Bayesian structural equation models of correlation matrices

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

Abstract

We present a method for Bayesian structural equation modeling of sample correlation matrices as correlation structures. The method transforms the sample correlation matrix to an unbounded vector using the matrix logarithm function. Bayesian inference about the unbounded vector is performed assuming a multivariate-normal likelihood, with a mean based on the transformed model-implied correlation matrix, and a covariance assumed to be of known form. Using Monte Carlo simulation, we examine the performance of the method with normal and ordinal indicators, as well as the capacity of the method to estimate models that account for misspecification. The performance of the approach is often adequate suggesting that the proposed method can be used for Bayesian analysis of correlation structures. We conclude with discussion of potential applications of the approach, as well as future directions needed to further develop the method.

Keywords: Bayesian SEM, correlation matrix, polychoric correlations, model misspecification

Structural equation modeling (SEM) is a popular statistical method for modeling covariance matrices under the assumption that the covariance matrices are structured. And the investigator is often interested in the structural parameters underlying the observed covariance structure. In practice, SEM is often performed with correlation matrices as opposed to covariance matrices. For example, only the correlation matrix may be available or computable, such as for latent continuous variables underlying observed ordinal indicators. Correlation matrices differ from covariance matrices primarily in that the diagonal elements of correlation matrices are known to be one. Consequently, procedures for analysis of covariance matrices should not be directly applied to correlation matrices.

In practice, most modellers analyze sample correlation matrices as if they are covariance

The Version of Record of this manuscript when published will be available in Structural Equation Modeling: A Multidisciplinary Journal (2024) https://doi.org/10.1080/10705511.2024.2306532

structures. There are two conditions for such an analysis to be valid: (i) the model is scale-invariant, and (ii) the model-implied covariance matrix must have a unit diagonal (Cudeck, 1989). Most psychometric models (without parameter constraints) are scale-invariant, thus inference about scale-free parameters will be adequate. Parameters that are not scale-free will need to be appropriately transformed for inference about them to be adequate. However, parameter constraints are sometimes necessary as part of theory testing, such that many models estimated in practice are not scale-invariant. And as models become more complex, it can be difficult to identify whether a model is scale-invariant and which parameters are scale-free. Hence, it is conceptually appealing to analyze sample correlation matrices as correlation structures. To analyze correlation matrices as correlation structures, observed variable variances need to be constrained to 1 during model estimation, i.e. observed variable residual variances are computed as opposed to estimated. Additionally, one needs an estimation method that implies that the variances of (and thus covariances with) observed variable variance parameters are 0. This can be accomplished in frequentist SEM using (variants of) weighted least squares estimation (Browne, 1984).

Bayesian SEM of covariance matrices is often performed assuming a Wishart likelihood, following from the assumption that the data are multivariate normal. However, it is inadequate to apply the Wishart likelihood to analysis of correlation matrices unless one intends to assume the sample correlation matrix is simply a re-scaled covariance structure and both conditions (i) and (ii) above are met.

In this paper, we present a Bayesian SEM for analysis of sample correlation matrices as correlation structures. The method is based on the transformation of the correlation matrix to an unbounded scale (Archakov & Hansen, 2021). Conceptually, the method may be described as a 'multivariate Fisher transformation' for correlation matrices. Although the likelihood is defined on an unbounded scale, the model-implied covariance matrix remains specified on the correlation scale hence parameters retain their typical definition. The method has promise for a variety of Bayesian SEMs of correlation matrices. For example, we demonstrate the adequacy of the method for Pearson, tetrachoric and polychoric correlation matrices.

In the next section of the paper, we briefly describe the method. Afterwards, we use Monte Carlo simulation studies to evaluate the adequacy of the method applied to several problems featuring Bayesian SEMs of correlation matrices. All scripts for simulation studies and data analyses

are available at `https://osf.io/ex8gw/`.

Matrix-logarithm transformation of correlation matrices

Much of this section restates results in Archakov and Hansen (2021), hereafter AH (2021). AH (2021) proposed the matrix logarithm transformation of correlation matrices which places correlation matrices on an unconstrained scale. The matrix logarithm of a correlation matrix, $\mathbf{P}$, (or any positive-definite matrix) is based on its eigendecomposition: $\log(\mathbf{P}) = \mathbf{Q}\log(\mathbf{\Lambda})\mathbf{Q}'$, where $\mathbf{Q}$ is an orthogonal matrix and $\log(\mathbf{\Lambda})$ is the diagonal matrix of the logarithm of eigenvalues of $\mathbf{P}$. Precisely, the strict half-vectorization function (vech$(\cdot)$, lower or upper triangle excluding diagonal) of the matrix-logarithm of a correlation matrix ($\boldsymbol{\gamma} = \text{vech}(\log(\mathbf{P}))$) contains all the information needed to reconstruct the correlation matrix. Different correlation matrices cannot result in the same $\boldsymbol{\gamma}$ and vice-versa. Finally, let $\gamma(\cdot) := \text{vech}(\log(\cdot))$, such that $\gamma^{-1}(\cdot)$ transforms the unconstrained vector back to its correlation matrix. AH (2021) provide an algorithm for computing $\gamma^{-1}(\cdot)$.

For our purposes, the asymptotic distribution of the transformed vector is of importance. Given a $p \times p$ population correlation matrix, $\mathbf{P}$, and $\boldsymbol{\rho} = \text{vech}(\mathbf{P})$ where $\boldsymbol{\rho}$ is a p^*-dimensional vector ($p^* = p \times (p-1)/2$), the asymptotic distribution of sample correlations, $\mathbf{r}$ (vech$(\mathbf{R})$), is $\mathcal{N}_{p^*}(\boldsymbol{\rho}, n^{-1}\mathbf{\Omega})$, where n is sample size and $\mathbf{\Omega}$ is assumed to be of known form in special cases (e.g. Browne & Shapiro, 1986; Neudecker & Wesselman, 1990). Based on the multivariate delta method, the asymptotic distribution of $\mathbf{y}$ $(= \gamma(\mathbf{R}))$ then is:

$$\mathbf{y} \sim \mathcal{N}_{p^*}(\boldsymbol{\gamma}, n^{-1}\mathbf{\Omega}_\gamma),\ \mathbf{\Omega}_\gamma = \mathcal{J}\mathbf{\Omega}\mathcal{J}',\ \mathcal{J} = \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\rho}} \tag{1}$$

AH (2021) provide an expression for $\mathcal{J}^{-1}$. Moreover, $\mathcal{J}^{-1}$ is often approximately diagonal (especially for larger p, Archakov, Hansen, & Luo, 2022), such that increasing/reducing the variance of specific elements in $\mathbf{y}$ will for the most part increase/decrease the variance of corresponding elements in $\mathbf{r}$.

A limited experiment to show desirable properties of the $\gamma(\cdot)$ transformation

Our goal is to develop a Bayesian SEM given input data, $\mathbf{R}$, and the transformed vector, $\mathbf{y}$. For Bayesian methods to be credible, the likelihood must represent a plausible data generation mechanism for the data, since data can only influence inference via the likelihood.[1] In our case, the multivariate normal assumption for $\mathbf{y}$ must be accurate. We describe the results of a small experiment to demonstrate this. We assume the following data generation process: $99 \times \mathbf{S}_i \sim \mathcal{W}_p(99, \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Delta})$ for $i \in \{1, \ldots, 5000\}$, i.e. there are 5000 covariance matrices for multivariate normal data each of sample size 100. $\mathbf{\Lambda}' = \left[\begin{smallmatrix} \lambda & \lambda & 0 & 0 \\ 0 & 0 & \lambda & \lambda \end{smallmatrix}\right]$ and $\mathbf{\Phi} = \left[\begin{smallmatrix} 1 & \\ \phi & 1 \end{smallmatrix}\right]$; and the diagonal of $\mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Delta}$ is constrained to be 1. There are 3 λ-conditions: $\lambda \in \{.4, .7, .9\}$ and two ϕ-conditions: $\phi \in \{0, .7\}$. Each generated covariance matrix ($\mathbf{S}_i$) is transformed to a sample correlation matrix, $\mathbf{R}_i$, after which we compute $\mathbf{y}_i = \gamma(\mathbf{R}_i)$.

