A comparison of different prior choices for estimating the influence of minor factors in Bayesian structural equation models

Abstract

Bayesian structural equation models may be estimated under the assumption that minor factors influence the population covariance matrix responsible for available data. This implies that a correct hypothesized model will have less than perfect fit to data. The influence of minor factors may be captured by regularized estimation of a full matrix of residual covariances. Extant literature has explored a normal prior for residual covariances, which returns an index of model fit. In this paper, we compare other shrinkage priors for estimating the influence of minor factors. We find that the generalized double-Pareto, a global-local prior, has optimal performance for estimating the influence of minor factors, but performs poorly for estimating the model fit index.

Keywords: model misspecification, Bayesian SEM, minor factors, CRMR, regularization, ridge, lasso, global-local

A comparison of different prior choices for estimating the influence of minor factors in Bayesian structural equation models

Model misspecification in Bayesian structural equation modeling (BSEM) is an active area of research. Levy (2011) introduced posterior predictive p-values computed using a likelihood ratio discrepancy measure to assess the presence of model misspecification. As with frequentist SEMs, PPP values can detect trivial misspecification especially for large samples, necessitating assessments of the degree or size of model misspecification. Accordingly, Levy (2011) also 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. Hoofs, van de Schoot, Jansen, and Kant (2018) evaluated the deviance information criteria for model comparisons, and proposed a Bayesian root mean square error of approximation (RMSEA). Garnier-Villarreal and Jorgensen (2020) adapted extant frequentist fit indices (RMSEA, CFI, TLI, ...) to the Bayesian context. These approaches are all similar to frequentist SEMs. Model parameters are estimated under the assumption that the model-implied covariance matrix is the population covariance matrix, and the size of misspecification is determined afterwards.

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 their approach, model parameters were estimated under the assumption that the model-implied covariance matrix is *not* the population covariance matrix. Consequently, uncertainty about model parameters reflects the degree of model misspecification. This approach is similar to the frequentist approach of Wu and Browne (2015) which estimated the RMSEA simultaneously with model parameters.

The conceptual rationale for assuming the model-implied covariance matrix is not the population covariance matrix underlying the available data is that *minor factors* which cannot be predicted a-priori influence the population covariance matrix (MacCallum &

Tucker, 1991). Uanhoro (2023b) operationalized minor factor influences by assuming the effect of minor factors (or residual covariances) are normally distributed with mean zero and a scale parameter that is learned from the data. The scale parameter when standardized is the root mean square error of standardized residual covariances, and can serve as an index of model fit.

The approach of Uanhoro (2023b) may be viewed as a Bayesian SEM that permits estimation of all residual covariances via ridge regularization. Under this framing, other regularization methods are immediately obvious, e.g. a lasso penalty (e.g. Chen, 2021). From a Bayesian perspective, different regularization methods relate to different prior choices for estimating residual covariances.

In this paper, we compare the effect of different prior choices for minor factor influences (or residual covariances) on model estimation. In the next section, we present different priors as well as their expected substantive implications for model estimation. Afterwards, we compare these priors using a simulation study.

A review of prior options

We briefly review the method in Uanhoro (2023b) using the example of confirmatory factor analysis. Assume the population covariance matrix ($\boldsymbol{\Sigma}$) underlying p indicators is:

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

where $\boldsymbol{\Lambda}$ is the loading matrix, $\boldsymbol{\Phi}$ is the inter-factor correlation matrix, $\boldsymbol{\Delta}$ is the usual residual covariance matrix (often diagonal), and $\boldsymbol{\Psi}$ is a full residual covariance matrix reflecting minor factor influences. Let ψ_{ij} $(i < j \leq p)$ be elements of the strict lower triangular part of $\boldsymbol{\Psi}$, then $\frac{\psi_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}} \sim \mathcal{N}(0, \tau)$, where τ is the root mean square error (RMSE) of standardized residual covariances (SRCs) that is learned from the data, and is akin to the CRMR.

