSEM.

In the next section, we elaborate the approach we recommend. Afterward, we conduct simulation studies that evaluate the validity of the approach, and compare the approach to extant methods. Then we demonstrate the approach with three datasets to further explore the behaviour of the approach, and conclude with a discussion and steps for further development.

All code for simulation studies, data analyses, Stan scripts and supplementary materials are available at `https://osf.io/29ydb/`.

Misspecification as a model parameter

Wu and Browne (2015) developed a frequentist approach for modeling misspecification as a parameter such that uncertainty about structural parameters ($\boldsymbol{\theta}$) reflects model misspecification. Ideally, their work would be a starting point for a similar Bayesian endeavour as ours. However, we approach the same goal somewhat differently primarily because their approach estimates the RMSEA as a measure of model misfit. We prefer a different approach due to challenges with interpreting the RMSEA (e.g. Chen, Curran, Bollen, Kirby, & Paxton, 2008; Savalei, 2012). That said, if a researcher is primarily interested in having uncertainty about structural parameters reflect the degree of model misspecification, then the approach of Wu and Browne (2015) should be sufficient and is readily extendable to the Bayesian context.

We lay out our proposal using the confirmatory factor analysis (CFA) model, though the proposal is valid for SEMs where the covariance matrix is sufficient. Our starting point is a Bayesian CFA model presented by Muthén and Asparouhov (2012), hereafter MA (2012):

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Psi} + \boldsymbol{\Delta} \tag{1}$$

where $\mathbf{\Lambda}$ is the loading matrix with some elements set to 0 on the basis of theoretical considerations and model identification constraints, $\mathbf{\Phi}$ is the inter-factor correlation matrix, $\mathbf{\Delta}$ is a diagonal matrix of residual variances, and $\mathbf{\Psi}$ is a full residual covariance matrix. Estimating $\mathbf{\Psi}$ without regularization will cause the model to be under-identified. MA (2012) estimate $\mathbf{\Psi}$ with a highly informative inverse-Wishart prior on $\mathbf{\Psi}$ that implies $\boldsymbol{\psi}$ – the vector of off-diagonal elements in $\mathbf{\Psi}$ – has a mean of 0 and a very small variance.

Substantively, this model assumes that major theoretical factors cause observed indicators to be correlated in a configuration assumed known a-priori ($\mathbf{\Lambda\Phi\Lambda}^{\mathsf{T}}$), while minor factors cause indicators to be trivially correlated in ways assumed unknown a-priori, $\mathbf{\Psi}$. This model is practically useful because it relaxes the assumption of local independence – often unrealistic for real world data – while permitting simple structure. Under a sound theory for observed indicators, the correlations induced by the minor factors should be trivial suggesting that a simple structure is indeed reasonable for the indicators. When the correlations induced by minor factors are non-trivial, the minor factors are not *minor* casting doubt over the hypothesized configuration of the observed indicators.

The preceding paragraph forms the basis for our approach. In MA (2012)'s original formulation, the modeler sets the variance of $\boldsymbol{\psi}$ by specifying a known inverse-Wishart prior on $\mathbf{\Psi}$. In our formulation, we assume this variance unknown, and attempt to learn the variance from the available data.

Prior strategies for $\mathbf{\Psi} + \mathbf{\Delta}$

$\mathbf{\Psi} + \mathbf{\Delta}$ forms the residual covariance matrix, so options for identifying $\mathbf{\Psi}$ must similarly address estimation of $\mathbf{\Delta}$. We consider two approaches for estimating both matrices. Additionally, we set priors for the data under the assumption that items are scaled with total variance of about one.

Method 1: Separation strategy

Let $\mathbf{\Psi} + \mathbf{\Delta} = \mathbf{DRD}$ such that $\mathbf{D}$ is a diagonal matrix of residual standard deviations and $\mathbf{R}$ is the residual correlation matrix. We model both $\mathbf{D}$ and $\mathbf{R}$ separately (*separation strategy,* Barnard, McCulloch, & Meng, 2000) which has the benefit over Wishart-type priors of setting the priors differently for the residual standard deviations and the residual correlations.

Prior distributions. Under this strategy, we assume that $\sqrt{\text{diag}(\mathbf{D})} \sim t^{+}(3, 0, 1)$, i.e. the residual standard deviations have a weakly-informative (Gelman, Jakulin, Pittau, & Su, 2008; Lemoine, 2019) half-t prior. And we assume $\mathbf{R} \sim \text{LKJ}(\eta)$, i.e. the correlation matrix is LKJ-distributed (Lewandowski, Kurowicka, & Joe, 2009) with shape parameter $\eta \in (0, \infty)$.[1] Accordingly, the marginal distribution of each residual correlation, r_{ij} in $\mathbf{R}$ is $\frac{r_{ij}+1}{2} \sim \text{beta}\left(\eta - 1 + {}^{p}/_{2}, \eta - 1 + {}^{p}/_{2}\right)$, where p is the number of items in $\mathbf{R}$. Practically, the residual correlations are assumed to have a mean of 0 and standard deviation, $\tau_r = \frac{1}{\sqrt{2\eta + p - 1}}$, i.e. $\mathbf{R} \to \mathbf{1}_{p\times p}$ as $\eta \to \infty$. The main parameter to be learned from the data is η, and we intend for it to be free to be as large as possible. Hence we assume ${}^{1}/_{\eta} \sim \mathcal{N}^{+}(0, 1)$, such that there is sufficient prior probability on very large values of η.

A standardized metric for model fit. τ_r helps assess model fit as it is the root mean squared error (RMSE) of residual correlations with the location assumed to be 0. As $\tau_r \to 0$, we can assume that residual correlations are increasingly trivial. However, τ_r is on the scale of residual standard deviations. To be able to use τ_r to compare different models fit to the same data, τ_r must be on the scale of the total variance. Accordingly, we rescale τ_r: $\tau_r'' = \tau_r \times \sqrt{\frac{1}{p(p-1)} \sum_{i=2}^{p} \sum_{j=1}^{i-1} d_{jj}^2 d_{ii}^2}$, where $d_{ii/jj}$ is the i/j-th diagonal element of $\mathbf{D}$, such that τ_r'' is the RMSE of residual covariances. Finally, to create a standardized metric of model fit, we standardize τ_r'': $\tau_r' = \tau_r'' \Big/ \sqrt{\frac{1}{p(p-1)} \sum_{i=2}^{p} \sum_{j=1}^{i-1} \sigma_{jj}\sigma_{ii}}$, where $\sigma_{ii/jj}$ is the i/j-th diagonal element of $\mathbf{\Sigma}(\boldsymbol{\theta})$. Hence, τ_r' is the root mean squared error of standardized residual covariances (SRCs).

Method 2: Hierarchical estimation of residual covariances

As an alternative, a simple approach for estimating residual covariances in $\mathbf{\Psi}$, ψ_{ij}, is to standardize the residual covariances and assume the standardized residual covariances (SRCs) to be normally distributed: $\frac{\psi_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}} \sim \mathcal{N}(0, \tau_{\psi}')$, where $\sigma_{ii/jj}$ is the i/j-th diagonal element of $\mathbf{\Sigma}(\boldsymbol{\theta})$ and the scale parameter, τ_{ψ}', is learned from the data: $\tau_{\psi}' \sim \mathcal{N}^{+}(0, 1)$.[2] Hence, τ_{ψ}' functions as the RMSE of SRCs with location assumed to be 0, and the prior for τ_{ψ}' is weakly-informative since SRCs will often be much lower than 1. We set the diagonal of $\mathbf{\Psi}$ to $\mathbf{0}_p$, and similar to the first method, we assume that the residual standard deviations are half-t: $\sqrt{\text{diag}(\mathbf{\Delta})} \sim t^{+}(3, 0, 1)$.

[1] When estimating this model using MCMC, it is useful to initialize $\mathbf{R}$ as $\mathbf{1}_{p\times p}$, otherwise the sampler may fail to adopt a positive-definite $(\mathbf{\Psi} + \mathbf{\Delta})$ matrix.

[2] When estimating this model using MCMC, it is useful to initialize the residual covariances at a number close to 0, otherwise the sampler may fail to adopt a positive-definite $(\mathbf{\Psi} + \mathbf{\Delta})$ matrix.

The rationale for this approach is that the SRCs are on average 0, but each SRC may differ from 0 in continuous gradations. Hence, this approach can be seen as a ridge penalty (Park & Casella, 2008) for SRCs where τ'_ψ is the shrinkage parameter. The connection to ridge regression opens up the possibility of other priors for SRCs. For example, sparsity may be desirable under the assumption that some SRCs are indeed 0 leading to alternative priors for SRCs such as double-exponential (lasso, Park & Casella, 2008), or global-local approaches that constrain some SRCs to 0 while allowing other SRCs to escape this constraint (e.g. horseshoe, Carvalho, Polson, & Scott, 2009). Broadly, Bayesian estimation accommodates a variety of options for estimating SRCs. We believe the assumption that SRCs have a center of 0 with continuous gradations away from 0 is most realistic, hence we maintain the normal distribution for SRCs above.

Similarities and differences between both methods

Both methods reflect an important observation about SEMs. It is commonplace in the SEM literature to assume that the population covariance matrix underlying a study is equivalent to a structured covariance matrix, i.e. $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^\mathsf{T} + \mathbf{\Delta}$. Misspecified SEMs are then generated by "wrongly" setting some structural parameters to incorrect values, e.g. assuming a cross-loading of 0.3 is 0. Differently, both methods above assume that misspecification (due to minor factors) is already present at the level of the population covariance matrix, i.e. $\mathbf{\Sigma}$ includes $\mathbf{\Psi}$ – MacCallum and Tucker (1991) suggested such a formulation for SEMs. Hence, even with a correct hypothesized structure, there is still error separate from sampling error. In both methods, $\mathbf{\Psi}$ contains this error and is determined by a parameter – η in method 1; τ'_ψ in method 2. The key gain over MA (2012) is that this parameter can be estimated from an observed covariance matrix, such that the parameter provides guidance on the magnitude of the effect of minor factors. Additionally, different studies with identical values of either η or τ'_ψ will have different residual covariances at the population level, i.e. the exact error due to minor factors is random, as opposed to being fixed.

Another similarity between both methods is that they produce fit indices, τ'_r and τ'_ψ or generically τ', that are similar to the CRMR. Although τ' focuses on standardized residual covariances (like the SRMR), τ' ignores the diagonal of the gap between the population covariance matrix and model-implied covariance matrix. In that respect, τ' is more similar to the CRMR.

With regard to differences, the first approach based on the separation strategy represents a

more credible data generation process~~, as it guarantees positive-definite population covariance matrices~~,[3] while the second approach will generate non positive-definite matrices as τ'_ψ gets increasingly large. However, that the second approach is not always credible for data generation does not imply it has no value for estimation.

Incorrect, part (i) in footnote is not pd.

Proposed Bayesian models

We now lay out the Bayesian models we use throughout this manuscript unless otherwise specified, and explain the model in the paragraphs that follow:

$$
\begin{aligned}
\mathbf{S} \sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\boldsymbol{\Sigma}\right),\ \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^\mathsf{T} + \boldsymbol{\Omega},\ \text{where } \boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\Delta} \\
\boldsymbol{\lambda} \sim \mathcal{N}(0, \tau_\lambda),\ \tau_\lambda \sim t^+(3,0,1),\ \boldsymbol{\Phi} \sim \text{LKJ}(1), \\
\sqrt{\text{diag}(\boldsymbol{\Omega})} \sim t^+(3,0,1),\ \frac{\delta_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \sim \text{Beta}(2,2) \times 2 - 1, \\
\text{Method 1:}\ (\boldsymbol{\Psi} + \boldsymbol{\Delta}^*) = \mathbf{DRD},\ \mathbf{R} \sim \text{LKJ}(\eta),\ \frac{1}{\eta} \sim \mathcal{N}^+(0,1) \\
\text{Method 2:}\ \text{diag}(\boldsymbol{\Psi}) = \mathbf{0}_p,\ \frac{\psi_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \sim \mathcal{N}(0, \tau'_\psi),\ \tau'_\psi \sim \mathcal{N}^+(0,1)
\end{aligned}
\tag{2}
$$

where $\mathbf{S}$ is the sample covariance matrix assumed Wishart with degrees of freedom set to sample size $(n) - 1$, and scale matrix assumed to be the population covariance matrix, $\boldsymbol{\Sigma}$, rescaled by the degrees of freedom. $\boldsymbol{\lambda}$ are non-structurally zero elements of $\boldsymbol{\Lambda}$, with a scaling parameter that is learned from the data. The residual covariance matrices $(\boldsymbol{\Psi}, \boldsymbol{\Delta})$ are summed to $\boldsymbol{\Omega}$. The root of the diagonal of $\boldsymbol{\Omega}$ represents residual standard deviations and are assumed half-t. Off-diagonal elements in $\boldsymbol{\Delta}$, δ_{ij}, represent purposely-specified or hypothesized residual covariances. δ_{ij} are standardized using the residual standard deviations $\left(\sqrt{\omega_{ii/jj}}\right)$, such that any hypothesized residual correlations are assumed beta-distributed with a boundary avoiding prior. Hence, our proposed models allow for both the a-priori specification of known residual covariances in $\boldsymbol{\Delta}$, and the influence of minor factors in $\boldsymbol{\Psi}$.

