Title

Comparing Estimation Methods for Psychometric Networks With Ordinal Data

Authors
Johal, Simran K
Rhemtulla, Mijke

Publication Date
2021-12-20

DOI
10.1037/met0000449

Abstract

Ordinal data are extremely common in psychological research, with variables often assessed using Likert-type scales that take on only a few values. At the same time, researchers are increasingly fitting network models to ordinal item-level data. Yet very little work has evaluated how network estimation techniques perform when data are ordinal. We use a Monte Carlo simulation to evaluate and compare the performance of three estimation methods applied to either Pearson or polychoric correlations: EBIC graphical lasso with regularized edge estimates ("EBIC"), BIC model selection with partial correlation edge estimates ("BIC"), and multiple regression with p-value-based edge selection and partial correlation edge estimates ("MR"). We vary the number and distribution of thresholds, distribution of the underlying continuous data, sample size, model size, and network density, and we evaluate results in terms of model structure (sensitivity and false positive rate) and edge weight bias.

Our results show that the effect of treating the data as ordinal versus continuous depends primarily on the number of levels in the data, and that estimation performance was affected by the sample size, the shape of the underlying distribution, and the symmetry of underlying thresholds. Furthermore, which estimation method is recommended depends on the research goals: MR methods tended to maximize sensitivity of edge detection, BIC approaches minimized false positives, and either one of these produced accurate edge weight estimates in sufficiently large samples. We identify some particularly difficult combinations of conditions for which no method produces stable results.

Keywords: network psychometrics; network models; ordinal data; nonnormal data

Translational Abstract

Psychometric network models are statistical models used to explore the associations among variables thought to mutually influence each other, and have been widely applied in psychopathology, politics, and intelligence. Although psychological researchers often collect data measured on an ordinal scale, meaning that variables can only take on a few values, little research has examined how network models perform when applied to such data, or which estimation technique researchers should use. In this paper, we conduct a simulation study evaluating and comparing three different estimation methods when applied to ordinal data, varying how the data was treated and characteristics of both the data and network model. The three approaches we compare are: EBIC graphical lasso ("EBIC"; a widely-used network estimation approach), BIC model selection ("BIC"; a popular nonregularized approach), and a multiple regression-based approach ("MR").

Our results indicate that if researchers are interested in estimating networks with ordinal data, then they should be aware that estimation performance of all three methods is affected by whether the data is treated as ordinal versus continuous, the sample size, the shape of the underlying continuous distribution, and the symmetry of the underlying thresholds. Furthermore, which estimation method we recommend depends on the research goal: MR methods maximize the ability to detect relations between variables, BIC methods minimize the detection of false positive relations, and either of these approaches accurately estimate the strength of the relations once the sample size is large enough.

Comparing Estimation Methods for Psychometric Networks with Ordinal Data

Psychological researchers are often interested in exploring associations among variables that are thought to mutually influence each other, and network models are an increasingly popular method to do so. Network models depict variables as *nodes* in a system, which are connected by *edges* that represent the associations among pairs of variables. Although initially proposed as an alternative to latent variable models for psychopathology (Borsboom & Cramer, 2013), network models have since been widely applied in a variety of subfields of psychology. For example, network models have been used to investigate the structure of intelligence (van der Maas et al., 2017), political attitudes (Dalege et al., 2017a, 2017b), and classroom dynamics (Abacioglu et al., 2019).

In most psychological network models, edges represent the unique relation between a pair of variables, controlling for other variables in the network. These edges are typically estimated using partial correlations, which are computed by inverting the matrix of Pearson correlations that represents the linear associations among variables. Yet many psychological variables are assessed using Likert-type scales and thus can only take on a few values (e.g., a participant responds on a 7-point scale ranging from strongly disagree to strongly agree). When data are ordinal, methodologists have recommended using polychoric correlations to estimate edge weights (Epskamp & Fried, 2018). Polychoric correlations, used when both variables are ordinal, assume that a latent continuous, normally distributed variable underlies each ordinal variable, and the relation between a continuous value and the observed ordinal response is determined by a set of thresholds. The polychoric correlation between two ordinal variables estimates the Pearson correlation that would be obtained if one had measured the underlying variables on their original continuous metric (Olsson, 1979).