This is a ridge-type prior, and implicitly assumes that minor factor influences are on

average zero, and deviate from 0 in a continuous manner. If one assumes that each residual covariance reflects a very large number of minor factors, each influencing the residual covariance in a tiny way, then the normality assumption follows from central limit theorem.

Logistic prior. A similar prior to the normal prior is the logistic prior, i.e. $\frac{\psi_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}} \sim \text{Logistic}(0, \tau')$, where $\tau = \tau' \frac{\pi}{\sqrt{3}}$ is the RMSE of SRCs. The logistic prior is largely similar to the normal prior, but the logistic distribution is more heavy-tailed and consequently more accommodating of outliers. Practically, the logistic prior maintains the realistic assumptions of the normal prior without overly shrinking large SRCs.

Lasso prior. Alternatively, a lasso prior could be assumed for the SRCs (e.g. Chen, 2021), i.e. $\frac{\psi_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}} \sim \text{Laplace}(0, \tau')$, where $\tau = \tau'\sqrt{2}$ is the RMSE of SRCs. A lasso prior follows from the assumption that SRCs are largely zero with a few large non-zero SRCs. This assumption is unrealistic because it is unlikely that any minor factor influences are exactly 0 for real data. However, this assumption may be practically useful when one believes a model is largely correct with a few notable residual covariances. Hence, the lasso prior is useful for identifying large SRCs.

Generalized double-Pareto prior. A feature of previously examined priors is that shrinkage is controlled by a global parameter – transformations of τ – leaving all SRCs subject to shrinkage towards zero. Ideally, we prefer a prior that strongly shrinks small SRCs, and has almost no impact on large SRCs. This would be ideal for identifying large SRCs. Practically, identification of large SRCs is important as they challenge the assumption that model misspecification is entirely due to "minor factors". A solution is to employ so-called *global-local* priors, that additionally estimate a unique (local) shrinkage parameter for each SRC. However, estimating a shrinkage parameter per SRC would increase model complexity, especially for models with a large number of indicators. The generalized double-Pareto prior (GDP; Armagan, Dunson, & Lee, 2013) is a global-local prior that can be parameterized using a single shrinkage parameter. The GDP distribution has Laplace-like spike at 0 (i.e. shrinkage to 0 for small SRCs) and Student-t tail behaviour

(i.e. heavy-tailed and accommodating of outliers). Hence, we may also assume SRCs have a GDP prior (mean-shape-scale notation): $\frac{\psi_{ij}}{\sqrt{\sigma_{jj}\sigma_{ii}}} \sim \text{GDP}(0, \alpha, \tau')$, where $\tau = \tau' \frac{\alpha\sqrt{2}}{(\alpha-1)(\alpha-2)}$ is the RMSE of SRCs ($\alpha > 2$).

In summary, the different priors have different implications, both normal and logistic priors assume continuous deviations of SRCs from zero, with the logistic prior being more accommodating of outlying SRCs. The lasso and GDP priors assume most SRCs are zero with limited exceptions, thus allowing for the identification of large SRCs, with the GDP prior allowing larger SRCs to escape shrinkage. The differences between the priors are captured in Figure 1. Based on the figure, the GDP prior can have comparable shrinkage to the lasso prior for small SRCs, while barely shrinking large SRCs.

Simulation study

We compared the different priors on their ability to identify unduly large residual covariances when the influence of minor factors is trivial. This study uses the design of simulation study 3 in Uanhoro (2023b) while comparing different prior choices. The data generation process (DGP) was:

$$
\begin{aligned}
&\mathbf{S} \sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\mathbf{\Sigma}\right),\ \mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^\mathsf{T} + \mathbf{\Omega},\ \text{where } \mathbf{\Omega} = \mathbf{\Psi} + \mathbf{\Delta} \\
&\mathbf{\Lambda}^\mathsf{T} = \begin{bmatrix} 0.7 & 0.8 & 0.6 & 0.9 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.75 & 0.85 & 0.6 \end{bmatrix},\ \mathbf{\Phi} = \begin{bmatrix} 1 & \\ .3 & 1 \end{bmatrix} \\
&\text{diag}(\mathbf{\Omega}) = \text{diag}(\mathbf{I}_{p\times p} - \mathbf{\Lambda\Phi\Lambda}),\ \frac{\delta_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} = [0.3, -0.3], \\
&(\mathbf{\Psi} + \mathbf{\Delta}^*) = \mathbf{DRD},\ \mathbf{R} \sim \text{LKJ}(\eta) \\
&n \in \{100, 300, 1000, 5000\}
\end{aligned}
\tag{2}
$$

Within each iteration, we randomly picked two pairs of residual covariances in $\mathbf{\Delta}$, δ_{ij} in the DGP, where each member of the pair belonged to a different factor. One residual correlation was set to .3, the other was set to $-.3$. The absolute values of residual covariances in $\mathbf{\Delta}$ range from 0.069 (when $i, j = 4, 9$) to 0.21 (when $i, j = 5, 10$). We chose a

value of η corresponding to $\tau = 0.025$ such that residual covariances due to minor factors mostly lie in the ± 0.05 interval, hence both pairs of residual covariances specified in $\mathbf{\Delta}$ should be larger than the elements in $\mathbf{\Psi}$.[1]

We simulated 500 datasets per condition. 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/xswy6/?view_only=4435656baf7740729054215f5bda64ba`.

Models

We estimated two model types:

1. A *complex* model: that assumed the correct factor configuration, and pre-specified the two randomly selected residual covariances in $\mathbf{\Delta}$.

2. A *simple* model: that also assumed the correct factor configuration, but did not pre-specify the two randomly selected residual covariances in $\mathbf{\Delta}$.

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 $\mathbf{\Psi}$. For both model types, we estimate four models by varying the residual covariance priors: normal, logistic, lasso, GDP.

Research questions

How do the four priors compare on:

1. the size of all SRCs excluding both randomly selected pairs in the *simple* model, hereafter *null* SRCs. Since null SRCs are largely small (± 0.05), the best model would have the smallest null SRCs on average.

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

2. the size of both randomly selected pairs of SRCs in the *simple* model, hereafter *non-null* SRCs. Since these SRCs represent missing residual covariances in the DGP (δ_{ij}), the best model would have the largest values for both SRCs.

3. the ability of the *simple* model to correctly identify the non-null SRCs as the largest SRCs. The best model would have the highest classification rate.

4. the ability of the *simple* and *complex* models to correctly estimate the structural parameters, and the ability of the *complex* model to estimate both randomly selected residual covariances, δ_{ij}. This is assessed using relative bias, coverage of 90% credible intervals and relative estimation efficiency. The best model would have acceptable bias, coverage and the highest estimation efficiency.

Simulation results

RQ 1. As shown in Figure 2 (left panel), the null SRCs were centered around zero for all priors. Compared to the normal prior, the mean absolute values of the null SRCs were 0.056 ($t_{44} = 38, p < .001$), 0.10 ($t_{44} = 27, p < .001$) and 0.16 ($t_{44} = 28, p < .001$) standard deviations lower for logistic, lasso and GDP priors respectively.[2] This matches the pattern of dispersion in Figure 2.

RQ 2. As shown in Figure 2 (right panel), the non-null SRCs were larger when the sample size was larger. Compared to the normal prior, the mean value of the non-null SRCs (with positive and negative SRCs sign-aligned) were 0.13 ($t_{24} = 8.4, p < .001$), 0.22 ($t_{24} = 9.8, p < .001$) and 0.32 ($t_{24} = 9.5, p < .001$) standard deviations higher for logistic, lasso and GDP priors respectively. This matches the pattern in Figure 2.