Permitting off-diagonal elements in both $\boldsymbol{\Psi}$ and $\boldsymbol{\Delta}$ complicates method 1 slightly. Precisely, we permit off-diagonal residual covariances in $\boldsymbol{\Delta}$ via parameter expansion (Merkle & Rosseel, 2018;

[3] ~~The sum of two positive-definite matrices: $(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^\mathsf{T})^{\text{(i)}} + (\boldsymbol{\Psi} + \boldsymbol{\Delta})^{\text{(ii)}}$, is itself positive-definite.~~

Palomo, Dunson, & Bollen, 2007), such that $\boldsymbol{\Delta}^*$ is a diagonal matrix of residual variances after accounting for hypothesized residual covariances in $\boldsymbol{\Delta}$. The remaining parameters and priors for methods 1 and 2 remain unchanged from their earlier presentation. Taken all together, these priors are mostly weakly informative (Lemoine, 2019) for indicators with total variances close to 1.

Guidelines for model evaluation

The methods return τ' which is similar to the CRMR. In the instance that the mean of SRCs is not 0, e.g. the modeler fixes loadings such that the model implied covariances are downwardly biased, both parameters estimate the RMSE of SRCs not the standard deviation of SRCs, such that such bias would be reflected in the returned index.

In the remainder of the paper, we follow guidelines in section 10 of Maydeu-Olivares (2017) for describing the magnitude of SRCs. If all SRCs are in ± 0.05, we decide the SRCs are negligible. And we classify SRCs exceeding the ± 0.1 interval as non-trivial. We also acknowledge that the data analyst is free to set their benchmark in a given application with justification. We now identify some uses for τ' with regard to model misspecification.

Benchmark for acceptable model fit. When $\tau' < 0.025$, most ($\approx 95\%$ of) SRCs will be in the ± 0.05 interval suggesting that most SRCs are negligible. When $\tau' < 0.05$, most SRCs will be in the ± 0.1 interval suggesting that most SRCs are small. Although these approximation are based on the normal distribution (method 2), the normal distribution reasonably approximates the beta distribution (method 1) when the beta shape parameters are equal (assumed for method 1) and exceed 10 – a condition that will hold in real world most datasets.

Relative model comparisons. So far, we have suggested that larger values of τ' mean that minor factors account for more of the relation between indicators. However, a hypothesized model could also be incorrect such that τ' reflects both the influence of minor factors and misspecification in the structured portion of the covariance matrix. Thus two models fit to the same data would differ in τ' if one model better reflects the true structured covariance matrix underlying the data, with the more correct hypothesized structure having the lower τ'. Differences ($\tau'_{.1} - \tau'_{.2}$) or ratios ($\tau'_{.1}/\tau'_{.2}$) for two models alongside credible intervals (based on the posterior distributions) provide an effect size of differences in model misspecification.

In addition to τ', examining $\boldsymbol{\Psi}$ allows for detecting large residual covariances i.e. relations

where "minor" factors play an out-sized role. First, we standardize this matrix to ease interpretations: $\boldsymbol{\Psi}' = \mathbf{D}^*\boldsymbol{\Psi}\mathbf{D}^*$, where $\mathbf{D}^*$ is a diagonal matrix with elements: $\sigma_{ii}^{-1/2}$, such that $\boldsymbol{\Psi}'$ is the SRC matrix. SRCs whose 90% (or 95%) credible intervals (CI) are fully contained within ± 0.1 may be judged to be trivially influenced by minor factors. SRCs whose 90% (or 95%) CI exclude 0 may be judged to be influenced by minor factors, though the influence may still be trivial depending on the estimate of the SRC. Absolute SRC estimates larger than 0.1 pose a challenge to the theory. SRCs with 90% intervals that both exceed ± 0.1 and contain 0 lead to inconclusive state – they are neither clearly small nor distinct from 0.

When SRCs cannot be judged trivial, it is necessary to identify the cause of the non-trivial SRCs. Clusters of SRCs may point to modifications of the hypothesized structure. See Saris, Satorra, and van der Veld (2009) for a discussion of next steps on identifying model misspecifications.

We now present results from simulation studies to test the validity of the proposed approaches.

Simulation studies

We conducted three simulation studies and summarize their results. In study 1, we compared our proposed approach to extant approaches (standard BSEM, approximate-zero (Muthén & Asparouhov, 2012)) under a relatively simple data generation process that assumed varying levels of influence of minor factors on the population covariance matrix. We found that the different approaches yielded negligible parameter bias for structural parameters, $\boldsymbol{\theta}$. However, only the proposed approaches correctly quantified uncertainty about $\boldsymbol{\theta}$ across conditions. Based on the results of this study, we opted for the hierarchical estimation of residual covariances as our preferred method in the remainder of the paper.

In study 2, we further examined our proposed approach under a more typical data generation process that assumed varying levels of influence of minor factors on the population covariance matrix. Additionally, we were interested in the ability of τ' to select between a correct and a wrong model (that wrongly assumed unidimensionality of a two-factor model). We found that our proposed approach yielded adequate parameter recovery (bias and inference) for $\boldsymbol{\theta}$ and τ' was capable of distinguishing between a correct and a wrong model especially for larger samples.

Finally in study 3, we assessed the ability of the proposed approach to identify overly large residual covariances in $\mathbf{\Delta}$ in the data generating process that were not pre-specified by the modeler in $\mathbf{\Delta}$. The proposed approach was able to identify such residual covariances in $\mathbf{\Psi}$, thus the approach is capable of identifying notable missing residual covariances. Moreover, the proposed approach is capable of *accurately* estimating hypothesized residual covariances in $\mathbf{\Delta}$ simultaneously with $\mathbf{\Psi}$.

With each simulation study, each design condition was replicated 1000 times. The MCMC engine was Stan (Carpenter et al., 2017); for each model, there were 1000 warmup iterations, then 1000 iterations were retained for inference across 3 chains. We retrieved $\widehat{R}$ (Vehtari, Gelman, Simpson, Carpenter, & Bürkner, 2020) – it exceeded 1.05 less than 1% of the time for all parameters in each design condition for each model within each study suggesting MCMC convergence for parameters across almost all replications across studies.

Simulation study 1

Data generation process

We set out to demonstrate the relative advantage of our approach over extant SEM estimation approaches. We assumed the separation-strategy as the true model for the data since this model is both realistic – it includes the influence of minor factors – and produces positive-definite covariance matrices. The data generation process (DGP) assumed parallel indicators for each factor:

$$\begin{aligned} \mathbf{S} &\sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\mathbf{\Sigma}\right),\ \mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^\mathsf{T} + \mathbf{\Omega} \\ \mathbf{\Lambda}^\mathsf{T} &= \left[\begin{smallmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{smallmatrix}\right],\ \mathbf{\Phi} = \left[\begin{smallmatrix} 1 \\ .3 & 1 \end{smallmatrix}\right] \\ \mathbf{\Omega} &= \mathbf{DRD},\ \mathbf{D} = 0.6 \times \mathbf{I}_{p\times p},\ \mathbf{R} \sim \text{LKJ}(\eta) \end{aligned} \tag{3}$$

where $p = 10$, there were two correlated factors, the true total variance of indicators was 1, and the input data for analysis was $\mathbf{S}$. As can be seen from the DGP, the exact residual covariances are random, i.e. vary across studies given a set level of misspecification represented by η.

Conditions

We varied two factors in this study: (i) sample size, n, by small (100), typical (300), large (1000), and very large (5000); (ii) the size of minor factor influences or misspecification parameter, $\tau' \in \{0.025, 0.05, 0.08\}$. Recalling the heuristic that most SRCs ($\approx 95\%$) will fall in $\pm 2\tau'$, these choices for τ' correspond to models where most SRCs are negligible (within ± 0.05), acceptable (within ± 0.1), and several are non-trivial (exceeding ± 0.1). Based on the relations in the separation-strategy section, $\eta = \frac{1}{2}\left[\left(\frac{0.6^2}{\tau'_r \times 1^2}\right)^2 - p + 1\right]$, i.e. $\eta \in \{99.18, 21.42, 5.625\}$ corresponds to $\tau' \in \{0.025, 0.05, 0.08\}$.

Models

We estimated the following models within each replication:

1. A standard frequentist model: parallel-indicators CFA with all residual covariances assumed null ($\mathbf{\Psi} = \mathbf{0}_{p \times p}$).

2. A standard Bayesian model (hereafter *baseline*): the corresponding Bayesian model with priors: $\lambda \sim \mathcal{N}(0, 1)$, $\sqrt{\delta} \sim t^+(3, 0, 1)$, $\mathbf{\Phi} \sim \text{LKJ}(1)$, where δ is the residual variance.

3. An approximate-zero (hereafter *AZ*) model: tau-equivalent CFA using the framework defined by MA (2012). This model had the same priors as the *baseline* model. The extra parameter, $\mathbf{\Psi}$, which estimates the influence of minor factor, was assumed to be inverse-Wishart with degrees of freedom set to $p + 6$, which implies a known standard deviation for residual covariances of about 0.1. Although we constrained the diagonal of $\mathbf{\Delta}$ to be identical, this model cannot have parallel-indicators because there is no way to constrain the diagonal of $\mathbf{\Psi}$ to be identical across items using the inverse-Wishart distribution.

4. The separation-strategy (hereafter *LKJ*) approach which corresponds to the DGP, with method specific priors as in line *Method 1* of equation 2 and shared priors as in the *baseline* model.

5. The hierarchical estimation of residual covariances (hereafter *normal*) approach, with method specific priors as in line *Method 2* of equation 2 and shared priors as in the *baseline* model.

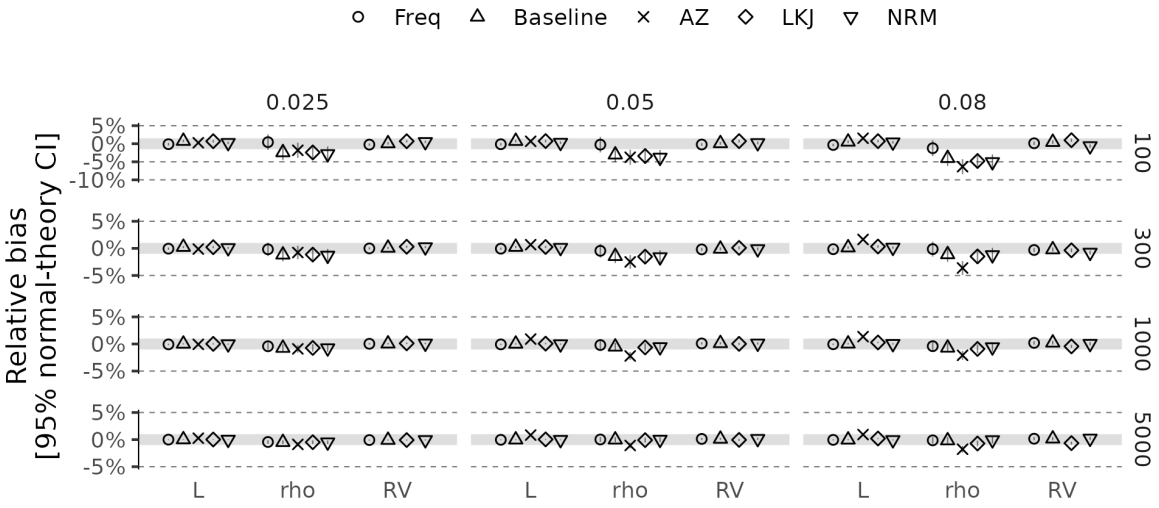
Estimands and performance metrics

We were interested in the recovery of the structural parameters: $\lambda, \delta, \phi_{12}$, and the misspecification parameter, τ' – we assessed all these parameters for the LKJ and normal models. For the frequentist model, we assessed all the structural parameters and computed the unbiased CRMR (Maydeu-Olivares, 2017). For the baseline model, we assessed all the structural parameters and computed the root mean squared error of realized SRCs using the approach laid out by Levy (2011). For the AZ approach, we assessed only λ and ϕ_{12} since the residual variance was split between two matrices. There is no prescribed RMR-type metric for this model.

We had two primary assessment metrics: bias and the empirical coverage rate (ECR) of the 90% confidence interval. We transformed bias to relative bias deeming relative bias within ±5% as ideal and ±10% as acceptable. We assessed empirical coverage rate of the 90% credible intervals. We set (87.5%, 92.5%) and (85%, 95%) as ideal and acceptable limits for the 90% ECR respectively.

Figure 1

Study 1: Relative bias of structural parameters



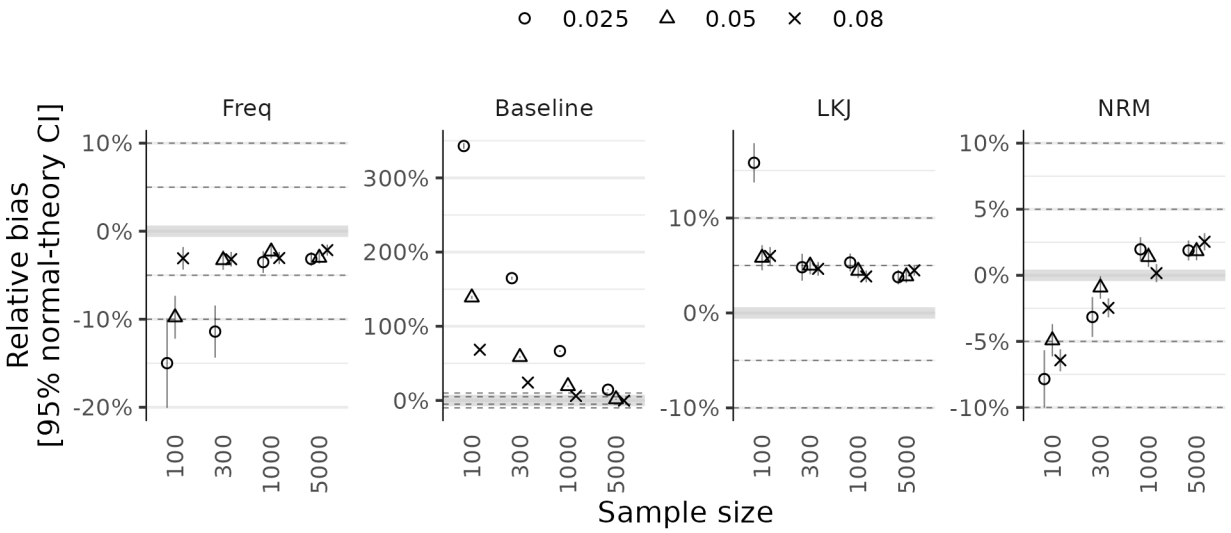
Note. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal condition. Horizontal dashed lines represent these bounds. L is the loading; RV: residual variance; rho is the interfactor correlation.