Under all six conditions, $\mathbf{y}_i$ should be multivariate normal – this was examined using Henze-Zirkler tests of multivariate normality (Henze & Zirkler, 1990) which yielded p-values between .069 and .67. Thus, we are unable to reject the null of multivariate normality of $\mathbf{y}_i$ vectors across conditions. On the other hand, when $\lambda = .4$, the Henze-Zirkler test p-values for $\mathbf{r}_i$ yielded .043 and .006; otherwise, $p < .001$. As one might expect, larger correlations (induced by large values of λ) are less likely to be normally distributed. The $\gamma(\cdot)$ transformation is able to resolve this problem, justifying the application of this likelihood in Bayesian modeling of $\mathbf{y}$.

An additional property of $\gamma(\cdot)$ is that $\mathbf{\Omega}_\gamma$ is often approximately diagonal, with select non-zero elements. For example, large values of λ and ϕ can create large correlations between elements of $\mathbf{r}_i$, while the correlations between elements of $\mathbf{y}_i$ are negligible, see rows 1 and 2 of Table 1. Additionally, large λ and negligible ϕ should create dispersion in $\text{Var}(\sqrt{n}\mathbf{r}_i)$ across elements of $\mathbf{r}_i$, see row 3 of Table 1. However, $\gamma(\cdot)$ although not variance-stabilizing shrinks the variances of different elements towards each other (row 4 of Table 1). These properties of $\mathbf{\Omega}_\gamma$ are explored in depth by AH (2021) and Archakov et al. (2022), and may be useful for incorporating regularization when using the $\gamma(\cdot)$ transformation.

[1] This is in contrast to frequentist SEM which contains a variety of *robust* methods for valid frequentist inference under misspecified likelihoods or models.

Table 1

Descriptions of covariance matrices of $\mathbf{r}_i$ *&* $\mathbf{y}_i$ *when* $\lambda = .9$

ϕ	Statistic	Min.	25th perc.	Median	Mean	75th perc.	Max.	P($\lvert x \rvert < .05$)
.7	Cor($\mathbf{r}_i$)	.20	.35	.36	.47	.64	.73	
.7	Cor($\mathbf{y}_i$)	−.032	−.010	−.001	.029	.020	.20	$^{12}/_{15} = 80\%$
ϕ	Diagonal elements of covariance matrix							
.0	Var($\sqrt{n}\mathbf{r}_i$)	0.12	1.02	1.04	1.01	1.03	0.12	
.0	Var($\sqrt{n}\mathbf{y}_i$)	1.04	0.87	0.88	0.87	0.88	1.03	

A Bayesian correlation structures analysis

Assuming $\mathbf{R}$ is a $p \times p$ sample correlation matrix, $\mathbf{r} = \text{vech}(\mathbf{R})$ and $\mathbf{r}$ has asymptotic covariance matrix, $\mathbf{\Omega}$, then the estimated SEM for $\mathbf{R}$ will be:

$$\mathbf{y} \sim \mathcal{N}_{p^*}(\boldsymbol{\gamma},\ n^{-1}\widehat{\mathbf{\Omega}}_\gamma),\ \mathbf{y} = \gamma(\mathbf{R}),\ \boldsymbol{\gamma} = \gamma(\mathbf{P}(\boldsymbol{\theta})),\ \widehat{\mathbf{\Omega}}_\gamma = \widehat{\mathcal{J}}\widehat{\mathbf{\Omega}}\widehat{\mathcal{J}}',\ \widehat{\mathcal{J}} = \frac{\partial \mathbf{y}}{\partial \mathbf{r}} \tag{2}$$

where $\mathbf{P}(\boldsymbol{\theta})$ is the model-implied correlation matrix i.e. $\mathbf{P}(\boldsymbol{\theta})$ may be of any form e.g. $\mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Delta}$ in the case of confirmatory factor models, with the restriction that diagonal($\mathbf{P}(\boldsymbol{\theta})$) = $\mathbf{1}_p$. And $\widehat{\mathbf{\Omega}}$ is an available consistent estimate of $\mathbf{\Omega}$. This allows for Bayesian SEMs with tetrachoric and polychoric correlation matrices as inputs. Replacing $\widehat{\mathbf{\Omega}}_\gamma$ with $\mathbf{\Omega}_\gamma$ in equation 2 results in a data generation process (DGP) for correlation matrices that always produces valid correlation matrices. This makes this DGP uniquely different from a process that assumes correlations or their Fisher z-transformations are multivariate normal.

We next use simulation studies to test the application of this method to some SEM problems. And for the remainder of the paper, we term the method the *log-correlation* method.

Simulation studies

In this section, we conduct three simulation studies to study the log-correlation approach to Bayesian SEM of correlation matrices. We describe procedures shared across the simulation studies. The Bayesian computation engine was Stan (Carpenter et al., 2017). We ran each model across 3 chains, requested 2000 posterior samples, and retained the final 1000 samples, resulting in 3000 posterior samples for inference per parameter in each model. The following priors were retained across all models:

$$\boldsymbol{\lambda} \sim \text{truncated-Normal}(0, \sigma_\lambda, -1, 1),\ \sigma_\lambda \sim \mathcal{N}^+(0, 0.25),\ \frac{\phi + 1}{2} \sim \text{Beta}(2, 2) \tag{3}$$

where $\boldsymbol{\lambda}$ are structurally non-zero loadings and ϕ is an interfactor correlation parameter. Loadings were assumed bound to the [-1, 1] interval. This restriction is not necessarily correct as standardized loadings in oblique factor models can exceed this interval. But without this restriction, the sampler often terminated early.

The log-correlation analysis model was based on equation 2, i.e. $\widehat{\boldsymbol{\Omega}}_\gamma$ was based on the observed data. Each simulation study had 1000 replications. And the primary goal of the studies was assessment of parameter recovery; we assessed (i) relative bias deeming absolute relative bias under 5% and 10% as ideal and acceptable respectively; (ii) empirical coverage rate of the 90% quantile interval deeming coverage between 85% and 95% as acceptable. We also explored the relative bias of posterior standard deviations (SDs) to (a) better help understand any problems with coverage; (b) assess how well the posterior SDs capture parameter uncertainty.

We additionally computed posterior predictive p-values for all Bayesian models. The discrepancy function was based on the multivariate normal likelihood on the transformed $\mathbf{y}$ vector. The distributions of pp p-values for the different simulation studies are reported in part A of the online supplementary materials. Finally, for each simulation study, over 99.9% of parameters had $\widehat{R}$ values under 1.01, suggesting no problems with parameter convergence across the simulation studies.