RQ 3. There was no statistically significant difference in the ability of the different priors to correctly identify the two largest SRCs, $\chi^2_3 = 0.015, p > .99$;[3] matching the

[2] Estimates were adjusted for sample size, cluster-robust standard errors were applied with clustering by residual covariance pair. Standard deviation was the raw unadjusted sample SD, yielding a conservative effect size. The same approach was retained for analysis of non-null SRCs.

[3] The analysis was based on a beta-binomial model of the number of times a model correctly identified both non-null SRCs, with adjustment for sample size which was a statistically significant factor,

pattern in Figure 3.

RQ 4. As shown in Figure 4, parameter recovery was largely adequate for structural parameters based on relative bias and coverage across different priors and for both the simple and complex models. The simple model sometimes had higher than nominal coverage for structural parameters, i.e. inference was sometimes conservative when there were missing residual covariances. Estimation of τ under the GDP prior was often upwardly biased and often very inefficient. Finally, estimation of τ under the normal prior was inefficient for the simple model.

Summary of simulation results

The GDP prior shrunk small minor factor influences the most, and shrunk large outlying SRCs the least, making it the most desirable prior for identification of overly large SRCs. However, the GDP prior often poorly estimated τ, making the GDP-based approach the worst option for estimating the model fit index. In this regard, the logistic approach was always adequate for estimating τ. Finally, estimation of structural parameters was often adequate across approaches, with different methods comparable in terms of bias, coverage and efficiency.

Conclusion

In this paper, we compared different prior choices for capturing the influence of minor factors on the population covariance matrix underlying available data. The GDP prior is a good choice for estimation of structural parameters as well as for estimation of the minor factor influences. Large SRCs are very weakly regularized while small SRCs are more strongly regularized. This can lead to greater distinction between trivial and non-trivial residual covariances, potentially permitting better diagnosis of model misspecification. However for determining the size of model misspecification, we recommend the use of the logistic model.

$\chi^2_3 = 23, p < .001$.

References

Armagan, A., Dunson, D. B., & Lee, J. (2013, January). Generalized double Pareto shrinkage. *Statistica Sinica*, *23*(1), 119–143.

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, J. (2021, July). A Bayesian Regularized Approach to Exploratory Factor Analysis in One Step. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 518–528. doi: 10.1080/10705511.2020.1854763

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

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

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

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

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

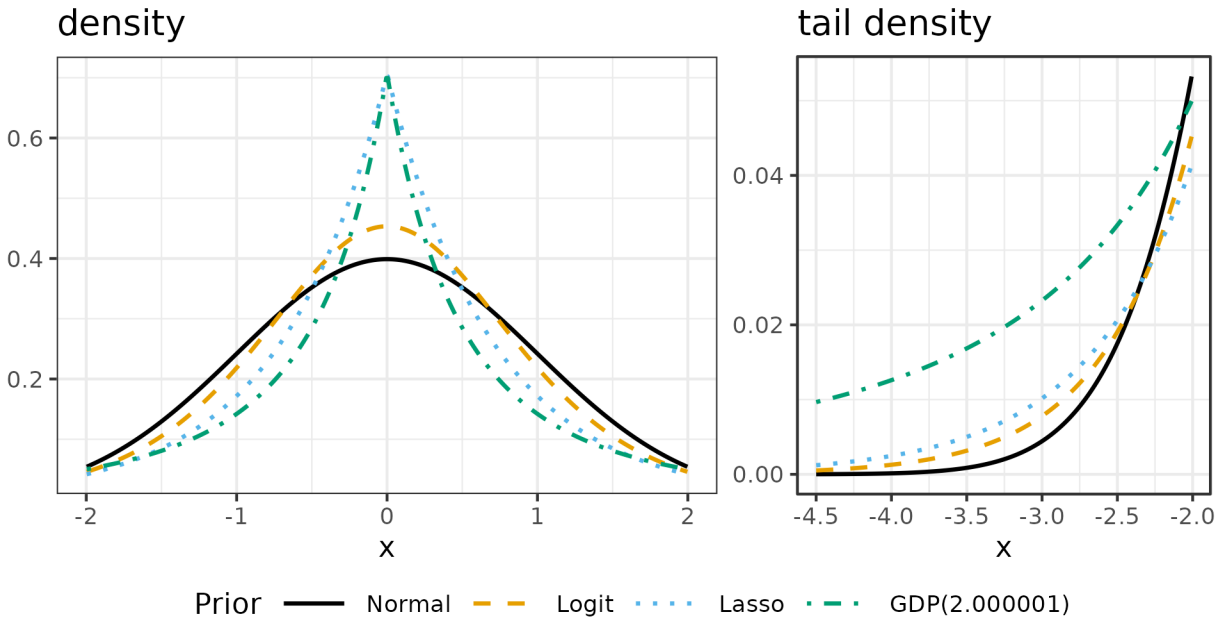
Uanhoro, J. O. (2023b, April). Modeling misspecification as a parameter in bayesian structural equation models. *Educational and Psychological Measurement*,