Simulation study 1 results

Relative bias for structural parameters. As shown in Figure 1, relative bias was ideal (±5%) for almost all parameters and acceptable (±10%) for all parameters across conditions and

Figure 2

Study 1: Relative bias of τ'



Note. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal condition. Horizontal dashed lines represent these bounds.

methods. The interfactor correlation was very slightly downwardly biased at small sample size – this bias was negligible at large sample size.

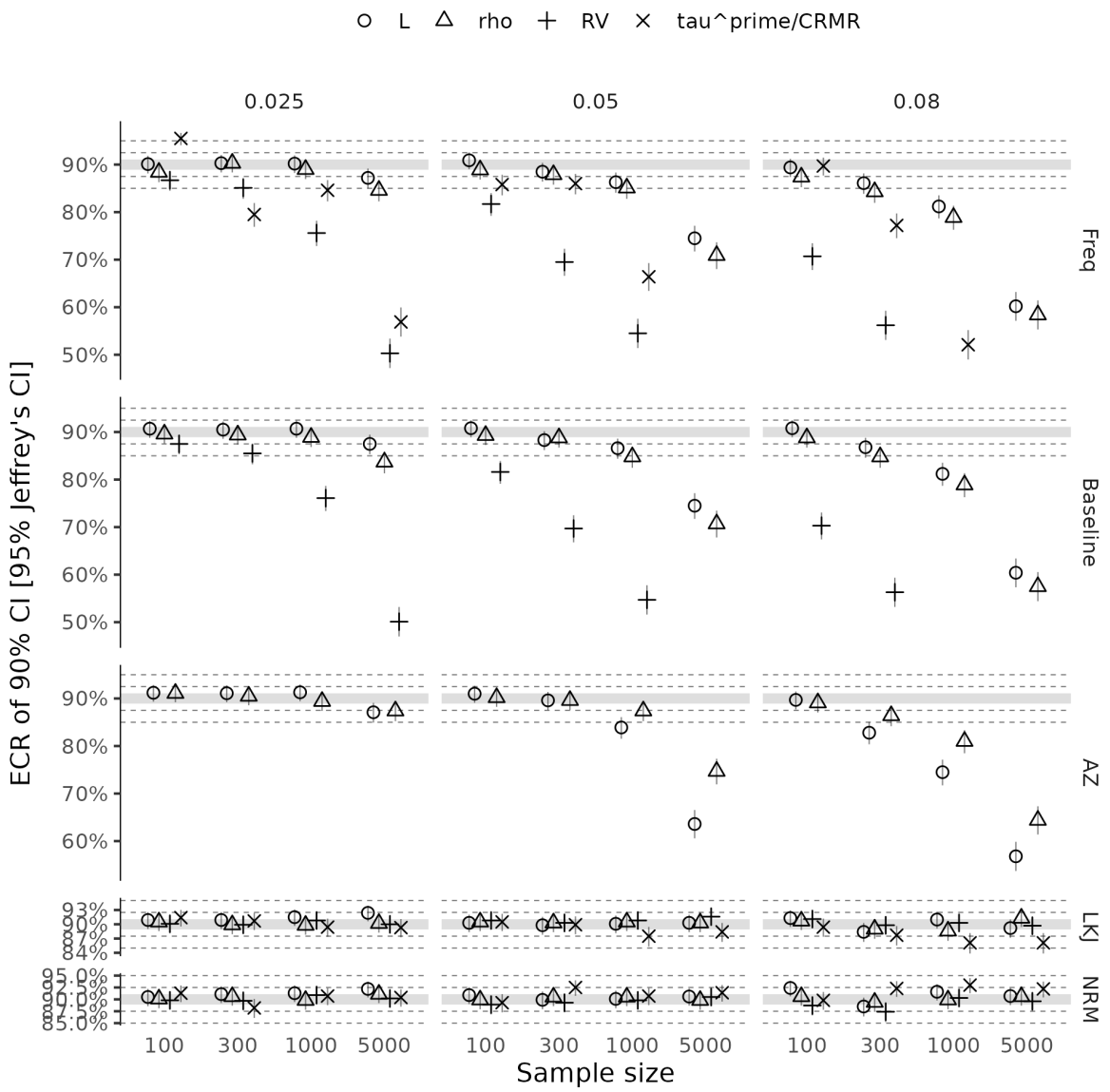
Relative bias for τ'. The most obvious problem as shown in Figure 2 was that computing the misspecification parameter on the basis of the distribution of realized values (Levy, 2011) in the baseline model resulted in highly biased estimates for the misspecification parameter. This bias was only acceptable at $n \geq 1000$ and high misspecification ($\tau' = .08$). On the other hand, both the LKJ and normal methods returned acceptable bias across conditions with one exception: the LKJ approach had abut 16% relative bias for the combination of small sample size ($n = 100$) and small misspecification ($\tau' = .025$).[4] Hence both proposed approaches appear consistent and rarely biased. The *unbiased CRMR* was downwardly biased when $\tau' = 0.025$ except when the sample size was large.

ECR for structural parameters and τ'. As shown in Figure 3, the frequentist and baseline models resulted in severe under-coverage (ECR $\ll$ 85%) relatively often. This problem increased with larger values of τ' and larger sample sizes. The AZ approach also demonstrated

[4]The downward bias of τ'_ψ (normal model) for smaller samples was practically trivial, e.g. $-7.85\%, -4.92\%, -6.43\%$ relative bias observed when $\tau = \{0.025, 0.05, 0.08\}$ and $n = 100$ means the average estimate of τ'_ψ was 0.023, 0.0475 and 0.0749 respectively.

Figure 3

Study 1: Empirical coverage rate of 90% CI of structural parameters and τ'.



Note. Ideally, all ECR estimates fall in (87.5%, 92.5%) interval, (85%, 95%) is a more liberal condition. Horizontal dashed lines represent these bounds. The y-axis is truncated at 50% – the ECR for τ' in the baseline model was less than 50% in several conditions. L is the loading; RV: residual variance; rho is the interfactor correlation.

under-coverage for the loading and interfactor correlation when $n \geq 1000$ and $\tau' \geq 0.05$. Finally, the ECR was adequate for both structural parameters and τ' for both proposed approaches.

Discussion of simulation study 1

Results suggests adequate parameter recovery and inference for the proposed models in a relatively simple CFA scenario under the assumption that minor factors influence the population covariance matrix. The only exception was notable upward bias for τ'_r at small sample size and small influence of minor factors. For this reason, of the proposed approaches, we continue with evaluation of the approach based on hierarchical estimation of the residual covariances (*normal*) approach.

As one might expect, ignoring the influence of minor factors – as the frequentist and baseline models do – leads to incorrect inference. Although structural parameter estimates remain unbiased, their uncertainty is under-estimated (see Figure C1 in Appendix C for comparison of empirical and model-based SEs). This problem is magnified for larger samples or larger influence of minor factors. We find a similar undesirable result for the AZ model which assumes the size of the influence of minor factor is known. This model also under-estimates the uncertainty about structural parameters leading to poor inference.

An additional finding is the downward bias of the unbiased CRMR under low misspecification for smaller samples – this finding was reported in the original presentation of the unbiased CRMR (Maydeu-Olivares, 2017, Table 1). Importantly, this combination of conditions returns acceptable bias for τ'_ψ returned by the normal model. This suggests that our method produces a fit index that is robust under certain scenarios that may affect the unbiased CRMR.

Finally, given the inability of extant approaches to produce adequate inference, we do not evaluate them going forward.

Simulation study 2

We set out to study the ability of the proposed approach to choose between a correct and wrong model, again under the assumption of minor factor influences. We modified the DGP from study 1 slightly:

$$
\begin{aligned}
&\mathbf{S} \sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\mathbf{\Sigma}\right),\ \mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^\mathsf{T} + \mathbf{\Omega} \\
&\mathbf{\Lambda}^\mathsf{T} = \left[\begin{smallmatrix} 0.7 & 0.8 & 0.6 & 0.9 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.75 & 0.85 & 0.6 \end{smallmatrix}\right],\ \mathbf{\Phi} = \left[\begin{smallmatrix} 1 & \\ .7 & 1 \end{smallmatrix}\right] \\
&\mathbf{\Omega} = \mathbf{DRD},\ \mathbf{D} = \text{diag-matrix}\left(\sqrt{\text{diag}(\mathbf{I}_{p\times p} - \mathbf{\Lambda\Phi\Lambda})}\right),\ \mathbf{R} \sim \text{LKJ}(\eta)
\end{aligned}
\tag{4}
$$

The interfactor correlation is now .7; the true total variance of indicators remained 1.

We varied the same factors from study 1: (i) sample size, $n \in \{100, 300, 1000, 5000\}$; (ii) the size of minor factor influences, $\tau' \in \{0.025, 0.05, 0.08\}$. Based on the relations in the separation-strategy section, $\eta = \frac{1}{2}\left[\left(\frac{\hat{\omega}}{\tau'_r \times 1^2}\right)^2 - p + 1\right]$, where $\hat{\omega}$ is the mean of the diagonal of $\mathbf{\Omega}$ i.e. $\eta \in \{170.345, 39.211, 12.575\}$ corresponds to $\tau' \in \{0.025, 0.05, 0.08\}$.

Models

We estimated two models in this study, both based on the hierarchical estimation of residual covariances or *normal* model:

1. A *correct* model: according to equation 2 (method 2) that assumed the correct factor configuration.

2. A *wrong* model: according to equation 2 (method 2) that wrongly assumed a unidimensional factor instead of two correlated factors.

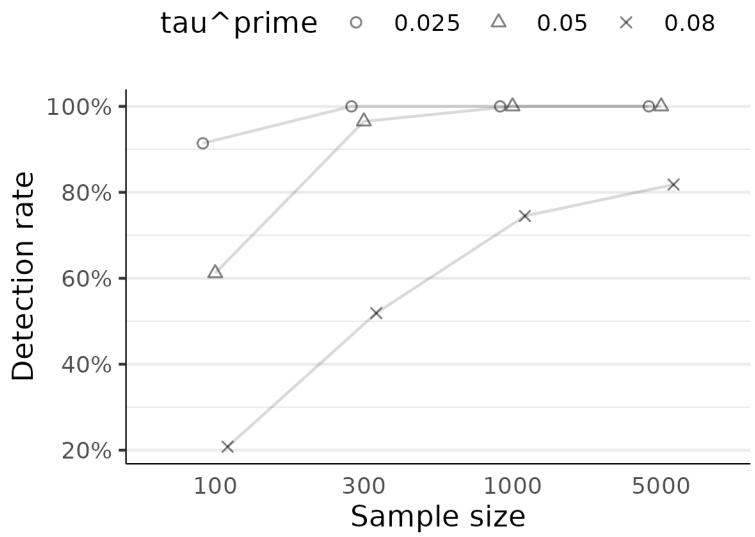
Wrongly assuming a unidimensional factor is a non-trivial misspecification. Given the DGP in equation 4, the wrong model should have expected τ' values of 0.083, 0.093 and 0.112 given the true population values of 0.025, 0.05 and 0.08. Hence, the wrong model would always be considered non-trivially misspecified. Ideally, τ' should identify the better fitting model, and should be able to choose between both models.

Estimands and performance metrics

We were interested in: (i) the recovery of the structural parameters and τ' in the correct model; and (ii) the ability of τ' to distinguish both models. For assessing parameter recovery, we maintained the same metrics and benchmarks from study 1: relative bias and ECR. For assessing

Figure 4

Study 2: Ability to distinguish correct from wrong model using τ'



the ability of τ' to distinguish both models, we subtracted τ' for the correct model from τ' for the wrong model and checked the proportion of times the 5th percentile of this difference exceeded 0.

Simulation study 2 results

Relative bias and ECR for structural parameters and τ'. Relative bias was acceptable ($\pm 10\%$) for all parameters across conditions with two exceptions: τ' was underestimated by 10.5% for both $n = 100, \tau' = 0.05$ and $n = 300, \tau' = 0.025$. Complete relative bias results are in Figure C2 in Appendix C. The ECR was acceptable ($90\% \pm 5\%$) for most parameters across conditions with a few exceptions where the ECR $> 95\%$. Complete ECR results are in Figure C3 in Appendix C

Distinguishing both models using τ'. As shown in Figure 4, τ' was able to distinguish both models to varying degrees, depending on sample size (more ability at large n) and the size of the influence of minor factors (more ability at lower τ'). When the influence of minor factors was low, τ' was almost always successful at identifying the correct model as the right model. Substantively, when the influence of minor factors is large, it can be difficult to conduct informative model comparisons. Importantly, τ' never led to the wrong conclusion, in fact τ' was never lower for the wrong model. We do not expect this result to hold when the degree of misspecification for

the major factors is less severe.