Estimating Networks in Psychological Research

Regardless of the correlation type used, the first step in calculating the partial correlation matrix, P, is to invert the variance-covariance matrix of the data, Σ**,** to obtain the precision matrix, **K**:

$$\Sigma^{-1} = K = \begin{bmatrix} k_{ii} & ¿ \\ \vdots & ¿\,¿\,k_{ij}\,¿\cdots¿\,k_{jj}\,¿ \end{bmatrix}.$$

The off-diagonal elements of this precision matrix are then standardized and the sign reversed to obtain the partial correlation ρ_{ij}

$$\rho_{ij} = \frac{-k_{ij}}{\sqrt{k_{ii} k_{jj}}}, i \neq j.$$

Due to the sampling variability present in a finite sample, even two conditionally independent variables are likely to have an estimated partial correlation that is nonzero (although it may be small). One goal of network estimation is to limit the number of false positives by setting some partial correlations to zero, thereby imposing sparsity on the precision matrix (Epskamp & Fried, 2018). Sparse networks are preferred for their ease of interpretation: by interpreting absent (zero) edges as likely conditional independencies, researchers may more readily identify network features such as central nodes (i.e., nodes that have more and stronger connections).

Numerous methods are available for introducing sparsity into the precision matrix, but the most commonly-used method in psychology involves l_1-regularization, or lasso, which introduces a penalty parameter that shrinks edge estimates toward zero, and shrinks many to exactly zero. This method was developed for use in high-dimensional settings where the number of variables was greater than the sample size, as methods that did not involve regularization (such as forward search algorithms used in regression models) could not guarantee a solution and were oftentimes impractical in these settings (Meinshausen & Bühlmann, 2006; Tibshirani, 1996).

Although psychological research often takes place in low-dimensional settings and therefore does not have the same restriction against nonregularized approaches, the lasso approach is a convenient way to obtain a sparse network with an easily interpretable structure. A specific variant of the l_1-regularization approach – called EBIC graphical lasso (see Network Estimation Methods section for technical details) – has been the most popular estimation technique in psychological applications. Since ordinal data violate the assumption of multivariate normality required by the lasso approach, it is recommended that the EBIC graphical lasso method is applied to the matrix of polychoric correlations instead of Pearson correlations when data are ordinal (Epskamp & Fried, 2018; Meinshausen & Bühlmann, 2006).

Epskamp (2017) showed that the EBIC graphical lasso approach works well with continuous data. This study evaluated performance by examining (a) how well EBIC graphical lasso was able to correctly identify which edges were present or absent (in other words, how well it recovered the true structure of the network) and (b) how accurately it estimated the values of the edges (i.e., how close the estimated edge weights were to the true partial correlations). The method performed well in determining which edges were truly absent, and its ability to determine which edges were truly present increased as sample size increased. Furthermore, EBIC graphical lasso accurately estimated the edge weights, especially at larger sample sizes.

Despite the promising performance of EBIC graphical lasso, recent work has indicated that some nonregularized approaches to imposing sparsity often perform as well as or better than EBIC graphical lasso in the low-dimensional settings common to psychological research. Williams and Rast (2019) compared EBIC graphical lasso to a nonregularized method that identified present edges based on either the 95% or 99% confidence intervals of the partial correlations. In their simulation, the ability of EBIC graphical lasso to correctly identify the presence of absent edges decreased with increasing sample sizes (especially in denser networks), whereas the confidence interval method showed consistent estimation.

Similarly, Williams et al. (2019) compared EBIC graphical lasso to nonregularized regression methods that selected the neighborhood of each node (i.e., the subset of remaining nodes that it is conditionally dependent on) based on forward search algorithms that determine which nodes to include by using either Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). They found that, when data were continuous, the regression-based methods typically outperformed EBIC graphical lasso in accurately detecting the edges of the network, and their performance only improved with increasing sample size. Furthermore, although EBIC graphical lasso produced more accurate edge estimates in larger networks, it had similar or worse performance (in terms of both accurate edge estimates and accurate structure recovery) in larger sample sizes and smaller networks.

These findings have drawn attention to the fact the EBIC graphical lasso – although widely used – may not always be the most appropriate choice in psychological settings. In light of this, nonregularized techniques are becoming increasingly preferred, and have been implemented as options in the primary psychological network software package *qgraph* (Epskamp et al., 2012), as well as the default of other network software packages (e.g., *GGMnonreg,* Williams et al., 2019). As more psychological researchers begin to adopt these approaches, it is worth examining how nonregularized techniques – as well as EBIC graphical lasso – perform with ordinal data, due to its ubiquity in psychological research. This paper focuses on how well EBIC graphical lasso and two nonregularized techniques – one based on the

BIC, and the other based on multiple regression – can accurately recover the network structure and edge weights when data are ordinal.

Little work has previously explored how EBIC graphical lasso or nonregularized methods perform when dealing with ordinal data. In particular, little is known about how these methods perform when the polychoric correlation matrix is used in the estimation process, as is the current recommendation for ordinal data, as opposed to the Pearson correlation matrix, which treats the ordinal data as though they were continuous. Although Williams et al. (2019) did include 5-point ordinal data in their comparisons, only one of their methods (the graphical lasso) used polychoric correlations, because their alternative methods were based on linear regression which assumed linear relations in the raw data. Their comparisons mainly focused on the performance of the estimation methods when the data were continuous or when the data were on a 5-point ordinal scale, and not on how performance was affected by the use of polychoric or Pearson correlation matrices. Epskamp (2017) compared the performance of EBIC graphical lasso when using the polychoric correlation matrix versus the Pearson correlation matrix, applied to 5-point scale data, and found that using the polychoric correlation matrix often resulted in worse structure recovery at smaller sample sizes, due to estimating very dense or fully connected networks. Once sample size increased to at least 250, the performance of the two types of correlations was comparable for both structure recovery and the accuracy of the edge estimates.