00131644231165306. (Publisher: SAGE Publications Inc) doi: 10.1177/00131644231165306

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

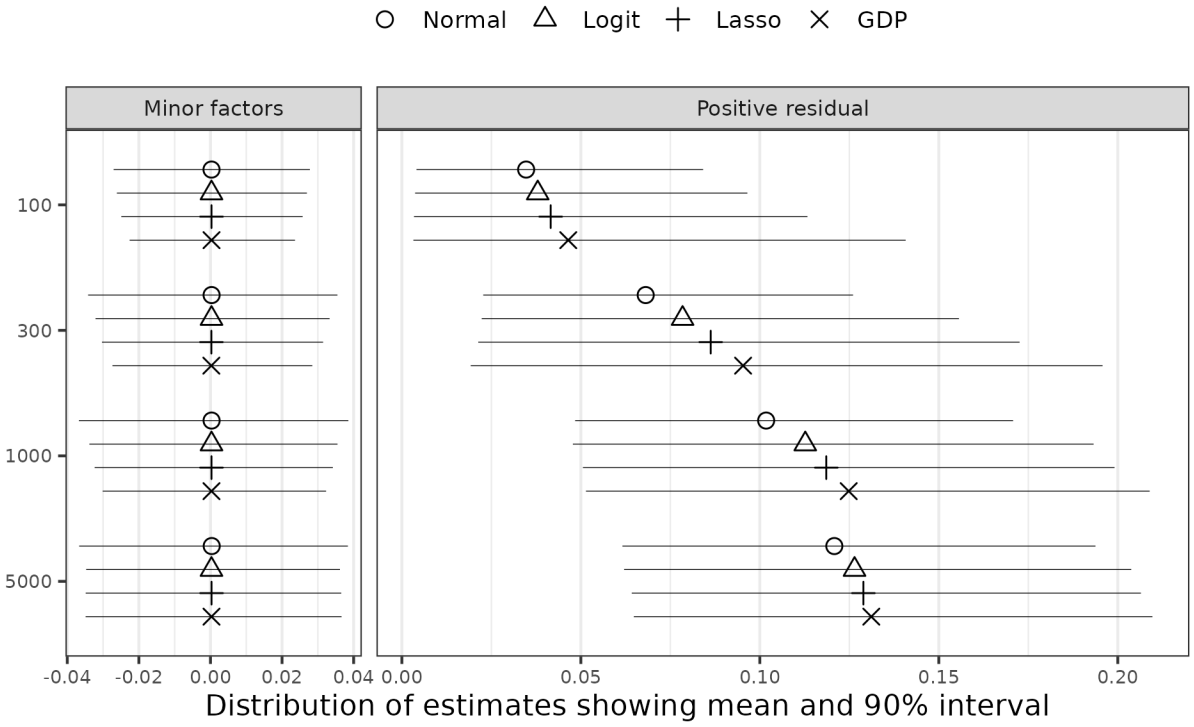
Density of different priors with mean 0 and standard deviation 1.