Summary of simulation study 2. The results lead to two conclusions. The normal model is able to recover structural parameters and τ' for a fairly typical CFA problem. Additionally, τ' is able to distinguish a correct model from a wrong model fit to the same data, assuming some influence of minor factors.

Simulation study 3

We were interested in the ability of the normal model to identify unduly large residual covariances when the influence of minor factors is trivial. We modified the DGP from study 2:

$$\begin{aligned}
\mathbf{S} &\sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\boldsymbol{\Sigma}\right),\ \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Omega},\ \text{where } \boldsymbol{\Omega} = \boldsymbol{\Psi} + \boldsymbol{\Delta} \\
\boldsymbol{\Lambda}^{\mathsf{T}} &= \left[\begin{smallmatrix} 0.7 & 0.8 & 0.6 & 0.9 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.75 & 0.85 & 0.6 \end{smallmatrix}\right],\ \boldsymbol{\Phi} = \left[\begin{smallmatrix} 1 & \\ .3 & 1 \end{smallmatrix}\right] \\
\text{diag}(\boldsymbol{\Omega}) &= \text{diag}(\mathbf{I}_{p\times p} - \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}),\ \frac{\delta_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} = [0.3, -0.3], \\
(\boldsymbol{\Psi} + \boldsymbol{\Delta}^*) &= \mathbf{DRD},\ \mathbf{R} \sim \text{LKJ}(\eta)
\end{aligned} \tag{5}$$

Within each iteration, we randomly picked two pairs of residual covariances in $\boldsymbol{\Delta}$, δ_{ij} in the DGP, where each member of the pair belonged to a different factor. In half the conditions, the pairs were non-null: one residual correlation set to .3 and the other set to $-.3$, hereafter *non-null delta* condition. In the other half of conditions, the pairs were null, hereafter *null delta* condition. In the non-null delta condition, the absolute values of residual covariances in $\boldsymbol{\Delta}$ range from 0.069 (when $i, j = 4, 9$) to 0.21 (when $i, j = 5, 10$). We assumed $\tau' = 0.025$ such that residual covariances due to minor factors mostly lie in the ± 0.05 interval, hence both pairs of residual covariances specified in $\boldsymbol{\Delta}$ should be larger than the elements in $\boldsymbol{\Psi}$ in the non-null delta condition.[5]

As noted in equation 2, we estimate off-diagonal residual covariances in $\boldsymbol{\Delta}$ via parameter expansion (Merkle & Rosseel, 2018; Palomo et al., 2007), such that $\boldsymbol{\Delta}^*$ is a diagonal matrix of residual variances after accounting for hypothesized residual covariances in $\boldsymbol{\Delta}$.

In addition to varying whether the residual covariances identified in $\boldsymbol{\Delta}$ were null or non-null, we varied the sample size, $n \in \{100, 300, 1000, 5000\}$.

[5]The residual covariances in $\boldsymbol{\Delta}$ are on the same scale as the residual covariances in $\boldsymbol{\Psi}$, since the population covariance matrix is standardized.

Models

We estimated two models in this study, both based on the *normal* model:

1. A *complex* model: according to equation 2 (method 2) that assumed the correct factor configuration, and pre-specified the two randomly selected residual covariances in $\mathbf{\Delta}$.
2. A *simple* model: according to equation 2 (method 2) that also assumed the correct factor configuration, but did not pre-specify the two randomly selected residual covariances in $\mathbf{\Delta}$.

In the null-delta condition, the *simple* model is correct while the *complex* model has two null parameters. In the non-null delta condition, the *simple* model is missing two parameters, while the *complex* model is correct. However, since the *simple* model estimates $\mathbf{\Psi}$, both parameters should stand out in $\mathbf{\Psi}$.

Estimands and performance metrics

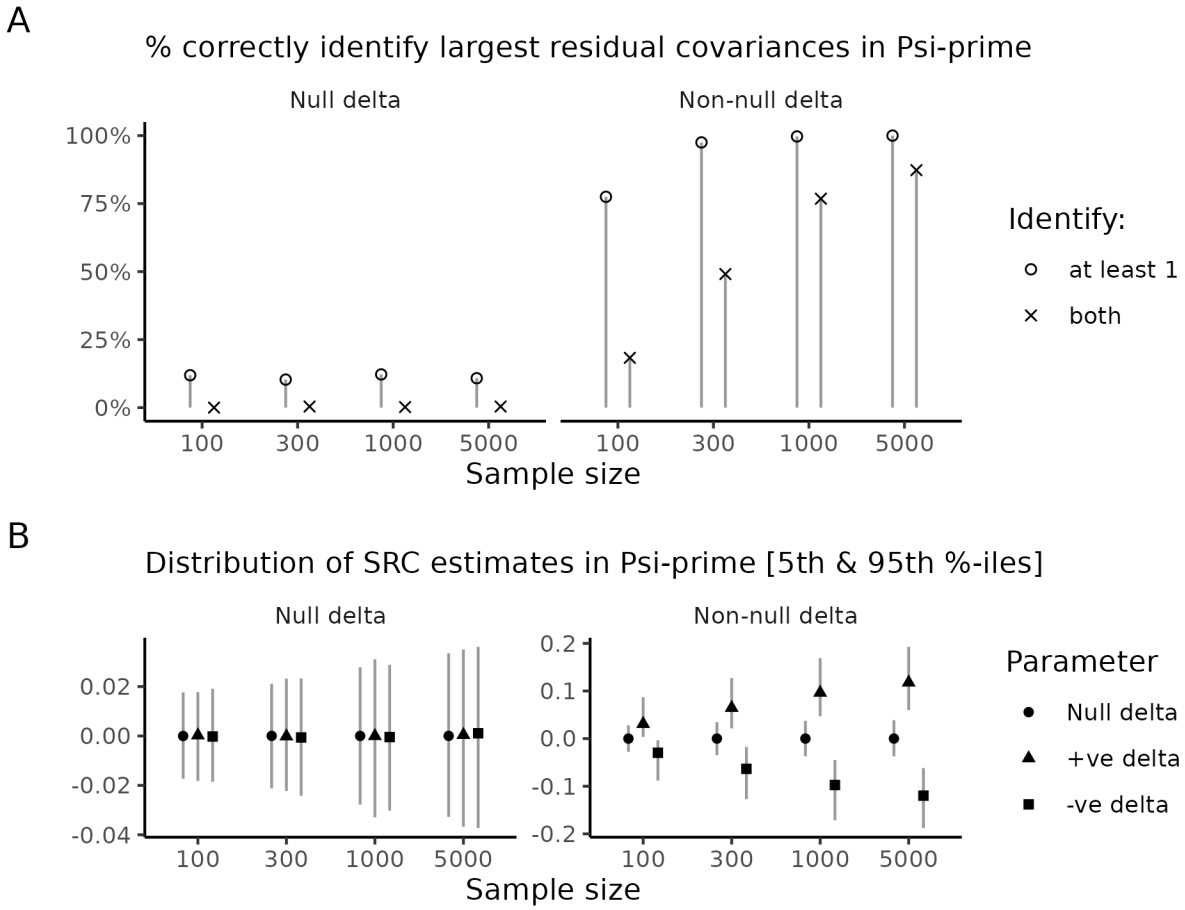
We were interested in: (i) the ability of the *simple* model to identify the pairs of randomly selected elements in $\mathbf{\Delta}$ as the largest SRCs in $\mathbf{\Psi}'$ (standardized $\mathbf{\Psi}$) in the non-null delta condition; (ii) the ability of the *complex* model to correctly estimate the randomly selected elements in $\mathbf{\Delta}$; and (iii) the ability of τ' to distinguish both models. For assessing aim (i), we examined the proportion of times the simple model identified the randomly selected pairs in $\mathbf{\Delta}$ as the largest SRCs in $\mathbf{\Psi}'$. For assessing aim (ii), we assessed the bias and ECR for both residual correlations estimated in $\mathbf{\Delta}$ in the complex model. For assessing aim (iii), we subtracted τ' for the complex model from τ' for the simple model and checked the proportion of times the 5th percentile of this difference exceeded 0.

Simulation study 3 results

***Simple* model: Identification of large SRCs.** In the null delta condition, the model identifies the pairs as the two largest SRCs in $\mathbf{\Psi}'$ at a low rate, see left panel of Figure 5A. In the non-null delta condition, the model identifies at least one of the pair at a very high rate, while the ability to identify both pairs increases with sample size, see right panel of Figure 5A. Additionally, the distribution of SRCs is always centered around 0 in the null delta condition (Figure 5B left panel), with variation that is a function of τ' and sample size. The two pairs of randomly selected

Figure 5

Study 3: Detecting large SRCs in simple model.



Note. Panel A. When using the *simple* model to identify the two largest SRCs in $\boldsymbol{\Psi}'$, do they correspond to the randomly selected pairs in $\boldsymbol{\Delta}$ in DGP? Panel B. Distribution of SRC estimates in $\boldsymbol{\Psi}'$ for non-randomly selected pairs, 1st, and 2nd randomly selected pairs in $\boldsymbol{\Delta}$ in DGP.

residual covariances result in estimates different from 0 in the non-null delta condition (Figure 5B right panel). Finally, the size of the SRC estimates increases with sample size, because the estimation of SRCs is regularized (shrunken) by their normal prior. These results suggest that $\boldsymbol{\Psi}'$ can be used to identify overly large residual covariances that were not pre-specified in $\boldsymbol{\Delta}$.

***Complex* model: Recovery of residual covariance parameters.** In the null-delta condition, the complex model estimates two redundant elements in $\boldsymbol{\Delta}$ with population parameter of 0, hence we are unable to compute their relative bias. These estimates had negligible bias, range = (-0.0049, 0.0024). In the non-null delta condition, the relative bias of both residual correlations ranged from -7.9% to 0.6%, suggesting acceptable bias for the residual correlations. Hence, the

model is capable of accurately estimating pre-specified residual correlations in $\mathbf{\Delta}$ alongside $\mathbf{\Psi}$. Additionally, the relative bias and ECR for structural parameters and τ' in the complex model were almost always acceptable and are reported in appendix C.

Distinguishing both models using τ'. In the null delta condition, the 90% interval for the difference in τ' for both models always included 0 across all conditions. Hence an investigator would rightly conclude that the *simple* model was just as acceptable as the *complex* model leading to a parsimonious solution. In the non-null delta condition, the 90% interval for the difference in τ' for both models excluded 0: 1.9%, 30.9%, 68.3% and 84.3% of the time for $n = 100, 300, 1000, 5000$ respectively – τ' was lower for the complex model. The ability to distinguish both models is markedly low for small samples.

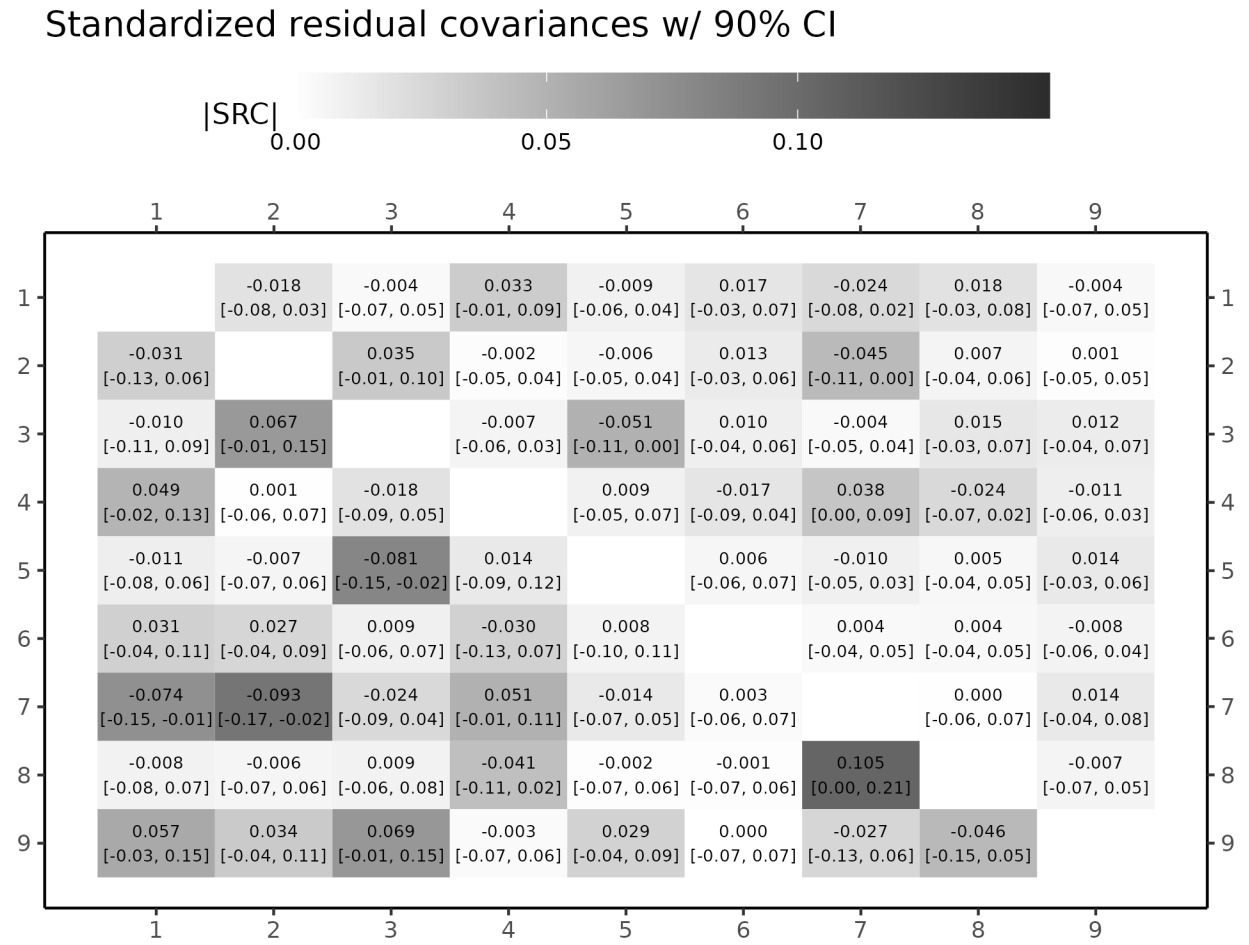
Summary of simulation study 3 results. These results demonstrate the ability to perform model diagnostics using $\mathbf{\Psi}'$ when the influence of minor factors is assumed. And as one might expect, sample size plays a major role in the ability of the approach to yield informative diagnostics. When overly large SRCs are missing from a model (as in the *simple* model), the approach is able to identify them especially when analyzing larger sample sizes. Moreover, the approach is able to estimate pre-specified residual covariances alongside the full residual covariance matrix as in the *complex* model. However, the ability to detect missing residual covariances was very low at small sample sizes.