Study 1: A study of the log-correlation method across varying conditions

In this study, we examined the log-correlation across a variety of conditions. The data generation process (DGP) for the study was:

$$
\begin{aligned}
\mathbf{y} \sim & \mathcal{N}_{p^*}(\boldsymbol{\gamma},\ n^{-1}\boldsymbol{\Omega}_\gamma),\ \boldsymbol{\gamma} = \gamma(\mathbf{P}(\boldsymbol{\theta})),\ \mathbf{R} = \gamma^{-1}(\mathbf{y}), \\
& \mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Delta},\ \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 & \boldsymbol{\lambda}_2 & \boldsymbol{\lambda}_3 \end{bmatrix}, \\
& \boldsymbol{\lambda}_1 = [\underbrace{\lambda, \ldots, \lambda}_{p/3 \text{ times}}, \underbrace{0, \ldots, 0}_{2p/3 \text{ times}}]',\ \boldsymbol{\lambda}_2 = [\underbrace{0, \ldots, 0}_{p/3 \text{ times}}, \underbrace{\lambda, \ldots, \lambda}_{p/3 \text{ times}}, \underbrace{0, \ldots, 0}_{p/3 \text{ times}}]', \\
& \boldsymbol{\lambda}_3 = [\underbrace{0, \ldots, 0}_{2p/3 \text{ times}}, \underbrace{\lambda, \ldots, \lambda}_{p/3 \text{ times}}]' \\
& \boldsymbol{\Phi} = \begin{bmatrix} 1 & & \\ \phi & 1 & \\ \phi & \phi & 1 \end{bmatrix},\ \boldsymbol{\Delta} = \text{diag-matrix}\left(\text{diagonal}\left(\mathbf{I}_p - \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\right)\right) \\
& n \in \{100, 300, 500, 2000\},\ p \in \{9, 18, 27\},\ \lambda \in \{.4, .7, .9\},\ \phi \in \{0, .3, .6\}
\end{aligned}
\tag{4}
$$

where $\mathbf{R}$ was the input for all analyses. In this DGP, the asymptotic covariance of $\mathbf{r}$, $\boldsymbol{\Omega}$ (used in calculating $\boldsymbol{\Omega}_\gamma$, equation 1), is based on $\boldsymbol{\rho}$ (e.g. equations 3–5 in Olkin & Finn, 1995). This DGP produces sample correlation matrices under assumptions of multivariate-normality.

The model was a three factor model. The design conditions of the study are given on the last line of equation 4. We wanted the conditions to reflect the variety of scenarios common in factor analysis models. Sample sizes (4 conditions) varied from small ($n = 100$) to large ($n = 2000$). The total number of indicators per factor (3 conditions) varied from relatively small ($p/3 = 3$) to moderately sized ($p/3 = 9$). The strength of factor loadings (3 conditions) varied from weak ($\lambda = .4$, 16% variance explained) to very strong ($\lambda = .9$, 81% variance explained). And the correlation between both factors (3 conditions) varied from non-existent ($\phi = 0$) to relatively large ($\phi = .6$) – large enough that a high-order factor could be credibly hypothesized. This resulted in 108 design conditions in total ($4 \times 3 \times 3 \times 3$).

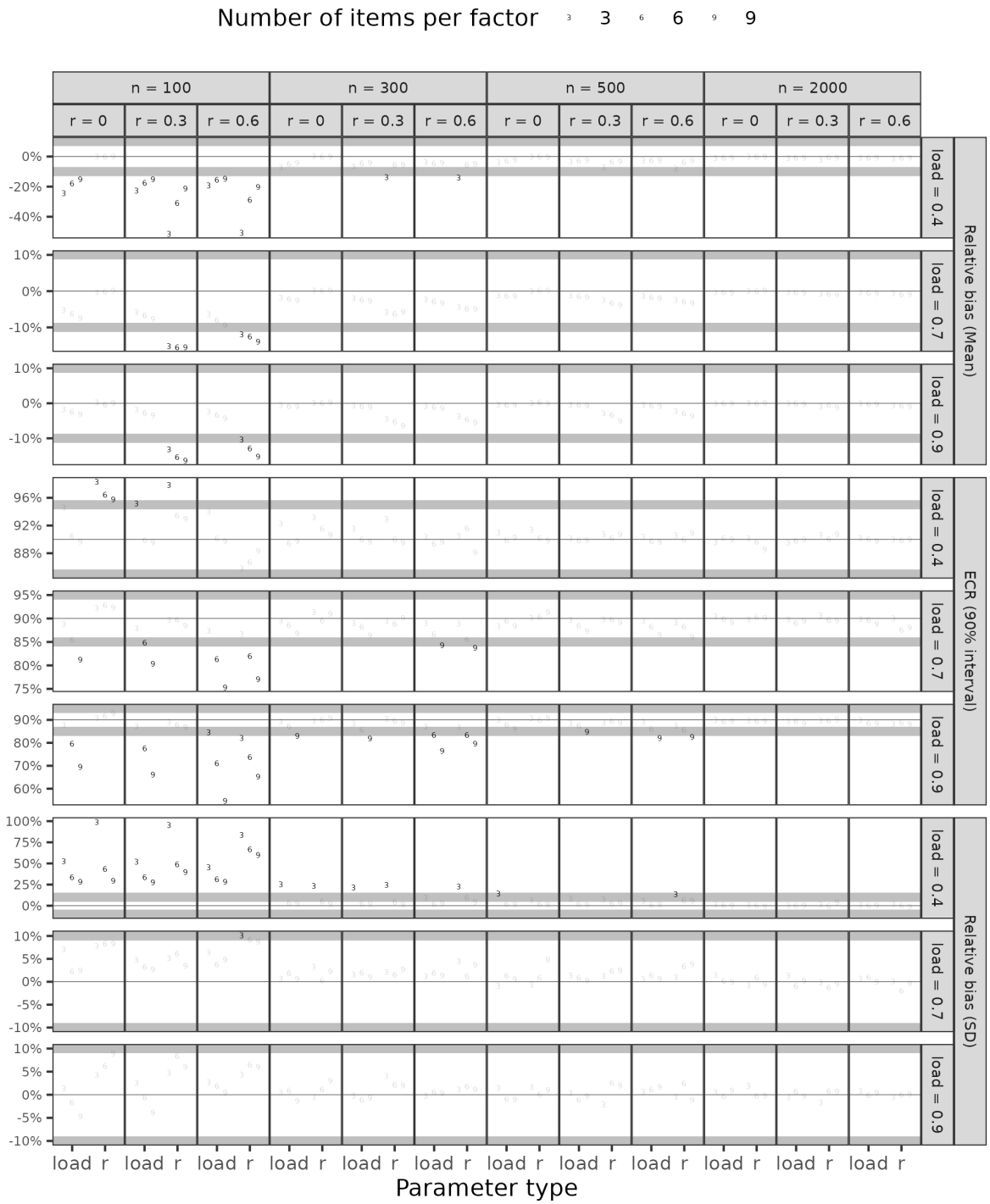
We were interested in parameter recovery for structural parameters: $\boldsymbol{\lambda}$ and $\boldsymbol{\phi}$. We assessed the recovery of both sets of parameters on average since they were constant within design condition.

Study 1 results

We report parameter recovery in Figure 1. The log-correlation method was largely adequate once sample size was at least 300. However, there was a pattern of under-coverage when the number of indicators was 27. When sample size was 100, the log-correlation method tended to be

Figure 1

Simulation study 1 parameter recovery



Note. load = average loading estimate, r = average interfactor correlation. We report bias not relative bias when $\phi = 0$. Thick lines show bounds for acceptable results as defined in the paper text. Estimates within these bounds are faded, while inadequate estimates are not to draw attention to inadequacies.

downwardly biased with several instances of under-coverage. The relative bias of posterior SDs was often adequate at this sample except when $\lambda = 0.4$. This suggests the patterns of under-coverage are related to the downward bias of parameter estimates at this sample size. In summary, the log-correlation method performs adequately for the sample sizes that would be recommended when assessing the quality of psychometric instruments. However, in line with commonplace advice on sample sizes, larger sample sizes are preferable when dealing with a larger number of indicators (e.g. Bentler & Chou, 1987; MacCallum, Widaman, Zhang, & Hong, 1999).

Study 2: Assessing the log-correlation method for categorical indicators

A common strategy for analysis of ordinal indicators in SEM is the two-step strategy: step 1, compute the polychoric correlation matrix alongside its asymptotic variance matrix; step 2, use the polychoric matrix and its asymptotic variance as input to estimate the SEM structured parameters. Hence, given the potential of the log-correlation method for correlation matrices broadly conceived, we also decided to assess the method when applied to polychoric correlations. This study is smaller in scope than study 1, as several more factors gain importance for analysis of ordinal data, and the application of any method to ordinal data is worth its own investigation. Hence our goal here is to provide some initial results to support the use of the log-correlation method with ordinal indicators.