However, even the comparisons in Epskamp (2017) were limited by focusing on ordinal data that were created by thresholding a normal distribution into 5 categories. Research examining ordinal estimation methods in structural equation modeling (SEM) has consistently found little difference between categorical and continuous estimation methods when there are at least 5 categories, especially when the underlying distribution is multivariate normal (Finney &

DiStefano, 2013; Rhemtulla et al., 2012). Using a categorical versus continuous estimation method tends to make the most difference when either the number of categories is less than 5, or the underlying distribution is nonnormal and/or the thresholds producing the observed ordinal distribution are highly asymmetric (Rhemtulla et al., 2012). The current study aims to evaluate and compare the performance of EBIC graphical lasso, a BIC nonregularized technique, and a multiple regression nonregularized technique when estimated on ordinal data with varying number of categories, underlying distributions, and thresholds, and using the polychoric versus the Pearson correlation matrix.

Network Estimation Methods

The goal of our simulation study was to examine the performance of the EBIC graphical lasso approach (EBIC) and two nonregularized approaches – one based on the Bayesian Information Criterion (BIC), and the other based on multiple regression (MR). A quick overview of these methods is presented in Table 1.

EBIC graphical lasso (EBIC)

This variant of the l_1-regularization method is a common network estimation technique in psychology. In contrast to the l_1-regularization approach described in Meinshausen and Bühlmann (2006), which penalizes the coefficients of nodewise regressions, the graphical lasso directly penalizes estimates of the elements of the precision matrix, by estimating the precision matrix, $\widehat{K}$, that maximizes the penalized likelihood function:

$$L(\widehat{K}) = \log(det(\widehat{K})) - tr(S\widehat{K}) - \lambda \sum |k_{ij}|,$$

where **S** is the sample covariance matrix and λ is a penalty parameter (Friedman et al., 2008). Penalizing the sum of absolute values of elements in the precision matrix shrinks these elements towards zero, and shrinks some to exactly zero. The edges that are estimated to be nonzero are biased estimates of the population partial correlations. Researchers commonly use the presence or absence of edges in the estimated network to make informal inferences about the presence or absence of edges in the corresponding population network (Williams et al., 2020).

The value of λ in the penalized likelihood affects the balance of false positives (i.e., spurious edges) and false negatives (i.e., absent true edges): higher values of λ will result in greater penalization, and thus more edge weights are estimated to have a value of zero. To balance the two, networks are estimated for a sequence of 100 λvalues, and the value of λ is chosen to minimize the extended Bayesian Information Criterion (EBIC) of the network, whose formula is given by:

$$EBIC = -2\left(L\left(\widehat{K}\right)\right) + E\log(n) + 4\gamma E\log(p),$$

where n represents the sample size, E the number of nonzero edges in the estimated network, and p the number of variables in the network. γ is a prespecified hyperparameter that controls the preference for simpler models, and is typically set to 0.5 for a more conservative approach (Epskamp & Fried, 2018; Foygel & Drton, 2010).

The continuous version of this method, which we refer to as cont-EBIC, uses the sample Pearson correlation matrix of the data as input. The categorical version, referred to as cat-EBIC, uses the sample polychoric correlation matrix.

Nonregularized Bayesian Information Criterion (BIC)

The nonregularized BIC approach, proposed by Epskamp (2018), is an iterative network structure estimation method, similar to the approach of Williams et al. (2019). It begins with the graphical lasso approach described in the previous section, but sets the EBIC hyperparameter γ

to 0, so that the EBIC formula reduces to the BIC. This was done in an effort to improve the ability of the graphical lasso to accurately detect edges that are present in the network, while retaining its ability to avoid false positives (Epskamp, 2018). The BIC-optimal network structure is then refit *without regularization*, meaning that the nonzero edge values are re-estimated without shrinkage, producing unbiased estimates of nonzero paths while maintaining the conditional independence pattern of the graphical lasso-derived optimal network. That is, each nonzero edge weight is computed as a partial correlation between a pair of variables, conditioning on all other variables in the network. Edges are then iteratively added or removed from this refit network, and the network re-estimated, until the BIC can no longer be improved. A diagrammatic representation of this method is shown in Figure 1.

Like the EBIC graphical lasso approach, the continuous version of this method, which we call cont-BIC, uses the sample Pearson correlation matrix as input. The categorical version, which we call cat-BIC, uses the polychoric correlation matrix.