Demonstrations

Having shown that the approach produces reliable results, we now demonstrate the approach with real world data. We fit the hierarchical estimation of residual covariances approach. As with the simulation studies, the MCMC engine was Stan. For each model, there were 2000 warmup iterations, then 2000 iterations were retained for inference across 4 chains. Sampler-specific diagnostics (Betancourt, 2017) were adequate for all estimated models. Across all models and parameters, the maximum $\widehat{R}$ never exceeded 1.01 suggesting parameter convergence for all parameters across models.

Figure 6

Holzinger-Swineford model: Standardized residual covariances in Ψ' matrix



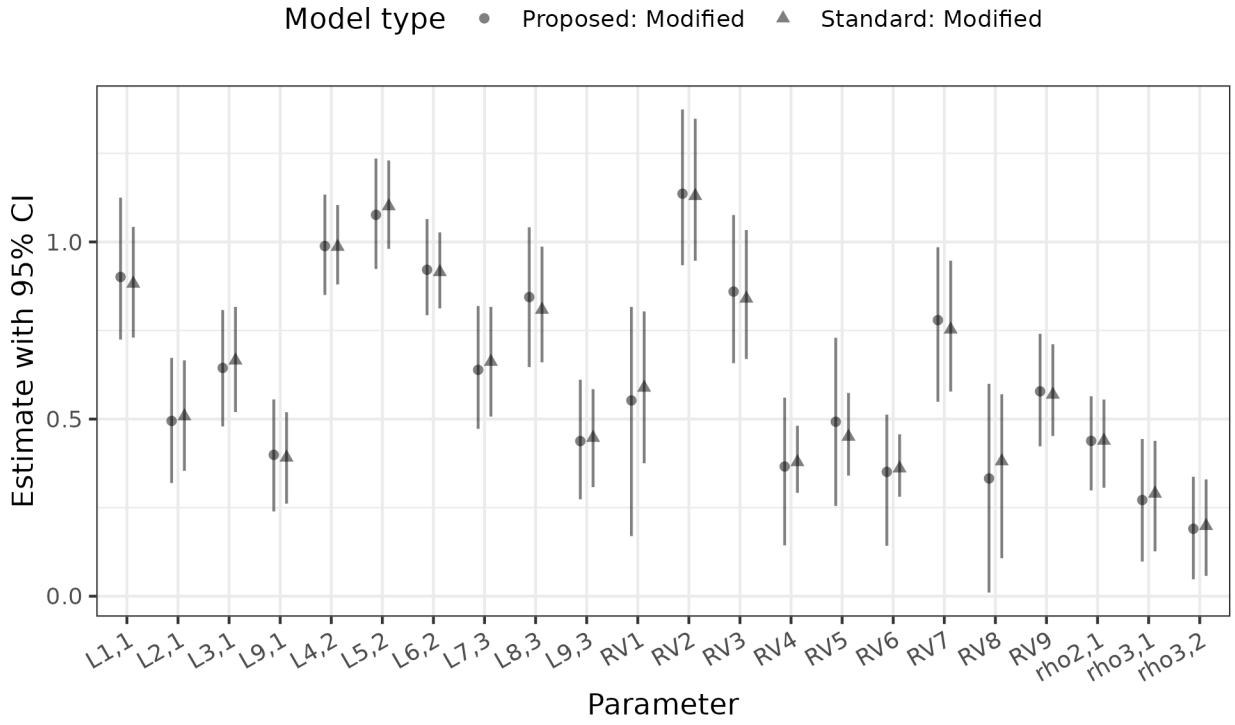
Note. Baseline model estimates in strict lower triangular part of matrix, modified model estimates in strict upper triangular part of matrix.

Holzinger-Swineford example

We demonstrate the approach with the reduced Holzinger-Swineford dataset (n = 301, Rosseel, 2012). We assumed the following factor structure as a *baseline* model: $\mathbf{\Lambda}_*^\mathsf{T} = \begin{bmatrix} \times & \times & \times & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \times & \times & \times & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \times & \times & \times \end{bmatrix}$, where $\times$ represents estimated loadings and $\cdot$ are loadings constrained to 0. We assumed all factors (namely visual, verbal, and speed factors) were correlated. τ' was 0.064, 95% CI [0.039, 0.098], such that one might expect some SRCs to exceed the ±0.1 desired interval. The unbiased CRMR (Maydeu-Olivares, 2017) for the corresponding frequentist model fit with lavaan (Rosseel, 2012) was 0.065, 95% CI [.047, .083]. The CRMR frequentist-CI is narrower because it ignores uncertainty due to minor factors. We inspected patterns in the SRCs (strict lower triangu-

Figure 7

Final fitted Holzinger-Swineford model showing parameter estimates and 95% CI across fitted models, $n = 301$.



Note. LX.Y are loadings of factor Y on test X; RV: residual variances; rho: interfactor correlations. Standard estimates are from a model with no $\mathbf{\Psi}$ matrix.

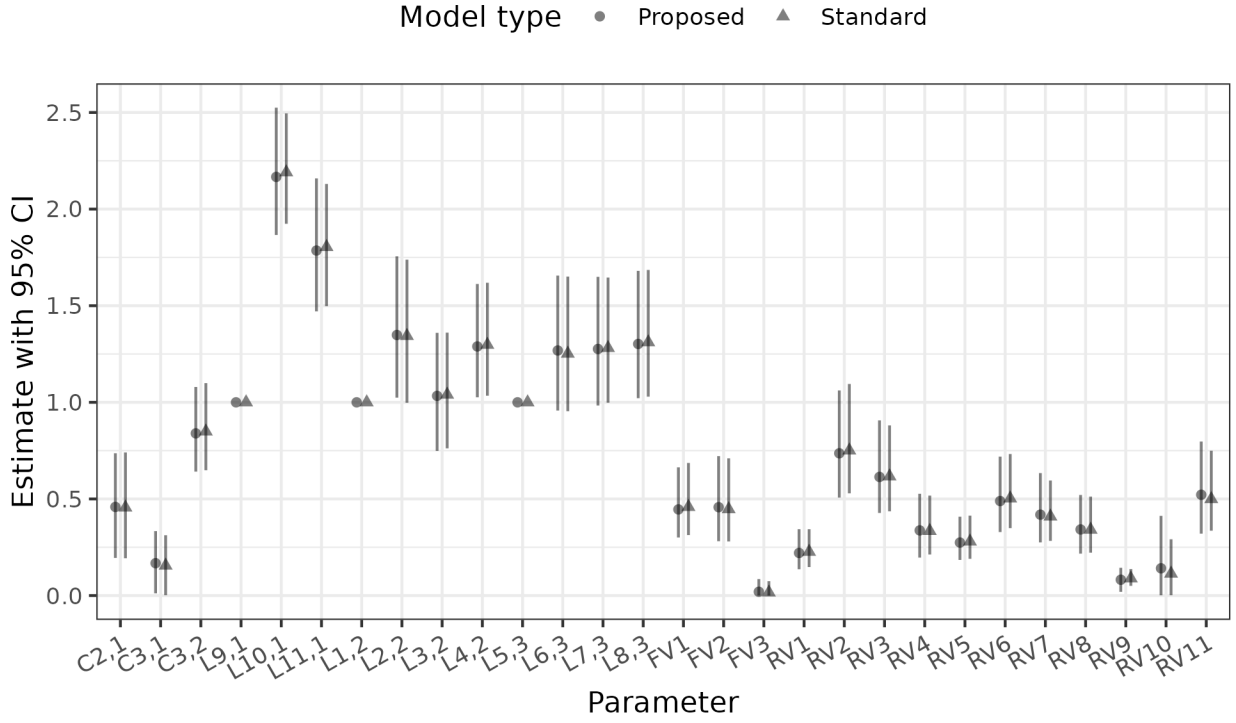
lar part of Figure 6). We preferred to modify the factor structure as opposed to including residual correlations. Based on the consistently positive residual relation between test 9 (discrimination straight and curved capitals) and tests 1–3 (visual factor), we included test 9 as an indicator of the visual factor i.e. $\mathbf{\Lambda}_*^\intercal = \begin{bmatrix} \times & \times & \times & \cdot & \cdot & \cdot & \cdot & \cdot & \times \\ \cdot & \cdot & \cdot & \times & \times & \times & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \times & \times & \times \end{bmatrix}$ – this modification makes sense on face value.[6] For the *modified* model, τ' was 0.037, 95% CI [0.014, 0.064], suggesting SRCs were mostly contained within the ±0.1 desired interval. The unbiased CRMR for the corresponding frequentist model was 0.039, 95% CI [0.023, 0.055]. Additionally, based on a comparison of posterior distributions for τ', there was 92% chance that τ' reduced from the baseline model to the modified model. We inspected patterns in the SRCs (strict upper triangular part of Figure 6) and almost all SRC estimates were within ±0.05 with 90% CIs almost fully contained in the (-0.1, 0.1) interval. Given these results, we judged the effect of minor factors to be trivial. For comparison, we also fit a model that ig-

[6]This is a subjective choice, other valid choices exist. See Saris et al. (2009) for elaboration of considerations in model modifications.

nored the effect of minor factors (i.e. no $\mathbf{\Psi}$ matrix) – a *standard* BSEM. Both the modified and standard models had highly similar point estimates for all parameters, but the standard model had narrower credible intervals (Figure 7). These results reflect the findings from simulation study 1, i.e. accounting for minor factors rightly increases uncertainty about structural parameters.

Figure 8

Political Democracy model showing parameter estimates and 95% CI across fitted models, $n = 75$.



Note. CY.X are latent regression coefficients of factor Y regressed on factor X. LX.Y are loadings of factor Y on item X; FV: factor variances; RV: residual variances. Standard estimates are from a model with no $\mathbf{\Psi}$ matrix. One marker loading per factor was set to 1.

Political Democracy example

We demonstrate the approach with the Political Democracy dataset ($n = 75$, Bollen, 1989). There were 11 items: four ratings data from 1960, four ratings data from 1965 and three economic measures from 1960. We assumed the following structured model: $\mathbf{\Lambda}_*^\intercal = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \times & \times \\ 1 & \times & \times & \times & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \times & \times & \times & \cdot & \cdot & \cdot \end{bmatrix}$, $\mathbf{B}_* = \begin{bmatrix} \cdot & \cdot & \cdot \\ \times & \cdot & \cdot \\ \times & \times & \cdot \end{bmatrix}$ for latent regression, and estimated factor (residual) variances and item residual variances – the Bayesian model is in appendix B. The ratings data were divided by 3 so that their variances were closer to 1. τ' was 0.020, 95% CI [0.001, 0.043], and suggesting that all SRCs were

within the ±0.1 desired interval. The unbiased CRMR for the corresponding frequentist model was 0.038, 95% CI [0.014, 0.061]. These frequentist values are notably larger than τ'. However, the frequentist tests of *exact* (based on the CRMR) and *close* fit (based on the unbiased CRMR) were not statistically significant, $p = .067$ & .81 respectively. Hence in a frequentist context, there would be insufficient statistical power to claim misfit on the basis of this family of fit statistics. We also inspected patterns in the SRCs (Figure D1 in Appendix D), and all SRC estimates were within ±0.05 with 90% intervals contained in (-0.1, 0.1). And we conclude this model acceptable.