The DGP was:

$$
\begin{aligned}
&\mathbf{X}^* \sim \mathcal{N}_p\left(\boldsymbol{\kappa},\ \mathbf{P}(\boldsymbol{\theta})\right),\ \mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Delta},\ \boldsymbol{\Lambda}' = \left[\begin{smallmatrix} 0.7 & 0.8 & 0.6 & 0.9 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.75 & 0.85 & 0.6 \end{smallmatrix}\right], \\
&\quad \boldsymbol{\Phi} = \left[\begin{smallmatrix} 1 & \\ .3 & 1 \end{smallmatrix}\right],\ \boldsymbol{\Delta} = \text{diag-matrix}\left(\text{diagonal}\left(\mathbf{I}_p - \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\right)\right), \\
&\quad \kappa_j = \Phi^{-1}(.15) + \frac{j-1}{10-1}\left[\Phi^{-1}(.85) - \Phi^{-1}(.15)\right] \text{ for } j \in \{1, \ldots, 10\}
\end{aligned}
\tag{5}
$$

where $\mathbf{X}^*$ is an $n \times p$ multivariate-normal matrix of latent continuous data, and $\Phi(\cdot)^{-1}$ is the standard normal quantile function such that the latent variables range in their means from low to high. We varied the number of observed response categories: $k \in \{2, 3, 5\}$, reflecting common number of response categories for ordinal indicators (e.g. multiple-choice and Likert response scales). We set the threshold vector, $\boldsymbol{\tau}_k$, defined for each value of k as: $\boldsymbol{\tau}_k = \Phi^{-1}\left(\frac{[0, 1, \ldots, k]}{k}\right)$. The observed data, $\mathbf{X}$, with k response categories were then computed from $\mathbf{X}^*$ and $\boldsymbol{\tau}_k$: $x_{ij} =$

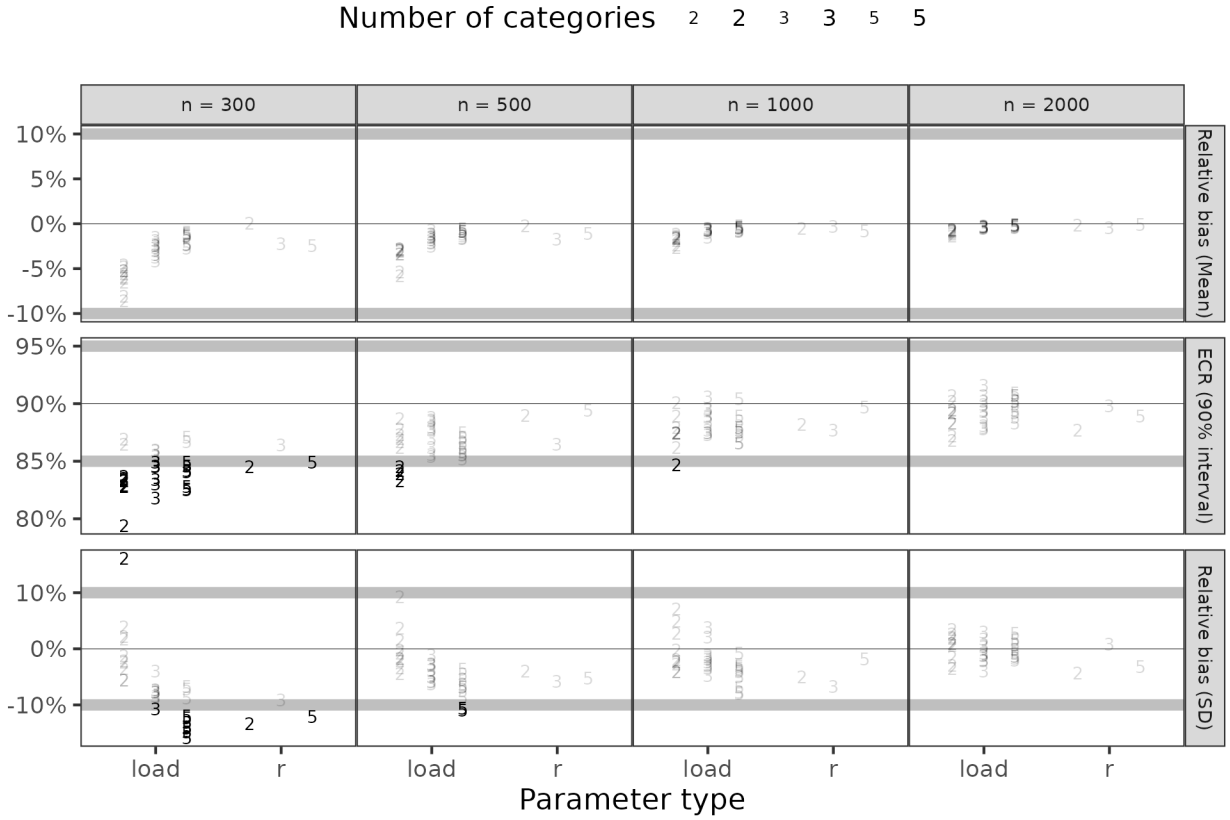
c if $\tau_{k,c} < x^*_{ij} < \tau_{k,c+1}$ for $i \in \{1, \ldots, n\},\ j \in \{1, \ldots, p\}$ and $c \in \{1, \ldots, k\}$.

For each iteration in the simulation study, the polychoric correlation matrix and its asymptotic covariance were calculated using lavaan (version 0-6.16, Rosseel, 2012) using a two-step estimator (Jöreskog, 1994) and were the input data for the Bayesian log-correlation model.

We also varied sample size: $n \in \{300, 500, 1000, 2000\}$. Compared to the earlier studies, we excluded $n = 100$ because we have low expectations that categorical data analysis methods will be adequate when $n = 100$ and included $n = 1000$ so as to better discriminate model performance at more adequate sample sizes for categorical data.

Figure 2

Simulation study 2 parameter recovery



Note. load = 10 loading estimates, r = interfactor correlation. Thick lines show bounds for acceptable results as defined in the paper text. Estimates within these bounds are faded, while inadequate estimates are not to draw attention to inadequacies.

Study 2 results

We report parameter recovery in Figure 2. Relative bias was always acceptable though bias lessened with increased sample size and at increased number of response categories. When sample size was 300, there was under-coverage for most loading parameters and the interfactor correlation. When sample size was 500 or above, the loadings for some binary indicators demonstrated under-coverage, however, at 2000, both bias and coverage were adequate regardless of the number of response categories. These results suggest the adequacy of the log-correlation method applied to Bayesian SEM of tetrachoric and polychoric correlation matrices, with some caveats. Larger sample sizes are needed for smaller number of response categories. These challenges likely stem from problems estimating polychoric correlations, as such problems are more common when the number of categories is smaller (Flora & Curran, 2004). Additionally, we can expect estimation to be less adequate when the number of indicators is larger due to challenges estimating the asymptotic covariance of the polychoric correlations (Browne, 1984; Flora & Curran, 2004).

Study 3: Assessing the log-correlation method for modeling misspecification

One way to think of Bayesian SEMs is to consider them as an alternative estimation method, often deployed whenever frequentist SEMs prove inadequate. An alternative is to appreciate Bayesian SEMs for the unique opportunities they provide. Muthén and Asparouhov (2012) demonstrated two such opportunities: estimation of a full residual covariance matrix, and estimation of all cross-loadings. In their demonstration, the modeler uses small variance priors to estimate parameters that would normally be unidentifiable by attempting to constrain them to 0. The product is more flexible SEMs that are often more realistic descriptions of the patterns in data. Uanhoro (2023b) further developed the estimation of the full residual covariance matrix by placing a prior on the standard deviation of residual covariances instead of requiring the modeler to set the scale of a small variance prior. After transformation, the standard deviation of residual covariances serves as a goodness of fit index similar to the correlation root mean squared residual (CRMR). Additionally, the method falls within an alternative SEM tradition that believes real data already contain deviations from hypothesized structures at the population level (e.g. MacCallum & Tucker, 1991; Tucker, Koopman, & Linn, 1969), and that such deviations are random varying

across different studies (e.g. Robitzsch, 2023; Wu & Browne, 2015).