Nonregularized multiple regression (MR)

The nonregularized multiple regression approach begins by estimating p multiple regressions, each of which predicts one variable in the network from the set of all others. A weighted adjacency matrix is constructed based on the results of these regressions: if the ith variable is a significant predictor of the jth variable at $\alpha = .05$ and vice versa, then element $\hat{w}_{ij}$ of the weighted adjacency matrix is set to be the nonzero partial correlation between those two variables, and is set to 0 otherwise. The AND-rule used here (as opposed to the OR-rule, where an edge is nonzero if at least one of the two regression coefficients is significant) has been used in other regression-based approaches to network estimation (e.g., van Borkulo et al., 2014). Note that while many nodewise regression methods, such as those used in Meinshausen and Bühlmann

(2006) and Williams et al. (2019), estimate the edge weights conditional only on the variables that were significant predictors, the edge weights in the weighted adjacency matrix are conditioned on *all* other variables in the network, as done in Williams and Rast (2019).

The continuous version of this approach, which we call cont-MR, uses linear nodewise regressions to estimate the structure, and the nonzero edge weights are Pearson partial correlations. The categorical version of this approach, cat-MR, estimates nodewise regression coefficients using the categorical least squares estimator that is commonly employed for estimating structural equation models with categorical data. This method estimates regression coefficients by applying unweighted least squares estimation to the polychoric correlation matrix, and applies a robust correction to the standard errors based on the asymptotic covariance matrix of the polychoric correlations (Muthén et al., 1997). The nonzero values in the weighted adjacency matrix are polychoric partial correlations.

Method

Conditions

To compare the performance of the three methods under multiple conditions – particularly conditions that represent some assumption violation – we varied characteristics relating to properties of the network and of the ordinal data.

Network Size

Previous research looking at published network analyses on psychological variables found that most of these networks contained between 10 and 30 variables (Wysocki & Rhemtulla, 2019). Therefore, to ensure that the network models examined here were similar to those used in psychological research, we simulated networks with either p = 10 or p = 20 nodes.

It is important to consider different network sizes in this simulation because including more or fewer variables can lead to vastly different results; for example, Williams et al. (2019) found that EBIC graphical lasso estimated edge weights more accurately than nonregularized methods for 20-node networks, but this pattern reversed for 10-node networks.

Network Connectivity

In order to produce consistent neighborhood selection, the graphical lasso assumes that the true network is sparse, with few connections between nodes (Meinshausen & Bühlmann, 2006). When the generating network meets this assumption of sparsity (e.g., when between 20 and 40 percent of possible edges are present), EBIC graphical lasso has been shown to adequately recover the network structure (Epskamp & Fried, 2018; van Borkulo et al., 2014). However, as the proportion of present edges (the connectivity) increases above 40%, the performance of EBIC graphical lasso decreases, such that the false positive rate increases with increasing n (Williams et al., 2019; Williams & Rast, 2019). Because most networks estimated in psychological settings are expected to be dense, with connectivity greater than 50% (Wysocki & Rhemtulla, 2019), it is important to consider the performance of these estimation methods under different (and increasing) levels of connectivity. To this end, we varied the level of connectivity in networks, such that connectivity was either high (70% of possible edges were present), medium (50% of possible edges were present), or low (30% of possible edges were present). The values of high and medium connectivity (70% and 50%) were within the range of larger connectivity values assessed in previous simulation studies, and represent a violation of the sparsity assumption (Williams & Rast, 2019; Williams et al., 2019; Wysocki & Rhemtulla, 2019). The value for low connectivity (30%) was chosen to represent a sparse network that is in line with the assumptions of lasso, and falls within the range of connectivity values where EBIC graphical lasso has shown adequate performance.

Sample Size

The performance of each estimation technique is also expected to vary as sample size increases. For example, if the underlying network is dense, EBIC graphical lasso tends to produce more false positives as the sample size increases, whereas the nonregularized methods assessed in both Williams and Rast (2019) and Williams et al. (2019) were either unaffected or improved with larger samples. Furthermore, when dealing with ordinal data, lower sample sizes can lead to densely connected networks that are hard to interpret (Epskamp, 2017).

To evaluate the performance of the three estimation techniques, we compared across sample sizes of 100, 250, 500, 1,000, and 3,000. This comparison allows us to evaluate the consistency of each method as sample size increases. These sample sizes were chosen to be representative of sample sizes in psychological research, and to allow us to replicate low-dimensional settings where sample sizes tend to be (much) larger than the number of variables.