Note. The GDP shape parameter is in parenthesis. The normal and logistic priors are bell-shaped, while the lasso and GDP priors are spiked at 0. Larger peaks suggest increased shrinkage of small SRCs, with spikes suggesting shrinkage to 0. Larger tail density suggests less shrinkage for large SRCs.

Figure 2

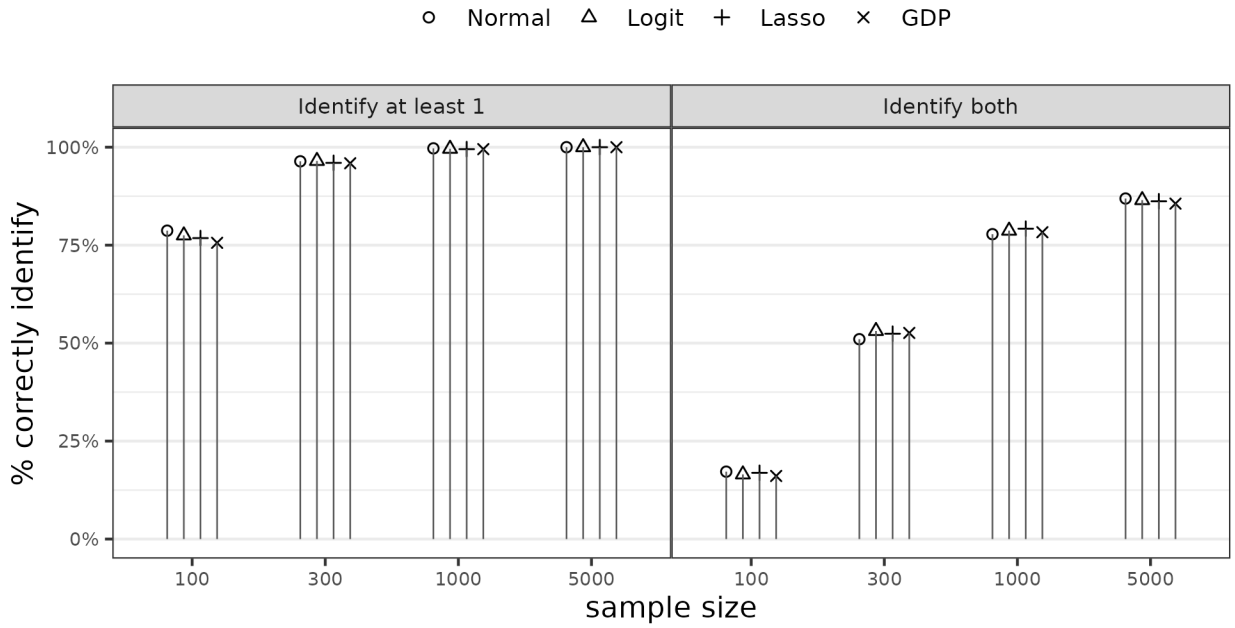
Distribution of standardized residual covariances from simple *model.*



Note. The left panel shows the distribution of residuals for all SRCs except both pairs of randomly selected residual covariances, the *null SRCs*. The right panel shows the distribution for the positive pair of randomly selected residual covariances – the distribution for the negative residual is the negative of the positive residual distribution.