Given this background, we were also interested in the ability of the log-correlation method to incorporate the modeling of misspecification. Practically, this would permit the log-correlation method to return: (i) a goodness of fit index; (ii) uncertainty intervals for structural parameters that reflect the amount of model misspecification. We characterize this misspecification as reflecting the influence of minor factors which vary across replications of a study. As with study 2, the purpose of this study is to provide initial results into the application of the log-correlation method for models under misspecification. The exact DGP was:

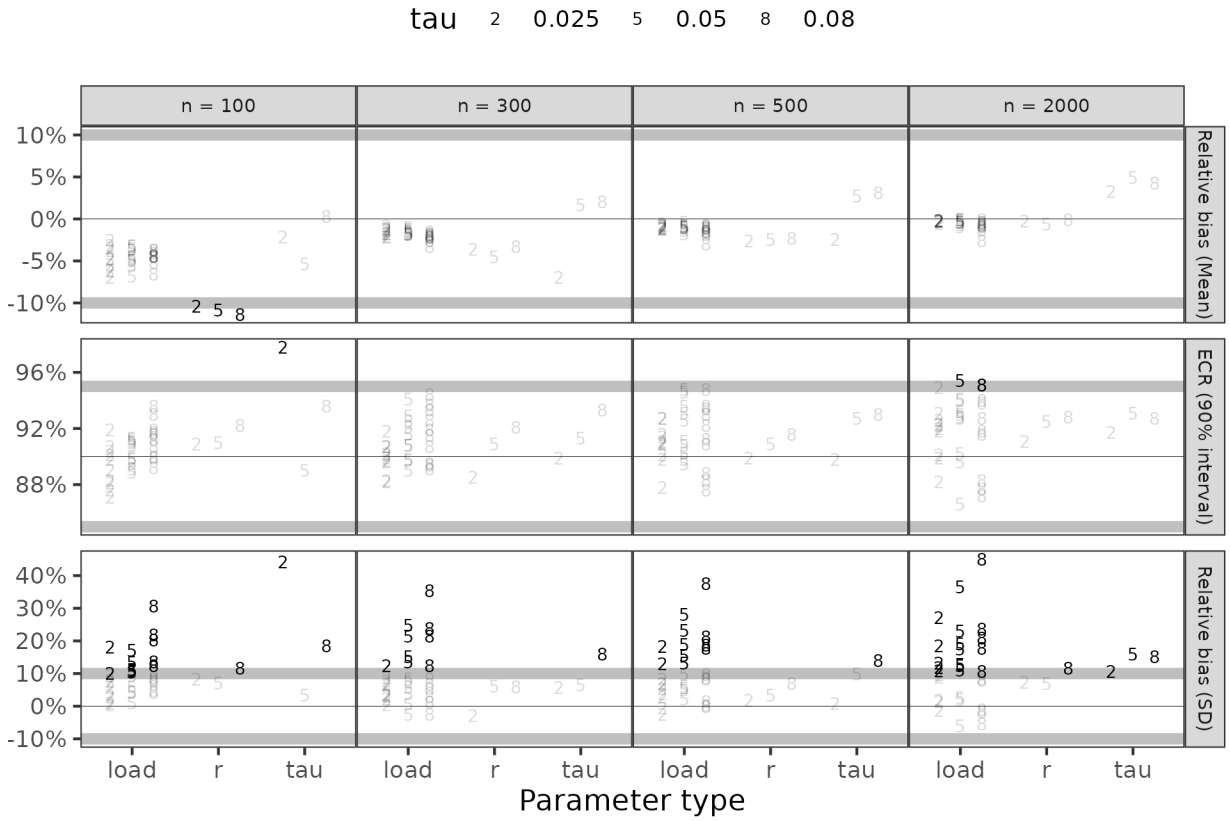
$$
\begin{aligned}
&\mathbf{y} \sim \mathcal{N}_{p^*}(\boldsymbol{\gamma},\ n^{-1}\boldsymbol{\Omega}_\gamma),\ \boldsymbol{\gamma} = \gamma(\mathbf{P}(\boldsymbol{\theta})),\ \mathbf{R} = \gamma^{-1}(\mathbf{y}),\\
&\quad \mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Delta},\ \boldsymbol{\Lambda}' = \left[\begin{smallmatrix} 0.7 & 0.8 & 0.6 & 0.9 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8 & 0.7 & 0.75 & 0.85 & 0.6 \end{smallmatrix}\right],\ \boldsymbol{\Phi} = \left[\begin{smallmatrix} 1 & \\ .3 & 1 \end{smallmatrix}\right]\\
&\quad \boldsymbol{\Delta} = \mathbf{DMD},\ \mathbf{D} = \text{diag-matrix}\left(\text{diagonal}(\mathbf{I}_p - \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda})^{\frac{1}{2}}\right),\ \mathbf{M} \sim \text{LKJ}(\eta)
\end{aligned}
\tag{6}
$$

where the residual variance matrix ($\boldsymbol{\Delta}$) reflects the correlations induced by minor factors in $\mathbf{M}$. Since $\mathbf{M}$ is simulated from the LKJ distribution, none of its off-diagonal elements (minor factor influences) are exactly 0, such that any traditional SEM will have less than perfect fit. We varied the typical size of minor factor influences (τ, akin to the correlation root mean squared residual) by altering the value of η. Precisely: $\eta = \frac{1}{2}\left[\left(\frac{\bar{\delta}}{\tau}\right)^2 - p + 1\right]$, where $\bar{\delta} = \text{mean}(\text{diagonal}(\boldsymbol{\Delta}))$, i.e. setting $\eta = \{170.345, 39.211, 12.575\}$ corresponds to $\tau = \{0.025, 0.05, 0.08\}$.

Accordingly, the model-implied correlation matrix in the log-correlation analysis model was assumed to reflect minor factor influences. Differently from Uanhoro (2023b) who assumed the residual correlations in $\boldsymbol{\Delta}$ to be normally distributed, we assumed them to logistic distributed: $\text{vech}(\boldsymbol{\Delta}) \sim \text{Logistic}\left(0, \tau\frac{\sqrt{3}}{\pi}\right)$, implying a standard deviation of τ. Compared to a normal distribution assumption, the logistic prior – as a consequence of being more heavy-tailed – more strongly shrinks residual correlations close to zero, and less strongly shrinks residual correlations far from zero. For the prior on τ, we assumed $\tau \sim \mathcal{N}^+(0, 1)$ though it would be reasonable to assume a much smaller standard deviation in this prior as $0 \leq \tau \ll 1$. Finally, we also varied sample size: $n \in \{100, 300, 500, 2000\}$.

Figure 3

Simulation study 3 parameter recovery



Note. load = 10 loading estimates, r = interfactor correlation, tau = τ. Thick lines show bounds for acceptable results as defined in the paper text. Estimates within these bounds are faded, while inadequate estimates are not to draw attention to inadequacies.

Study 3 results

We report parameter recovery in Figure 3. When sample size was 300 or above, bias and coverage were almost always acceptable. The exceptions were one loading parameter when $n = 2000$ and $\tau \in \{.05, .08\}$. There were notable problems when sample size was 100: the interfactor correlation was notably downwardly biased; and the interval about τ over-covered when $\tau = 0.025$. Across all conditions, the posterior SDs were often upwardly biased. Given the adequate coverage of intervals (for sample sizes of at least 300) and the primacy of intervals over posterior standard deviations for summarizing parameter distributions in Bayesian analysis, the undesirable pattern of upwardly biased posterior standard deviations is not a problem.

Having shown the flexibility of the log-correlation method for a variety of problems, we next

demonstrate the approach with real data.

Data demonstration

The log-correlation method for correlation structure analysis is implemented in the minorbsem R package (Uanhoro, 2023a), a Bayesian SEM package built atop Stan and the package is used for the data demonstration.