Underlying Distribution

In order to simulate ordinal data, continuous data were simulated from either a multivariate normal or multivariate non-normal distribution, then discretized into the appropriate number of categories. We varied the underlying distribution because both EBIC graphical lasso and polychoric correlations assume underlying normality (Meinshausen & Bühlmann, 2006; Olsson, 1979) – an assumption that is unlikely to hold in psychology, where the majority of variables are non-normally distributed (Cain et al., 2016; Micceri, 1989). The violation of the normality assumption introduces possible bias into the estimates of the polychoric correlations.

In the SEM literature, multivariate non-normal data are oftentimes generated using the Vale-Maurelli method (Vale & Maurelli, 1983; e.g., Curran et al., 1996; Flora & Curran, 2004; Rhemtulla et al., 2012). The Vale-Maurelli method uses Fleishman polynomials to transform data generated from a multivariate normal distribution so that the underlying population distribution becomes multivariate non-normal with the desired covariance matrix, skewness, and kurtosis (Fleishman, 1978; Vale & Maurelli, 1983). These continuous data are then discretized to produce ordinal data. Using this method, previous literature has shown the relatively modest impact of underlying non-normality on SEM estimates based on polychoric correlations (Flora & Curran, 2004; Li, 2016; Rhemtulla et al., 2012). Thus, in line with previous work examining the impact of underlying non-normality on ordinal data, we used the Vale-Maurelli method to simulate non-normal data with a univariate skewness of 2 and univariate kurtosis of 7 to represent "moderate" non-normality (Curran et al., 1996; Olvera Astivia, 2020).

Recent work has criticized the approach of discretizing data generated from the Vale-Maurelli method to achieve ordinal data, as well as the Vale-Maurelli method in general. In order for a multivariate distribution to be non-normal, either the marginal distributions must be non-normal or the copula of the multivariate distribution must be non-normal (Foldnes & Olsson, 2016). The copula of data generated using the Vale-Maurelli method is similar to the multivariate normal copula, and thus the data may not be as non-normally distributed as intended (Olvera Astivia, 2020). Furthermore, work by Grønneberg and Foldnes (2019) and Foldnes and Grønneberg (2020) has demonstrated that discretizing data produced by the Vale-Maurelli method is numerically equivalent to discretizing the intermediate data generated by the multivariate normal distribution, only with different thresholds. In other words, the observed ordinal data produced by discretizing data with a multivariate non-normal distribution could also have been produced by discretizing data with a multivariate normal distribution. This calls into question whether the estimation results are actually based on data with an underlying non-normal distribution.

To address this limitation, alternative simulation methods have been proposed, based on either vines and copulas, or on independent generator variables (Foldnes & Olsson, 2016; Grønneberg & Foldnes, 2017). To compare the effect that different types of underlying non-normality have on estimation performance, we also simulated non-normal data with univariate skewness of 2 and univariate kurtosis of 7, but based on the independent generator method. The independent generator method finds a matrix of coefficients whose product with its transpose equals the desired covariance matrix, as well as the skewness and kurtosis of the independent generator variables that, when transformed, will result in the desired skewness and kurtosis. Data are simulated from the independent generator variables, and then transformed using the matrix of coefficients to obtain a random non-normal sample (Foldnes & Olsson, 2016).

Number of Categories

The simulated continuous data were discretized into 3, 4, 5, and 7 categories. Although previous work has examined the performance of the EBIC graphical lasso method with 5 categories, we expanded the number of categories under consideration to determine when ordinal data can be treated as continuous (i.e., supplying the Pearson correlation matrix as input) and produce similar results. Comparisons of continuous and ordinal estimation methods for structural equation models have typically found that, when there are five or more categories, continuous estimation methods perform as well as or better than ordinal methods (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Rhemtulla et al., 2012). Thus, since anywhere from 3 to 7 categories are common in Likert-type items, we considered that range here. We did not consider binary variables, as such networks are typically estimated as Ising networks using the regularized logistic regression approach proposed by van Borkulo et al. (2014).

Threshold Symmetry

Previous research in SEM investigated the impact of category thresholds being symmetrically or asymmetrically distributed around the mean, and found that asymmetric thresholds can impact, for example, the accuracy of parameter estimates, test statistics, and power (Lei, 2009; Rhemtulla et al., 2012). In the case of network estimation, asymmetric thresholds can lead to biased estimates of the polychoric correlations, which can then compound to even greater bias in the partial correlation estimates (Epskamp & Fried, 2018).

We considered three types of threshold symmetry: symmetric thresholds that evenly divided the space between -2.5 and 2.5 standard deviations around a mean of 0, moderately asymmetric thresholds where the majority of responses fell to the left of center, and extremely asymmetric thresholds where the first category had the greatest proportion of cases (and then the proportion continued to decrease for each of the remaining categories). The exact threshold values are the same as those used in Rhemtulla et al. (2012), and are presented in Table 2.