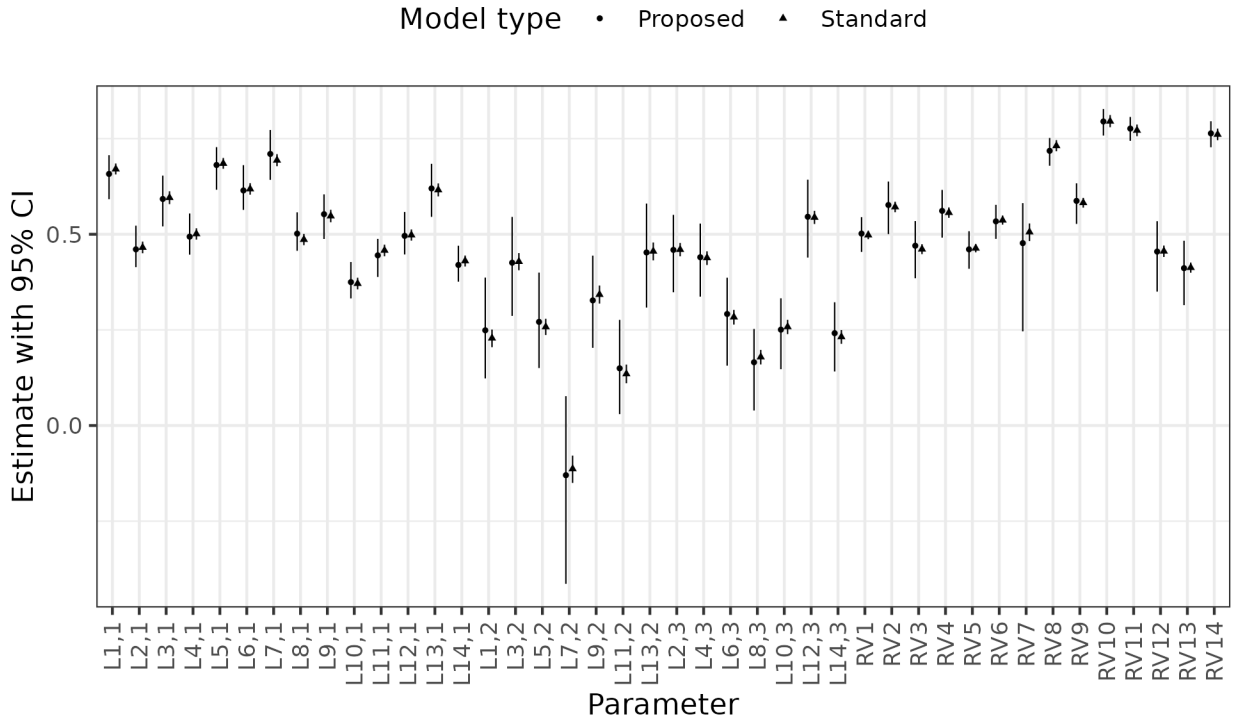
For comparison, we fit a model that ignored the effect of minor factors (i.e. no $\mathbf{\Psi}$ matrix) – a *standard* BSEM. Both models had highly similar point estimates for all parameters, and the standard model did not always have narrower credible intervals (Figure 8), especially when compared to the first demonstration. This matches the simulation study 1 finding that: when the influence of minor factors is smaller (smaller τ') and sample size is small, the increase in uncertainty about structural parameters due to accounting for minor factor is lessened.

Hospital Anxiety and Depression scale example

For our final example, we demonstrate the approach with the Hospital Anxiety and Depression scale (Zigmond & Snaith, 1983). We use the pooled correlation matrix computed by Norton, Cosco, Doyle, Done, and Sacker (2013) in the context of a meta-analytic CFA as the input data, $n = 21820$. We assumed the 14 items followed a bifactor structure: $\mathbf{\Lambda}_*^\intercal = \begin{bmatrix} \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\ \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot \\ \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times & \cdot & \times \end{bmatrix}$, representing general, anxiety and depression factors – all constrained orthogonal. τ' was 0.028, 95% CI [0.023, 0.033], and all SRC intervals were fully contained the ±0.1 desired interval (Figure D2 in Appendix D). The unbiased CRMR for the corresponding frequentist model was 0.023, 95% CI [0.023, 0.024]. Hence, we conclude this model acceptable. For comparison, we fit a *standard* BSEM. Both models had highly similar point estimates for all parameters, but the standard model had markedly narrower credible intervals (Figure 9). This matches the simulation study 1 finding that for very large samples, the increase in uncertainty about structural parameters due to accounting for minor factor is magnified.

Figure 9

Hospital Anxiety and Depression scale bifactor model showing parameter estimates and 95% CI across fitted models, $n = 21820$.



Note. LX.Y are loadings of factor Y on item X; RV: residual variances. Standard estimates are from a model with no $\mathbf{\Psi}$ matrix.

Discussion

We have presented a uniquely Bayesian approach to model fit, demonstrated the approach using real world data, and shown that the approach is valid using simulation studies. The approach leverages Bayesian computation to model misspecification as a parameter, and permits detection of non-trivial residual covariances. Practically, our approach would require fitting a new class of BSEMs beyond what is available in pre-packaged BSEM software. For this reason, we have provided code to support adoption.

A potentially unappealing feature of the proposed approach is the increased uncertainty about estimates – an increase that is more marked at large sample sizes. However, we see this feature as a positive. A typical SEM implicitly assumes that the model configuration is correct. Uncertainty associated with specifying an incorrect model configuration is not reflected in the uncertainty about structural parameters. Given that models are at best approximations to reality,

the typical SEM can be said to have overly optimistic uncertainty about structural parameters. MA (2012) capture this misspecification in the model by introducing $\mathbf{\Psi}$, which reflects the influence of minor factors, under the assumption that the parameters underlying $\mathbf{\Psi}$ are known. We relax this assumption and estimate $\mathbf{\Psi}$ such that the size of the influence of minor factors is estimated by the model. Simulation results demonstrated that our proposed approach yields more reliable inference for structural parameters than fixing the size of the influence of minor factors a-priori. In this regard, our approach is similar to the frequentist approach of Wu and Browne (2015) that models the RMSEA as a parameter. Both approaches produce structural parameters that reflect uncertainty due to model misspecification Alternatively stated: our proposed approach models the degree of incorrectness of the model. Hence, the uncertainty associated with specifying an incorrect model (which always exists in practice) is reflected in the uncertainty about structural parameters. Moreover, one application of our approach is the creation of realistic misspecified covariance structures, with varying levels of misspecification.

Another aspect of our work is that the proposed approach returns a misspecification parameter, τ', that is highly similar to the CRMR. Given that τ' had acceptable bias and reliable inference across simulation conditions, τ' may be considered as a credible fit index in Bayesian SEMs. Based on the simulation results, the variety of conditions under which τ' is reliable is wider than that of the unbiased CRMR which is less reliable for the combination of small sample size and low influence of minor factors (Maydeu-Olivares, 2017). Moreover, the uncertainty about the unbiased CRMR does not reflect the uncertainty due to model misspecification as τ' does.

A general concern with Bayesian methods is the influence priors have on the substantive results. In this paper, there are two sets of priors: (i) priors for structural parameters; and (ii) priors related to the estimation of τ' in both the normal and LKJ models. For structural parameters, we have opted for weakly-informative priors that are reasonable defaults when the observed items are scaled to have a variance of 1. Sometimes, researchers utilize more informative priors to augment relatively weak information in the data. Our expectation is that the use of informative priors will affect the proposed approach in ways that depend on the quality of the prior information (McNeish, 2016). When the informative priors increase efficiency about structural parameters, we can expect corresponding gains in the efficiency of τ' and standardized residual covariances; and vice-versa. We do not expect any gains to be obtained from using extremely diffuse priors unless the data

are very informative (Smid & Winter, 2020). For priors related to the estimation of τ' for both the normal and LKJ-models (for η), we hypothesize that alternative choices may yield improved results. Reasonable alternative priors include gamma, inverse-gamma, Student-t, and log-normal distributions and other distributions with positive-only support. The choice of prior for τ' is a question of optimizing the approach and we leave this question to future research.

Given the promise of Bayesian modeling of misspecification and τ' as a fit index in this initial presentation, we consider some potential extensions. We intend to explore applications to Bayesian SEMs with non-continuous indicators such as for models with binary, ordinal or mixed indicators. Moreover, distributions with scale matrices are also candidates for extension, e.g. a multivariate-t extension would be reasonable for handling indicators with outliers. Additionally, we hope to extend the work of Wu and Browne (2015) to a Bayesian context, after-which the approach can be formally compared to our proposed approach. Finally, when presenting our approach, we mentioned the possibility of sparsity-inducing methods (e.g. lasso, global-local approaches) applied to the estimation of residual covariances – we also intend to explore this in a future study.

One limitation of our proposed approach is that τ' does a poor job of distinguishing models at small sample sizes. One way to resolve this is to use generic information criteria (IC) for model selection as IC have already been applied to Bayesian SEMs (e.g. Cain & Zhang, 2019; Merkle & Rosseel, 2018).[7] We consider using approximate Bayesian leave-one-out cross-validation to distinguish between proposed models. The results are promising – see implementation details and simulation results in Appendix A.

References

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281–1311.

Betancourt, M. (2017, January). *A conceptual introduction to Hamiltonian Monte Carlo.* (arXiv: 1701.02434)

Bollen, K. A. (1989). *Structural equations with latent variables.* Hoboken, NJ, USA: John Wiley & Sons, Inc. (Pages: 514) doi: 10.1002/9781118619179

Bürkner, P.-C., & Vuorre, M. (2019, March). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. doi: 10.1177/2515245918823199

[7] We thank an anonymous reviewer for making this suggestion.

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling*, *26*(1), 39–50. doi: 10.1080/10705511.2018.1490648

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

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009, April). Handling sparsity via the horseshoe. In *Proceedings of the twelth international conference on artificial intelligence and statistics* (pp. 73–80). PMLR. Retrieved from `http://proceedings.mlr.press/v5/carvalho09a.html`

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

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 355–366. doi: 10.1111/1467-9868.00236

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

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008, December). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. doi: 10.1214/08-AOAS191

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

Lee, S. Y., Song, X. Y., & Tang, N. S. (2007, July). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 404–434. doi: 10.1080/10705510701301511

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. doi: 10.1111/oik.05985

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

Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling.* Boca Raton, FL: Chapman and Hall/CRC. (Pages: 466)

Lewandowski, D., Kurowicka, D., & Joe, H. (2009, October). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. doi:

10.1016/J.JMVA.2009.04.008

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502–511. doi: 10.1037/0033-2909.109.3.502

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558. doi: 10.1007/s11336-016-9552-7

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling*, *23*(5). doi: 10.1080/10705511.2016.1186549

Merkle, E. C., & Rosseel, Y. (2018, June). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*(4), 1–30. doi: 10.18637/jss.v085.i04

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

Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013, January). The hospital anxiety and depression scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, *74*(1), 74–81. doi: 10.1016/j.jpsychores.2012.10.010

Ogasawara, H. (2001, September). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika*, *66*(3), 421–436. doi: 10.1007/BF02294443

Palomo, J., Dunson, D. B., & Bollen, K. (2007). Bayesian structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 163–188). Elsevier/North-Holland. (Section: 8)

Park, T., & Casella, G. (2008, June). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. doi: 10.1198/016214508000000337

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–20. doi: 10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. doi: 10.1080/10705510903203433

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

Scheines, R., Hoijtink, H., & Boomsma, A. (1999, March). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*(1), 37–52. doi: 10.1007/BF02294318

Smid, S. C., & Winter, S. D. (2020, December). Dangers of the defaults: A tutorial on the impact of default priors when using Bayesian SEM with small samples. *Frontiers in Psychology*, *11*, 611963.

doi: 10.3389/fpsyg.2020.611963

Song, X. Y., & Lee, S. Y. (2012). *Basic and advanced Bayesian structural equation modeling*. Wiley. (Pages: 396) doi: 10.1002/9781118358887

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. doi: 10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 1–28. doi: 10.1214/20-BA1221

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

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

Appendix A

Elaboration of model selection based on LOO-CV

For studies 2 and 3, we also used approximate Bayesian leave-one-out cross-validation (LOO-CV, Vehtari, Gelman, & Gabry, 2017) to choose between models. There are two important points to note for LOO-CV for the recommended approach:

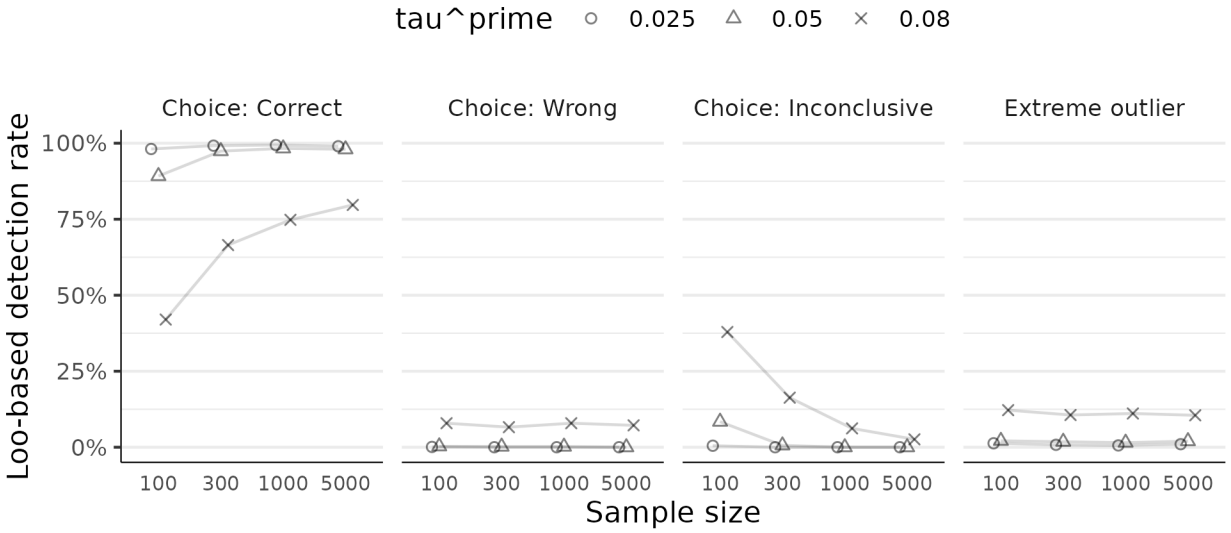
1. The presentation of the approach was based on the Wishart-likelihood applied to the sample covariance matrix. For LOO computations, the log-likelihood was the multivariate normal log-likelihood applied to the (demeaned) data.