A typical example

In this demonstration, we show the log-correlation method applied to a typical example, while including unique features of Bayesian SEM. The example is from chapter 1 of Mulaik (2009). Two hundred and twenty-five soldiers were asked to rate the meaning of "firing my rifle in combat" using 14 scales. The 14 scales were assumed to reflect two factors: the degree of *emotional* concern for the danger of the situation; and the degree of *optimism* in being able to cope with the situation. The exact data is the correlation matrix of the 14 scales, available in the accompanying demonstration script.

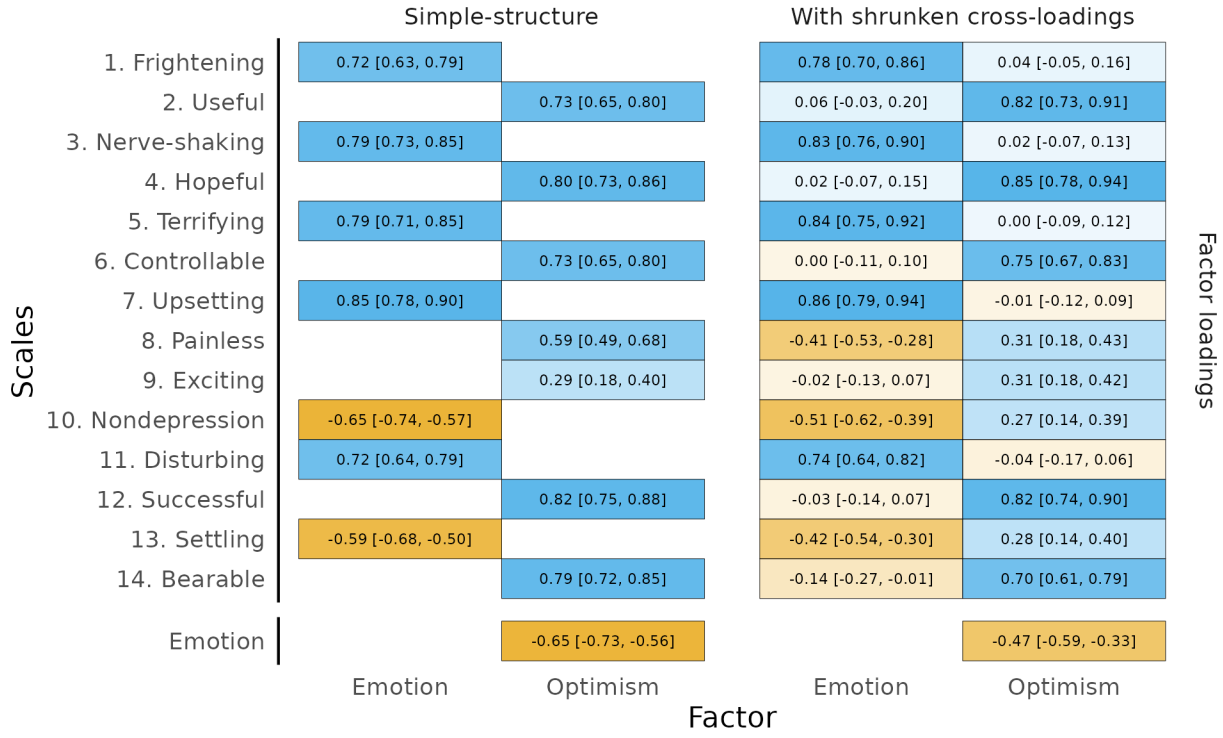
We fit the models under the assumption of minor factor influences – in this situation, the posterior predictive p-value (PPp-value) would often be adequate, so we examine the average size of residual correlations (τ) and the residual correlation matrix to identify patterns of misspecification.

We began with the hypothesized two-factor structure with half the scales reflecting emotion and the other half reflecting optimism – the scales are listed in Figure 4. As shown in the left panel of Figure 4, the loading between the optimism factor and the exciting scale was 0.29; all other loadings were above 0.55 on average suggesting reasonable alignment between factor and scales. The correlation between the factors (-.65) suggests a high negative relation between emotion and optimism factors. In terms of misspecification, the PPp-value was .056 and τ was .055, 90% CrI [.042, .071]. On examining the individual residual correlations, there were multiple notable large values (see lower-triangular matrix in Figure 5). These residual correlations either reflect high influence of minor factors or an inadequate hypothesized factor structure.

As an alternative, we estimated a model with all cross-loadings estimated but shrunken to zero (e.g. J. Chen, Guo, Zhang, & Pan, 2021; Muthén & Asparouhov, 2012) – the parameter estimates are shown on the right panel of Figure 4. Several cross-loadings were successfully shrunken

Figure 4

Mulaik example: structural parameters, preliminary models

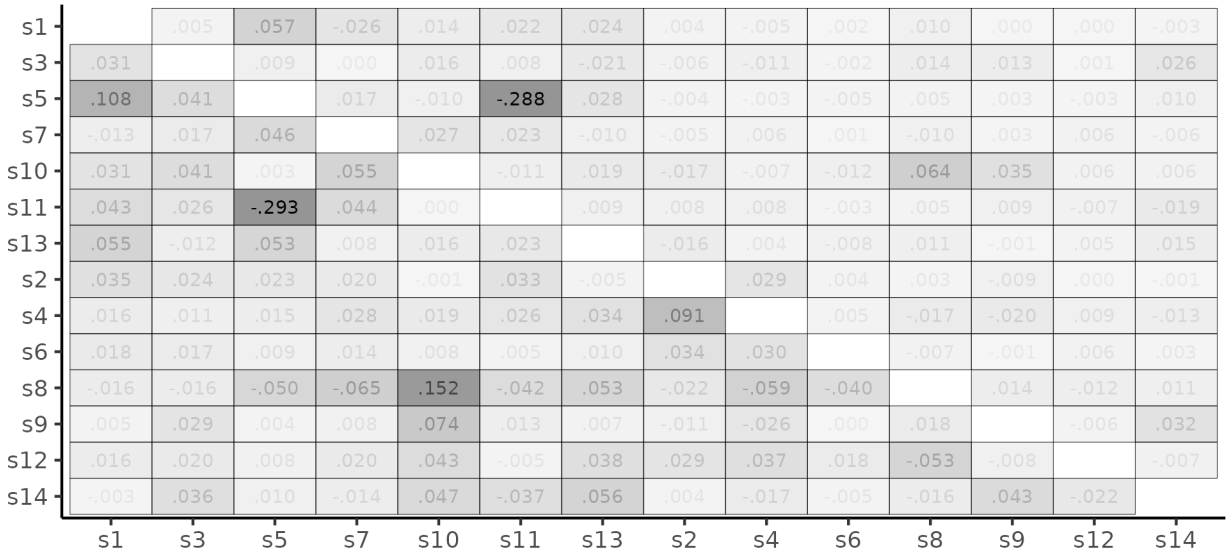


Note. The first fourten rows of data contain factor loadings, while the last row of data contains interfactor correlations. The text in each cell follows the structure: posterior-median [with 90% quantile interval]. The cell background colour is more faded when the estimate is closer to zero. As an example, the loading from emotion to frightening was 0.72, 90% CrI [0.63, 0.79].

to almost 0 (< 0.1). However, the painless, nondepression, settling and bearable scales had non-ignorable cross-loadings, while exciting did not load strongly on either factor. Additionally, the correlation between the factors reduced (-.65 $\rightarrow$ -.47) suggesting that the earlier model artificially inflated the interfactor correlation as a result of constraining cross-loadings to zero (e.g. Ferrando & Lorenzo-Seva, 2000). In terms of misspecification, the PPp-value was .564 and τ was .034, 90% CrI [.026, .044]. On examining the individual residual correlations, all residual correlations reduced from the first model and were small (<.1; see upper-triangular matrix in Figure 5) except the residual correlation between the terrifying and disturbing scales (-.29). We could add this residual correlation to the hypothesized model acknowledging that it likely reflects a real pattern in the data. However, we chose not to do this because the model with minor factor influences already incorporates this residual correlation and adding it to the hypothesized structure did not alter other

Figure 5

Mulaik example: Residual correlation matrices, preliminary models



Note. The lower-triangular matrix is from the standard two-factor model. The upper-triangular matrix is from the two-factor model with shrunken cross-loadings. Larger estimates are bolded. 's' in the axis labels means scale, so s1 is the first scale, frightening.

structural parameters.