Simulation Procedure

In summary, the factors we varied in this study were:

- <u>Network size</u>: 10 nodes, 20 nodes
- <u>Level of network connectivity</u>: High (70%), medium (50%), low (30%)
- <u>Sample size</u>: 100, 250, 500, 1,000, 3,000
- <u>Underlying normality</u>: Normal, non-normal (Vale-Maurelli), non-normal (independent generators)

- Number of categories: 3, 4, 5, 7
- Threshold symmetry: Symmetric, moderately asymmetric, extremely asymmetric

These factors were fully crossed, resulting in a total of 1,080 conditions. For each condition, we simulated 1,000 datasets as follows:

1. A bank of polychoric partial correlations was created using the *bfi* dataset available in the R package *psych*, which assessed participants' personality traits using a 6-point response scale (Revelle, 2019). Following the procedure of Epskamp (2017), all partial correlations with absolute value less than .05 were removed from this bank.
2. The appropriate number of partial correlations were sampled from this bank to be the nonzero edges of the network, where the appropriate number was the integer closest to

$$\frac{p(p-1)}{2} * connectivity.$$

If the resulting partial correlation matrix was not positive definite, then the sampling procedure was repeated until a positive definite matrix was obtained.

3. For each population network, an ordinal dataset was simulated that met the prespecified conditions formed by crossing the factors described above. We then calculated the polychoric and Pearson correlation matrices to use as input for the network estimation techniques – if this correlation matrix was not positive definite, the nearest positive definite matrix was found using the algorithm of Higham (2002).
4. Networks were estimated using the 6 network estimation techniques: cont-EBIC, cat-EBIC, cont-BIC, cat-BIC, cont-MR, and cat-MR.

Network Estimation Software

Continuous data were simulated using the *mvrnorm* function of the *MASS* package (Venables & Ripley, 2002), *mvrnonnorm* function of the *semTools* package (Jorgensen et al., 2020), and *rIG* function of the *covsim* package (Grønnenberg & Foldnes, 2017) for underlying normal, non-normal (Vale-Maurelli), and non-normal (independent generators) distributions, respectively. Polychoric partial correlation matrices were estimated using the *cor_auto* function in the *qgraph* package, and if necessary, the nearest positive-definite matrix was found using the *nearPD* function of the *Matrix* package (Bates & Maechler, 2019; Epskamp et al., 2012). Network estimation with the EBIC and BIC methods were implemented using the *EBICglasso* and *ggmModSelect* functions, respectively, in the *qgraph* package. Finally, the regressions of the MR approach were run in *lavaan*, and if the data were being treated as ordinal, the outcome variable was declared to be ordered and the estimator was specified as "ULSMV" (Rosseel, 2012)

The code for the entire simulation, and figures demonstrating full results, can be found at https://osf.io/t4gb9.

Performance Measures

In order to evaluate the performance of the methods, we compared how well each method was able to accurately recover the network structure – that is, the pattern of present and absent edges – and to accurately estimate edge strength. We also compared the model convergence rates of each estimator to assess the ability of each method to find a solution in difficult conditions.

Convergence

We assessed the convergence rates for each method across all conditions. Each estimation method could fail to converge for different reasons – for the MR method, a replication could fail to converge if at least one of the *p* regressions did not produce a solution, whereas the BIC method could fail to converge if a solution was not found within a given time limit (due to convergence issues within the *glasso* function that is used within *ggmModSelect,* Friedman et al., 2019; e.g., Mazumder & Hastie, 2012).

In addition to convergence rates, we noted whether the correlation matrices were non-positive definite. One noted issue with polychoric correlation matrices is that their pairwise construction means the final matrix may not be positive definite, an issue which Pearson correlation matrices avoid except in instances when variables are highly redundant or invariant (Lorenzo-Seva & Ferrando, 2020). If the correlation matrix is not positive definite, then it cannot be inverted to find the precision matrix. To avoid this problem, the nearest positive definite matrix was found instead, but the partial correlation matrix that results from this procedure may not be stable, and could lead to a densely connected network with edge weights near -1 and 1 (Epskamp & Fried, 2018). This instability in the partial correlation matrix can impact the estimates of our chosen performance measures – in particular, densely connected networks can affect estimates of sensitivity and false positive rate, while large edge weights can affect the estimates of bias. Therefore, the results reported below exclude replications where the original correlation matrix was not positive definite, and we note where this exclusion noticeably affected the outcomes.

Sensitivity

Sensitivity, also known as the true positive rate, quantifies how well a method correctly detects edges present in the network. It is calculated as

$$\text{Sensitivity} = \frac{\mathit{True\ Positives}}{\mathit{True\ Positives} + \mathit{False\ Negatives}},$$

where a true positive is an edge that is present in the population and estimated to be present by the network estimation technique, and a false negative is an edge that is present in the population but estimated to be absent. Sensitivity ranges between 0 and 1, with a score of 1 indicating that all edges that were present in the population network were estimated as such in the sample network. Thus, an estimated network can achieve perfect sensitivity if it is fully connected.