2. The model-implied covariance matrix fit using the approach we recommend will be near-identical across different models because the models include $\mathbf{\Psi}$ which estimates all residual covariances. Hence, when computing the case-wise log-likelihood, it is better to exclude $\mathbf{\Psi}$ from the computation of the model-implied covariance or all models will have near identical fit to the data. Note that the structural parameters tend to be more variable due to accounting for minor factors and this variability will be reflected in the computed LOO information criteria (LOOIC).

To determine whether LOO-CV was able to distinguish models, within each iteration, we examined whether the difference in LOO between the two fitted models exceeded 1.96 times the

Figure A1

Study 2: LOOIC-based model selection



standard error of the difference (e.g. Bürkner & Vuorre, 2019). When this criterion was met, the model with lower LOOIC was the better fitting model. When this criterion was unmet, the model selection was judged *inconclusive*.

The LOO-CV procedure sometimes produced extremely large differences between models suggesting a problem with the procedure for such iterations. Any test that had a LOOIC difference more than four median absolute deviations (rescaled to the standard deviation scale) away from the median LOOIC-difference (within each condition) was judged an outlying test. Such tests were excluded from further analysis.

Hence, there were four mutually exclusive outcomes of LOO-CV: choosing the *correct* model, *wrong* model, an *inconclusive* decision, and *extreme outlier* cases.

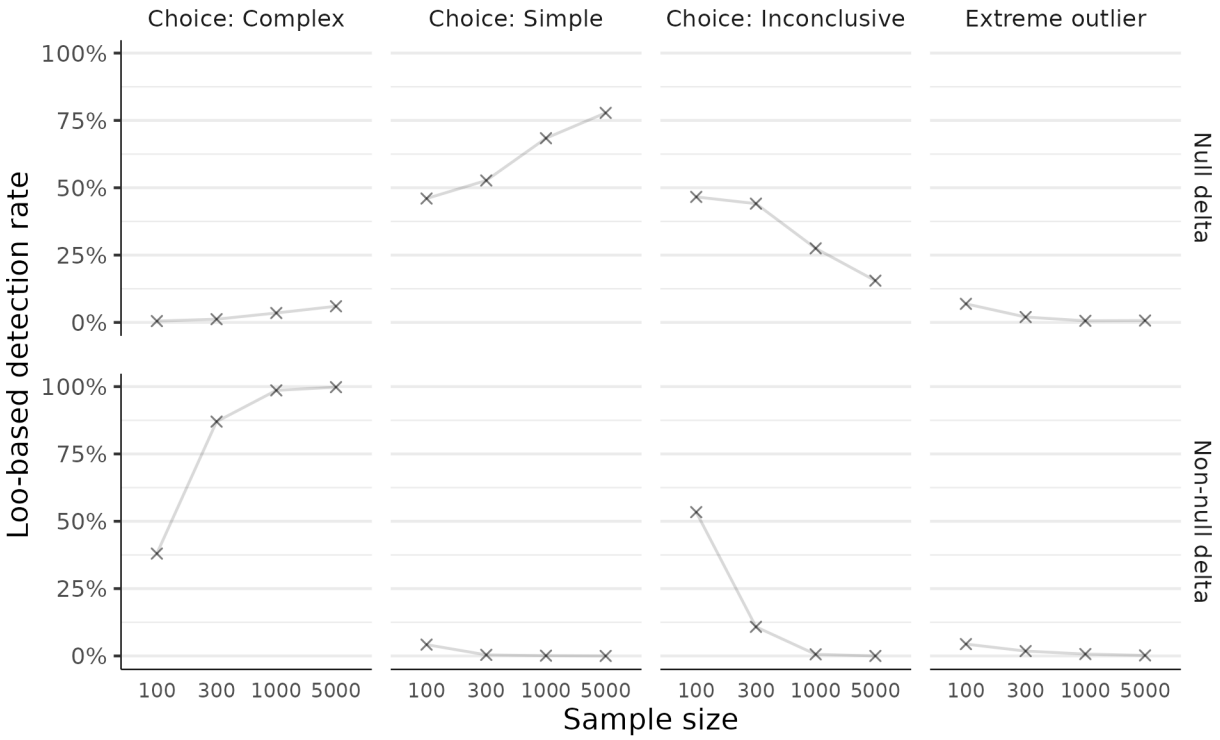
Study 2 LOO-CV results

Summarizing the simulation conditions, the data were generated under a two-factor structure. The *correct* model fit two-factors to the data, the *wrong* model assumed the data were unidimensional. Hence, the *correct* model was always the right choice.

Results are in Figure A1. The LOO-CV procedure never selected the wrong model except when τ' was 0.08 – under this condition, the procedure made the incorrect decision about 6 – 8%

Figure A2

Study 3: LOOIC-based model selection



Note. In the null delta condition (top panel), the *simple* model was the more parsimonious choice. In the non-null delta condition (bottom panel), the *complex* model was the right choice.

of the time regardless of sample size. The LOO-CV procedure also returned extreme outlying tests about 10 – 12% of the time when τ' was 0.08.

Relative to model selection based on τ', the procedure returned a much higher number of true positives. The procedure was almost always correct when τ' was negligible and often correct when τ' was large with increasing accuracy at larger sample sizes. However, there was a relatively large proportion of inconclusive results (38%) for the combination of small samples ($n = 100$) and high τ' (0.08).

In summary, as with earlier results, accurate model selection is more difficult when the influence of minor factors is relatively large. Unlike selection based on τ', LOO-CV is capable of misleading the researcher to select the incorrect model. However, LOO-CV also had a much higher rate of true-positives.

Study 3 LOO-CV results

Under the null delta condition, the *simple* was the correct model for the data, while the *complex* model estimated two residual covariances that were truly zero absent random fluctuations from zero due to $\boldsymbol{\Psi}$. Hence, under this condition, the *simple* model was the more parsimonious choice. As shown in the top panel of Figure A2, LOO-CV often selected this model or landed on an inconclusive result. However, as sample size increased, there was an increasing chance that LOO-CV picked the less parsimonious model rising to 6% when $n = 5000$.

Under the non-null delta condition, there were two residual covariances missing from the *simple*, while the *complex* model estimated both residual covariances. Hence, the *complex* model was the right choice. As shown in the bottom panel of Figure A2, LOO-CV often selected this model or landed on an inconclusive result. At $n = 100$, the proportion of inconclusive results exceeded the proportion of correct choices. Otherwise (for $n \geq 300$), the proportion of correct choices was very high (above 85%).

As with study 2, LOO-CV is much better at model selection than τ'. Moreover, outlying tests were a much smaller problem for these data. This is because the influence of minor factors was negligible in this study ($\tau' = 0.025$), and outlying tests in study 2 were more notable at large values of τ'.

Appendix B

Bayesian model for Political Democracy example

$$
\begin{aligned}
&\mathbf{S} \sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\boldsymbol{\Sigma}\right), \\
&\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{1}-\mathbf{B})^{-1}\boldsymbol{\Phi}(\mathbf{1}-\mathbf{B})^{-1\mathsf{T}}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Psi} + \boldsymbol{\Delta}, \\
&\boldsymbol{\lambda} \sim \mathcal{N}(0, \tau_\lambda),\ \tau_\lambda \sim t^+(3,0,1),\ \sqrt{\text{diag}(\boldsymbol{\Phi})} \sim t^+(3,0,1), \\
&\mathbf{B} = \begin{bmatrix} . & . & . \\ \beta_{21}\tau_{\beta 1} & . & . \\ \beta_{31}\tau_{\beta 2} & \beta_{32}\tau_{\beta 2} & . \end{bmatrix},\ [\beta_{21}, \beta_{31}, \beta_{32}] \sim \mathcal{N}(0,1),\ [\tau_{\beta 1}, \tau_{\beta 2}] \sim t^+(3,0,1), \\
&\sqrt{\text{diag}(\boldsymbol{\Delta})} \sim t^+(3,0,1),\ \text{diag}(\boldsymbol{\Psi}) = \mathbf{0}_p,\ \frac{\psi_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \sim \mathcal{N}(0, \tau'),\ \tau' \sim \mathcal{N}^+(0,1)
\end{aligned}
\tag{B1}
$$

where the coefficients have a normal prior with scale parameter ($\tau_{\beta.}$) that varies by outcome (row of $\mathbf{B}$) – akin to ridge regression, and the interfactor covariance matrix ($\boldsymbol{\Phi}$) is diagonal and factor

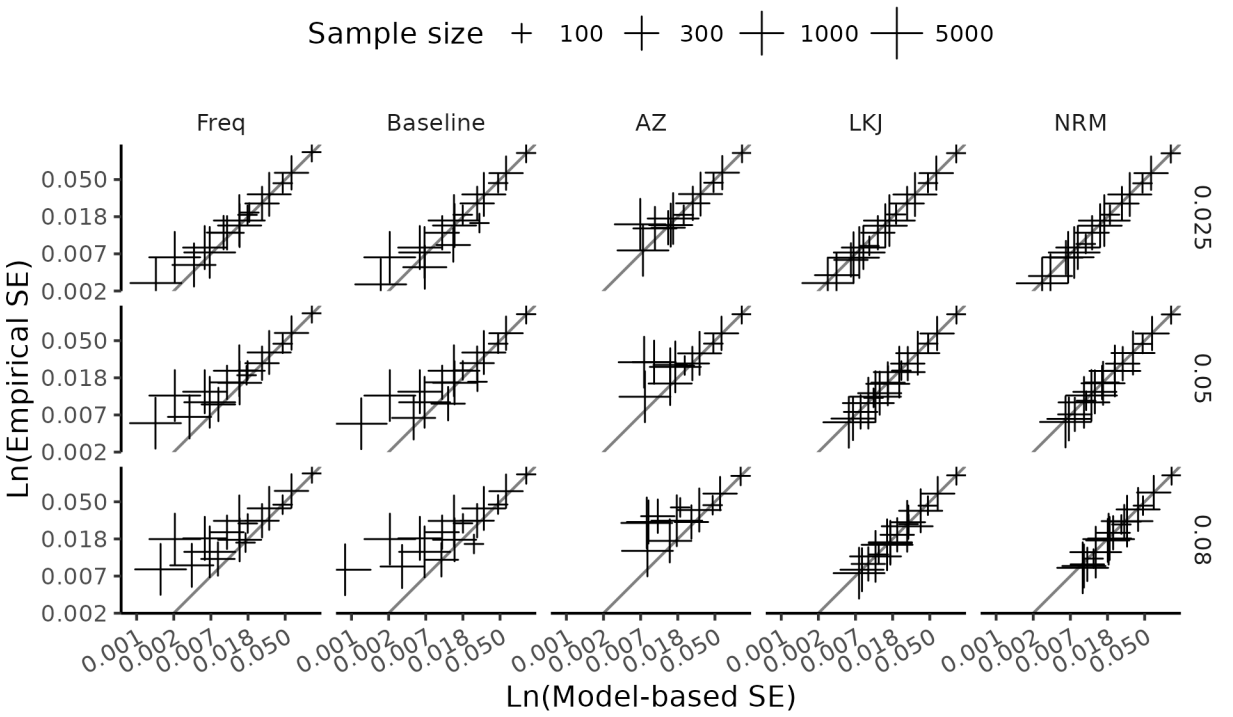
standard deviations are assumed half-t.

Appendix C

Additional results for simulation studies

Figure C1

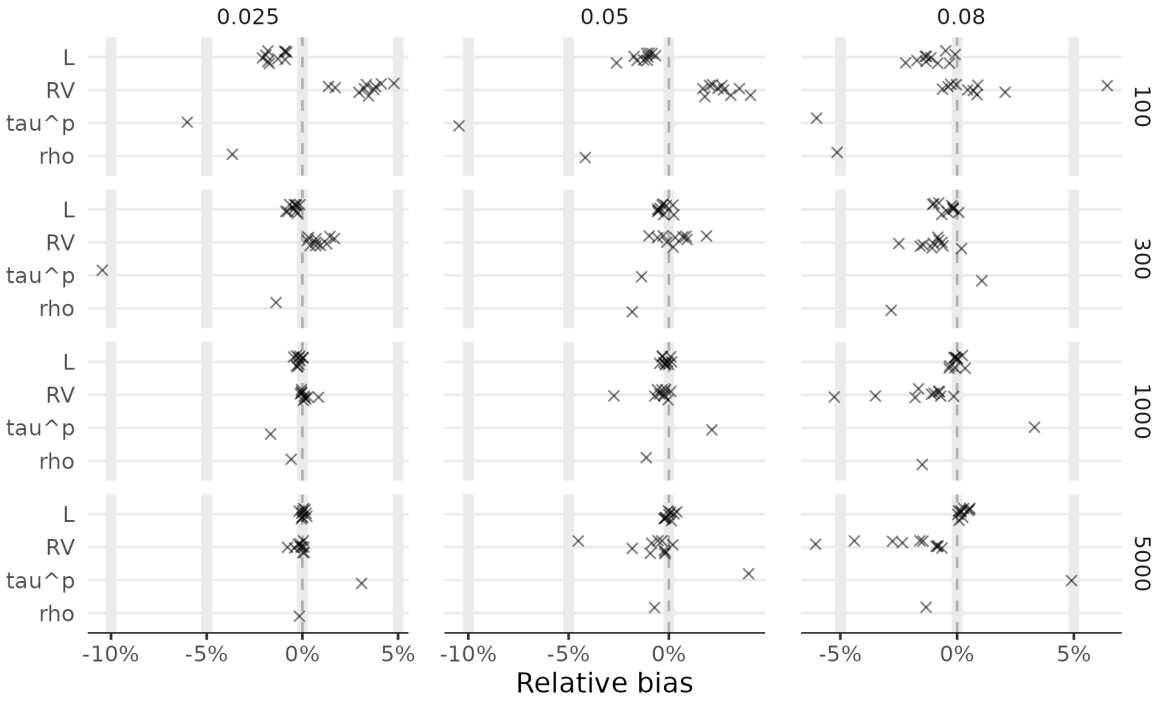
Study 1: Empirical versus model-based standard errors for different parameters



Note. Both axes are log-transformed so all points are clearly visible. Ideally, all points should be on the line (intercept = 0, slope = 1). But the model-reported standard error is often less than the standard deviation of estimates for the frequentist, baseline and AZ models.