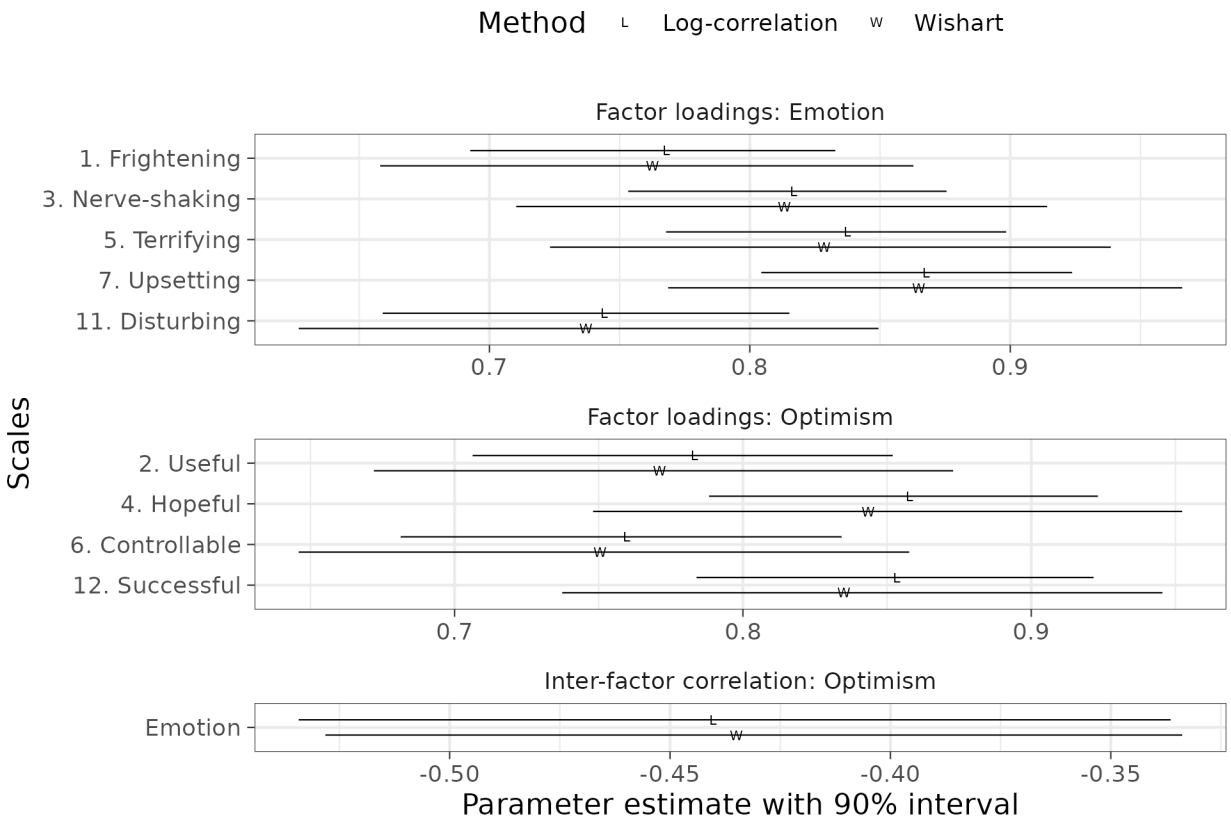
For our third model, we decide to drop weakly and ambiguously loading scales so we could maintain simple structure and strong measurement quality: scales 1, 3, 5, 7 and 11 were assumed to unambiguously reflect emotion and scales 2, 4, 6 and 12 were assumed to unambiguously reflect optimism – see parameter estimates in Figure 6. On fitting the model, the resulting interfactor correlation was -.44 (inline with the second model) and all factor loadings ranged between .74 and .87 suggesting strong alignment between scales and factors. In terms of misspecification, the PPp-value was .577 and τ was .047, 90% CrI [.033, .069]. Outside of the large residual correlation between the terrifying and disturbing scales (-.30), the largest residual correlation was .054, matching the pattern of low misspecification in the second model.[2]

For comparison, we also analyzed the correlation matrix as a covariance structure (i.e. assuming a Wishart distribution). Parameter estimates are in Figure 6. Although the covariance analysis returns similar estimates to the correlation analysis, the intervals returned by the covariance analysis are much larger, except for the interfactor correlation which is a scale-free parameter.

[2] τ was larger in this model likely because the number of small residual correlations reduced since we dropped 5 scales, and the lone large residual correlation remained causing the average size of residual correlations to be larger.

Figure 6

Mulaik example: structural parameters, final models



Note. The lower-triangular matrix is from the standard two-factor model. The upper-triangular matrix is from the two-factor model with shrunken cross-loadings. Larger estimates are bolded. 's' in the axis labels means scale, so s1 is the first scale, frightening.

The covariance structure factor loadings could be made scale-free via standardization since this particular model is scale-invariant.

An example with a Heywood case

We report this example only to demonstrate an atypical feature of correlation structure analysis shared by the Bayesian log-correlation approach. Typically, Bayesian SEMs place non-negative prior distributions on residual variance parameters such that negative residual variances are not possible. In frequentist SEMs, residual variances can be constrained to 0 to avoid such problems. But since the log-correlation approach only computes residual variances as opposed to estimating them, it is possible for computed residual variances to be negative – a situation known as a *Heywood case* (F. Chen, Bollen, Paxton, Curran, & Kirby, 2001). In single factor models

or models where each indicator only reflects a single construct, one can restrict factor loadings to ensure the explained variance in an observed variable does not exceed 1. However, in more complex models, multiple structural parameters can combine to produce an explained variance exceeding 1.

The data for this example is a correlation matrix of nine tests of ability for 696 respondents reported by Harman (1976) and obtained from the psych package in R (Revelle, 2023). The data are assumed to reflect a bifactor structure with the nine tests split equally between three specific factors. On analysis via the log-correlation method, the mixed arithmetic test had median loadings of .73 and .74 on its general factor and specific factor respectively, resulting in a total explained variance of 1.08 ($.73^2 + .74^2$) and a computed residual variance with 90% CrI [-0.32, 0.037]. A Bayesian covariance structure analysis avoids this problem since the estimated residual variance has a non-negative prior distribution, but the 90% CrI for the residual variance was [0.00, 0.053] with a clear mode of 0. A frequentist covariance structure analysis using lavaan also returned a Heywood case with 90% CI [-0.21, 0.087], but this problem can be "fixed" by constraining this parameter to 0.

Discussion

We have presented an approach to Bayesian SEM of correlation matrices. The approach assumes the strict half-vector of the matrix logarithm of the correlation matrix is multivariate normal, an assumption that should be easily met for a variety of commonplace SEM problems. The approach is widely applicable as it provides an option for analyzing any correlation matrix for which a consistent estimate of the asymptotic covariance matrix is available. Moreover, the approach can be utilized with frequentist SEM estimation, though this is not necessary given other methods in the frequentist SEM toolkit.

One of the extensions explored in this paper is Bayesian SEMs for ordinal data on the basis of the polychoric correlation matrix. Bayesian SEMs for ordinal data methods can be time-consuming, especially with large datasets. With the log-correlation method, the polychoric correlation matrix is first computed. The larger the sample size, the more reliable the polychoric correlations and their asymptotic covariance matrix and consequently: the log-correlation method. And model complexity is then related to the size of the correlation matrix as opposed to the sample size of the data. Further study is required to understand the suitability of the log-correlation method

for routine analysis of polychoric correlations, especially given relatively common problems with two-stage estimation of SEMs for ordinal variables, such as non-positive-definiteness and unreliable asymptotic covariance matrices (e.g. Zhang, Trichtinger, Lee, & Jiang, 2022).