False Positive Rate

The false positive rate quantifies how often each method mistakenly identifies an edge that is absent in the population as present in the estimated network. It is calculated as

$$\text{False Positive Rate} = 1 - \frac{\mathit{True\ Negatives}}{\mathit{True\ Negatives} + \mathit{False\ Positives}},$$

where a true negative is an edge that is absent in the population network and estimated to be absent, and a false positive is an edge that is absent in the population but estimated to be present. The false positive rate ranges between 0 and 1, where a score of 0 indicates that all edges that were absent in the true network were absent in the sample network (i.e., no false positives were made). Just as a fully connected estimated network will have perfect sensitivity, an estimated empty network will have a false positive rate of 0.

Bias

In order to assess how well each method estimated the edge weights, we calculated the relative bias within each condition as:

$$RB_{\hat{w}} = \frac{1}{nREP} \sum_{j=1}^{nREP} \frac{1}{nTPE_j} \sum_{i=1}^{nTPE} \frac{\hat{w}_{ij} - w_{ij}}{w_{ij}},$$

where *nREP* is the number of converged replications, *nTPE* is the number of true positive edges within a replication, $\widehat{w}$ is an estimated edge weight, and w is the true edge weight value. In order to avoid conflating edge detection with edge accuracy, we only calculated relative bias for edges that were correctly estimated to be present.

Results

Due to the overwhelming number of conditions, we describe here a selection of the most informative results. First, we only present the results for 10-node networks – typically, having 20 nodes instead of 10 led to the selection of denser networks, particularly for the EBIC methods. This resulted in higher sensitivity, but also a higher false positive rate. Next, we exclude some levels of connectivity. Connectivity mainly affected the EBIC methods – the higher the connectivity of the network, the greater the sensitivity and the false positive rate of these methods. The effect of connectivity was especially pronounced in the 20-node networks – not only was the discrepancy between the levels of connectivity greater, but now all six estimation methods were impacted (again, such that higher connectivity led to greater sensitivity and more false positives). Therefore, we chose to only present results for those networks where 50% of possible edges were nonzero, as this maintained the relations between the different estimation approaches without either overstating or understating the performance of the EBIC methods.

We also omitted the Vale-Maurelli and moderately asymmetric thresholds conditions from the results. Because the independent generators approach to nonnormality produced more dramatically different results, we present only those here. Similarly, the moderately asymmetric thresholds showed very similar performance to the conditions with symmetric thresholds, whereas the extremely asymmetric thresholds produced notably different results.

Finally, the results only include conditions with 3 or 7 levels, as the focus on the number of levels was mostly on how differences between treating the data as ordinal or continuous changed as the number of levels increased. Since these comparisons can be made by looking at the two extremes of the number of levels, we omitted 4 and 5 level conditions (although we make reference to them in the text as relevant).

Convergence Failures and Non-Positive-Definite Matrices

Convergence failures are reported in Table 3. Convergence failures were limited to the two nonregularized approaches, whereas the EBIC method converged in all replications. The overall number of convergence failures was much higher for the BIC method than the MR method; in fact, for the BIC method, 3 conditions failed to converge in any of the 1,000 iterations (these 3 conditions were all 20-node networks with 3-category data and n = 100, generated using the independent generator method).

For both the BIC method and the MR method, convergence failures were more common in 20-node networks than 10-node networks, as well as when the sample size was smaller, when thresholds were extremely asymmetric, or when the data were generated by the independent generator method. For the BIC method, convergence failures were more common when the data were treated as ordinal rather than continuous, and with fewer levels. The reverse was true for the MR method – convergence failures were more likely when data were treated as continuous, and with more levels.

With regards to the correlation matrix not being positive definite, this issue was (as expected) more common when treating the data as ordinal than as continuous. At most 10 out of 1,000 Pearson correlation matrices were non-positive definite, whereas in some conditions all

1,000 polychoric correlation matrices were non-positive definite. The proportion of iterations with non-positive definite matrices increased as the number of nodes in the network increased, the connectivity increased, the thresholds were extremely asymmetric, or the data were generated using the independent generator method. Conversely, the proportion of iterations with this issue decreased with increasing sample size and increasing number of levels.