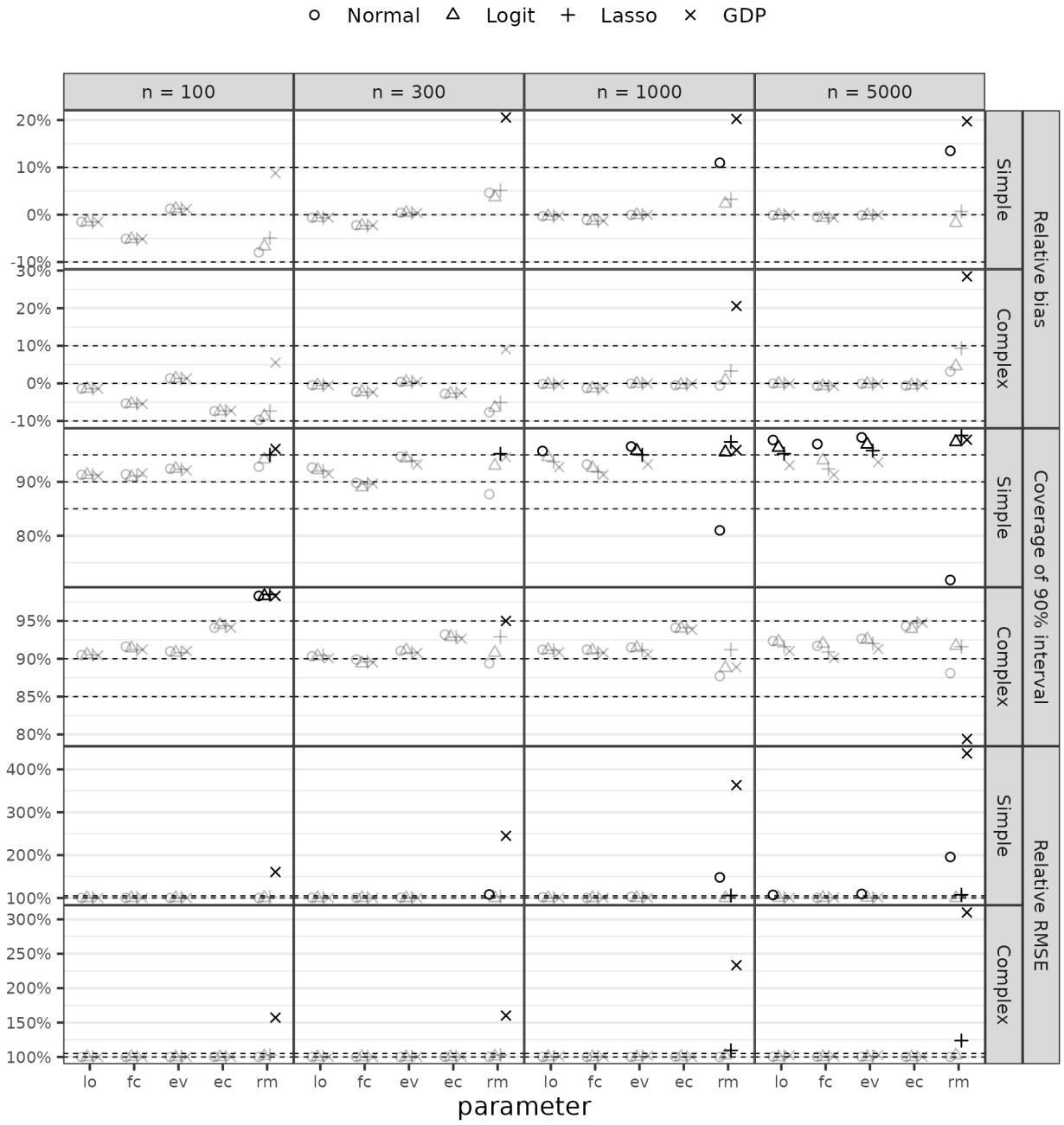
Figure 3

Detecting large SRCs in simple model.



Note. What percentage of the time does the *simple* model identify the two largest SRCs as the non-null SRCs?

Figure 4
Parameter estimation and recovery.



Note. lo = loading, fc = factor correlation, ev = error variance, ec = error correlation, rm = τ. $\pm$10% relative bias, 90% coverage in (85%, 95%), and relative RMSE (compared to minimum RMSE) $<$ 105% are deemed acceptable. Estimates that pass these criteria are faded. Certain relative RMSE values are missing as they were very high ($>$ 1.5). The population parameter for τ under the *simple* model was upwardly adjusted to account for both δ_{ij} parameters.