In addition to model estimation, the log-correlation method always represents a credible data generation process (DGP) for correlation matrices, with potential for use in simulation studies. Wishart-based DGPs can also be modified to produce valid correlation matrices, but such DGPs are restrictive. For example, consider Wishart-based hierarchical models for covariance matrices (Wu & Browne, 2015). This procedure with trivial modifications produces valid correlation matrices, but variation between correlation matrices is constrained by the Wishart distribution, implying a strong deterministic relation between the size of a correlation and its variance across different populations. In the log-correlation method, the variation of each correlation across populations is additionally influenced by a unique dispersion parameter, representing a more flexible process.

Based on the previous paragraph, the log-correlation method has potential for meta-analytic SEMs. Meta-analytic SEMs often take multiple correlation matrices as input data and attempt to identify the structured correlation matrix responsible for the observed data. That correlation rather than covariance matrices are often the input for such models makes the log-correlation method an appealing analytic approach. Moreover, one can estimate linear models on the transformed scale to accommodate predictor variables and complex data structures including dependent samples (e.g. Wilson, Polanin, & Lipsey, 2016). However, a common challenge with meta-analytic SEMs is the presence of missing correlations and variables across studies. The log-correlation method requires complete correlation matrices to compute $\boldsymbol{\gamma}$. This implies that missing correlations will require treatment (such as imputation) prior to the matrix logarithm operation. We intend to examine the application of the log-correlation method to meta-analytic SEMs in the future.

We also examined the ability of the log-correlation method to generate and estimate SEMs under misspecified models. For misspecification, we introduced a residual correlation matrix that varied randomly across replications. Alternatively, one can introduce random-error by generating the sample correlation matrix as: $\mathbf{R} = \gamma^{-1}(\mathbf{y}+\boldsymbol{\epsilon})$ where $\mathbf{y}$ is as usual, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \varepsilon)$. Since $\boldsymbol{\epsilon}$ varies randomly across repeated samples, this DGP also implies a misspecified SEM as the error cannot be predicted a-priori. Interestingly, on examining this process via Monte Carlo simulation, we found that the RMSEA from fitting a frequentist SEM to data generated under this process approximately

equals ε especially when ε is low. Moreover, a Bayesian model that matches this modified DGP often adequately recovers ε – results are in part B of the online supplementary materials. We intend to explore the connection between this misspecification model and the RMSEA in a future study.

Finally, the log-correlation method requires $\widehat{\boldsymbol{\Omega}}$, a consistent estimate of $\boldsymbol{\Omega}$. $\widehat{\boldsymbol{\Omega}}$ is unstable for smaller samples, or larger number of indicators. And these problems are worsened with polychoric correlations. In frequentist SEMs, $\widehat{\boldsymbol{\Omega}}$ is often considered unreliable such that only its diagonal is used for estimating point estimates, i.e. diagonal weighted least squares (Muthén, du Toit, & Spisic, 1997). In contrast, Bayesian methods typically supplement limited information in unreliable data with informative priors. To improve the performance of the log-correlation approach, it is necessary to identify optimal procedures for handling unreliable $\widehat{\boldsymbol{\Omega}}$. Potential options include: resampling-based estimates of $\widehat{\boldsymbol{\Omega}}/\widehat{\boldsymbol{\Omega}}_\gamma$ (e.g. Monroe, 2018), smoothing $\widehat{\boldsymbol{\Omega}}/\widehat{\boldsymbol{\Omega}}_\gamma$ (e.g. Yang & Yuan, 2019), computing $\widehat{\boldsymbol{\Omega}}/\widehat{\boldsymbol{\Omega}}_\gamma$ based on the model-implied correlations during model estimation, and the combination of multiple strategies. We intend to explore these strategies in a future study.

Disclosure Statement

There are no relevant financial or non-financial competing interests to report.

References

Archakov, I., & Hansen, P. R. (2021). A new parametrization of correlation matrices. *Econometrica*, *89*(4), 1699–1715. doi: 10.3982/ECTA16910

Archakov, I., Hansen, P. R., & Luo, Y. (2022, October). *A new method for generating random correlation matrices* (Tech. Rep. No. arXiv:2210.08147). arXiv. Retrieved 2023-08-23, from `http://arxiv.org/abs/2210.08147`

Bentler, P. M., & Chou, C. P. (1987, August). Practical issues in structural modeling. *Sociological methods and research*, *16*(1), 78–117. doi: 10.1177/0049124187016001004

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83. doi: 10.1111/j.2044-8317.1984.tb00789.x

Browne, M. W., & Shapiro, A. (1986, October). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra and its Applications*, *82*, 169–176. doi: 10.1016/0024-3795(86)90150-3

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

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. (2001, May). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, *29*(4), 468–508. Retrieved 2023-12-08, from `https://doi.org/10.1177/0049124101029004003` (Publisher: SAGE Publications Inc) doi: 10.1177/0049124101029004003

Chen, J., Guo, Z., Zhang, L., & Pan, J. (2021). A partially confirmatory approach to scale development with the Bayesian lasso. *Psychological Methods*, *26*, 210–235. (Place: US Publisher: American Psychological Association) doi: 10.1037/met0000293

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*(2), 317–327. doi: 10.1037/0033-2909.105.2.317

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psicológica*, *21*(2), 301–323. Retrieved from `https://www.uv.es/revispsi/articulos3.00/ferran7.pdf`

Flora, D. B., & Curran, P. J. (2004, December). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. doi: 10.1037/1082-989X.9.4.466

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago Press.

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, *19*(10), 3595–3617. doi: 10.1080/03610929008830400

Jöreskog, K. G. (1994, September). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*(3), 381–389. doi: 10.1007/BF02296131

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*(3), 502–511. doi: 10.1037/0033-2909.109.3.502

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. doi: 10.1037/1082-989X.4.1.84

Monroe, S. (2018, March). Contributions to estimation of polychoric correlations. *Multivariate Behavioral Research*, *53*(2), 247–266. doi: 10.1080/00273171.2017.1419851

Mulaik, S. A. (2009). *Foundations of factor analysis* (2nd ed.). New York: Chapman and Hall/CRC. Retrieved from `https://doi.org/10.1201/b15851` doi: 10.1201/b15851

Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible represen-

tation of substantive theory. *Psychological Methods*, *17*(3), 313–335. doi: 10.1037/a0026802

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997, November). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Retrieved from `https://www.statmodel.com/download/Article_075.pdf`

Neudecker, H., & Wesselman, A. M. (1990, January). The asymptotic variance matrix of the sample correlation matrix. *Linear Algebra and its Applications*, *127*, 589–599. doi: 10.1016/0024-3795(90)90363-H

Olkin, I., & Finn, J. D. (1995, July). Correlations redux. *Psychological Bulletin*, *118*(1), 155–164. doi: 10.1037/0033-2909.118.1.155

Revelle, W. (2023). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from `https://CRAN.R-project.org/package=psych` (R package version 2.3.3)

Robitzsch, A. (2023, June). Modeling model misspecification in structural equation models. *Stats*, *6*(2), 689–705. doi: 10.3390/stats6020044

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

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969, December). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421–459. Retrieved 2022-07-28, from `https://doi.org/10.1007/BF02290601` doi: 10.1007/BF02290601

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*. doi: 10.1177/00131644231165306

Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016, June). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods*, *7*(2), 121–139. doi: 10.1002/jrsm.1199

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

Yang, M., & Yuan, K.-H. (2019, January). Optimizing ridge generalized least squares for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(1), 24–38. doi: 10.1080/10705511.2018.1479853

Zhang, G., Trichtinger, L. A., Lee, D., & Jiang, G. (2022, March). PolychoricRM: A computationally efficient R function for estimating polychoric correlations and their asymptotic covariance matrix. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(2), 310–320. doi: 10.1080/10705511

.2021.1929996