Figure C2

Study 2: Relative bias of structural parameters and τ' under correct model.



Note. Each point is an estimated relative bias. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal condition. Vertical bars represent these bounds. L are loadings; RV: residual variances; tau^p is τ'; rho is the interfactor correlation. Points are slightly jittered vertically.

Figure C3

Study 2: Empirical coverage rate of 90% CI of structural parameters and τ' under correct model.

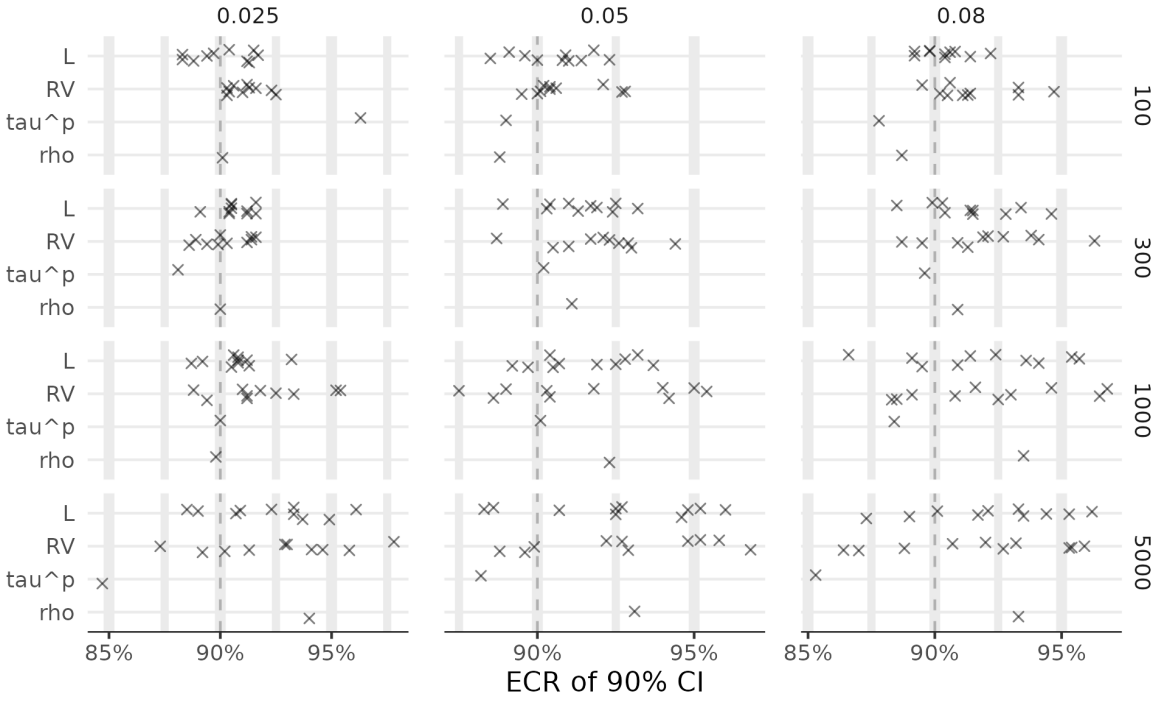
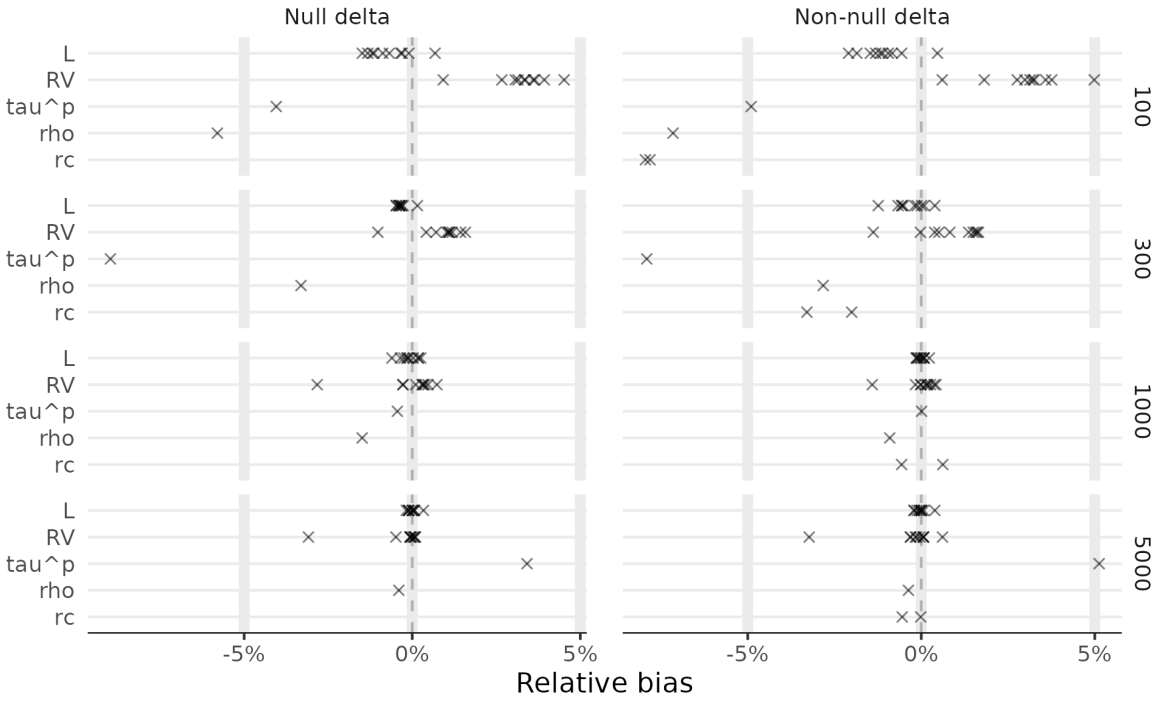


Figure C4

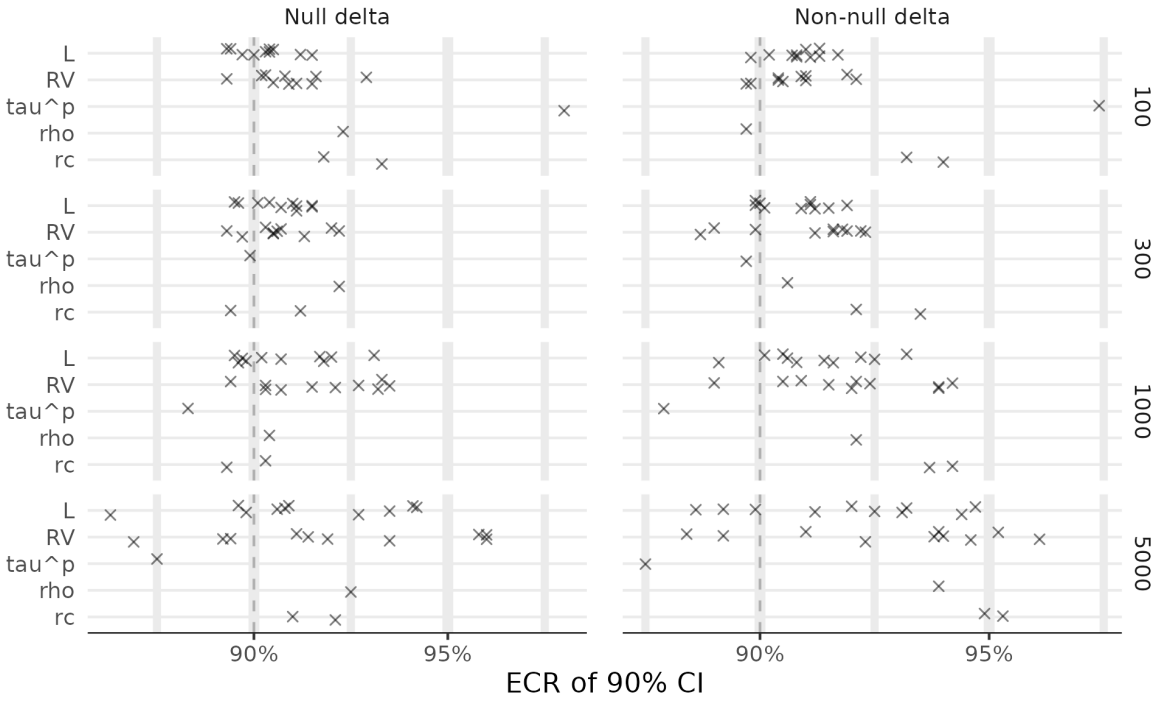
Study 3: Relative bias of structural parameters and τ' under complex *model.*



Note. Each point is an estimated relative bias. Ideally, all absolute RB estimates are under 5%, 10% is a more liberal condition. Vertical bars represent these bounds. L are loadings; RV: residual variances; tauˆp is τ'; rho is the interfactor correlation; rc are residual correlations. Points are slightly jittered vertically.

Figure C5

Study 3: Empirical coverage rate of 90% CI of structural parameters and τ' under complex *model.*



Note. Each point is an estimated ECR. Ideally, all ECR estimates fall in (87.5%, 92.5%) interval, (85%, 95%) is a more liberal condition. Vertical bars represent these bounds. L are loadings; RV: residual variances; tau^p is τ'; rho is the interfactor correlation; rc are residual correlations. Points are slightly jittered vertically.

Appendix D

Additional results for data demonstration

Figure D1

Political Democracy model: Standardized residual covariances in $\boldsymbol{\Psi}'$ *matrix*

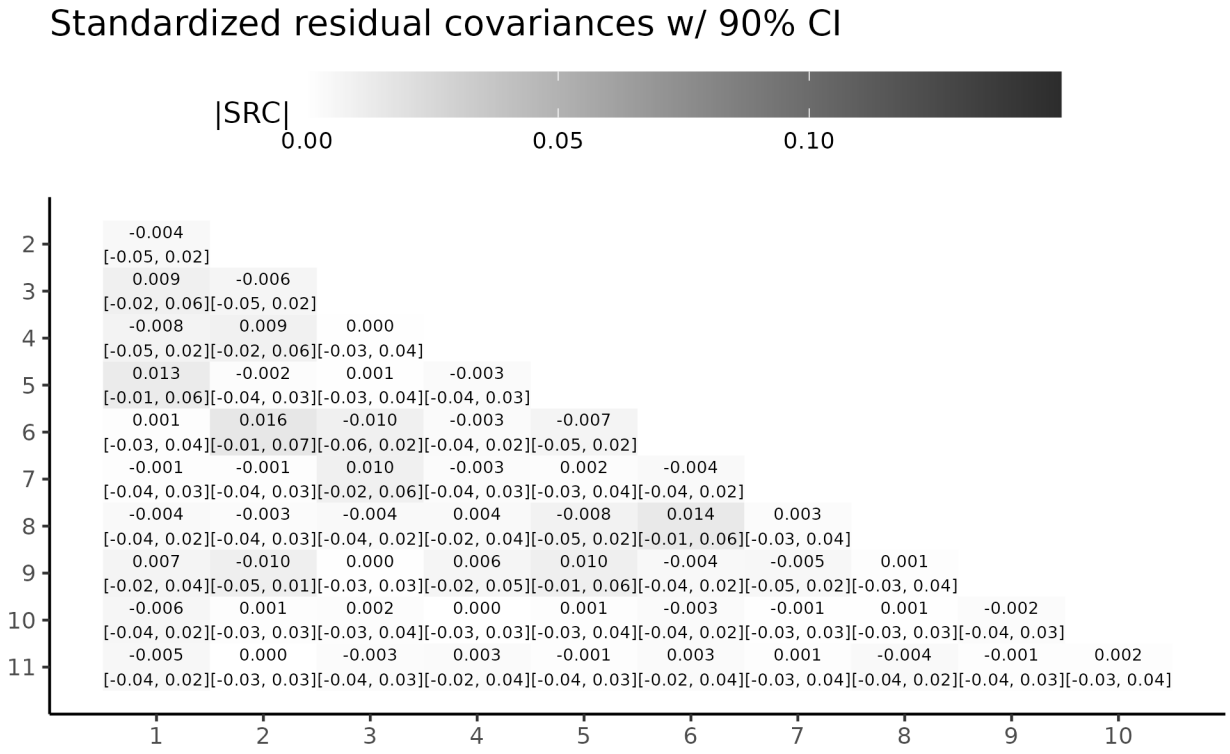


Figure D2

Hospital Anxiety and Depression scale example: Standardized residual covariances in $\boldsymbol{\Psi}'$ *matrix*