Sensitivity

The sensitivity results are displayed in Figures 2 and 3, with Figure 2 containing conditions with an underlying normal distribution, and Figure 3 an underlying non-normal distribution. Sensitivity for all methods tended to be low at small sample sizes (e.g., at $n = 100$ or $n = 250$, sensitivity generally ranged from below 5% to around 40%); however, the EBIC methods tended to perform the worst due to selecting empty or near-empty networks. Similarly, cat-MR selected empty or near-empty networks when the underlying distribution was non-normal and thresholds extremely asymmetric, resulting in the poor performance shown in the second column of Figure 3. Although sensitivity did increase as sample size increased, most methods barely achieved 70% sensitivity until $n = 1{,}000$, and it was not until $n = 3{,}000$ that all methods achieved above 70% sensitivity. Once sample size reached at least 3,000, all methods performed similarly well. One exception was cont-BIC and cont-EBIC in conditions that combined non-normal underlying distributions with extremely asymmetric thresholds – in such conditions, these methods tended to perform worse than other methods and did not reach 70% sensitivity even at $n = 3{,}000$. We also ran a single iteration of $n = 1{,}000{,}000$ to assess the asymptotic behavior of these methods, and found that all methods asymptotically achieved perfect sensitivity. The one exception was cat-MR in a condition that combined 3 levels, extremely asymmetric thresholds, and a non-normal underlying distribution, which selected an empty network and therefore had zero sensitivity.

It appears that the BIC and EBIC methods were most affected by treating the data as ordinal instead of continuous, with treating the data as ordinal resulting in higher sensitivity. This discrepancy is especially pronounced when the thresholds are extremely asymmetric and/or the underlying distribution is non-normal. As the number of levels increases, the discrepancy decreases, such that at 7 levels, treating the data as continuous gives similar results to treating the data as ordinal as long as the underlying distribution is normal. If the underlying distribution is non-normal, then even with 7 levels, treating the data as ordinal still results in higher sensitivity. The effect of the number of levels on reducing the gap between ordinal and continuous methods is made clearer in Figure 4, where sensitivity is plotted as a function of the number of levels for the condition with an underlying normal distribution, symmetric thresholds, and a sample size of 1,000 (although the pattern shown generally holds in other conditions as well). From Figure 4, it becomes clear that the performance of cat-BIC and cat-EBIC tends to be unaffected by the number of levels, and the increasing similarity in performance is due to the improved performance of cont-BIC and cont-EBIC as the number of levels increases.

Although the MR method is affected by treating the data as ordinal instead of continuous, this difference is not as extreme as that observed for the BIC or EBIC methods, and the difference goes in the opposite direction (i.e., cat-MR tends to show *lower* sensitivity than cont-MR). Furthermore, the difference does not change with the increasing number of levels – as can be seen in Figure 4, the lines representing cat-MR and cont-MR remain roughly parallel as the number of levels increases. Instead, the difference between cat-MR and cont-MR relies most on sample size – as the sample size increases (i.e., going down each column in Figures 2 and 3), the discrepancy decreases, and the two often give similar performance once the sample size is at least 500.

For all methods, there is also a general effect of threshold symmetry, such that performance worsens when the thresholds are extremely asymmetric. This effect is more pronounced when the underlying distribution is non-normal, but does not seem to affect any estimation method or data type more often than the rest. Similarly, performance tends to worsen when the underlying distribution is non-normal versus normal. These comparisons also reveal a potential problem with cat-BIC and cat-EBIC – these two methods have an unusually high sensitivity when the underlying distribution is non-normal or the thresholds are extremely asymmetric (or both), which is particularly obvious when the number of levels is low (e.g., the left two columns of Figure 3). This is likely due to selecting dense networks in these conditions, and indicates that these methods may potentially be unstable.

Overall, researchers interested in recovering as many present edges as possible (maximizing sensitivity) should be aware that if their data have few response options and the underlying distribution is non-normal with extremely asymmetric thresholds, then a sample size greater than 3,000 would be needed to produce reliable results. In general, researchers who wish to maximize sensitivity would be recommended to use either cat-MR or cont-MR (with cont-MR being preferred for sample sizes below 500), as these methods tended to result in similar or better sensitivity than other methods, but were stable across distribution type and threshold symmetry (unlike cat-BIC and cat-EBIC, which had high sensitivity but also a tendency to select dense networks).

False Positive Rate

The results for false positive rate are shown in Figures 5 and 6, with Figure 5 pertaining to conditions with an underlying normal distribution and Figure 6 pertaining to conditions from an underlying non-normal distribution. Overall, the false positive rates for all methods tended to be acceptable, with the average false positive rate being less than 30% in most conditions. Although the false positive rate is not as strongly affected by sample size as sensitivity is, some general trends emerged. The false positive rate of the BIC methods declined with increasing sample size, often nearing 0 by n = 3,000 (particularly cont-BIC). Conversely, the false positive rate of the EBIC methods *increased* with increasing sample size (often from almost 0 at n = 100 to around 20% or 30% at n = 3,000), which points to a problem with this method that has been noted before (Williams et al., 2019) – it is desirable that an estimation method produces fewer false positives as information increases, and the EBIC method fails to do so. The MR method also showed a very slight increase with increasing sample size, but this increase was, on average, only 5% for cat-MR and 2% for cont-MR across all conditions (compared to an average increase of 28% and 19% for cat-EBIC and cont-EBIC, respectively).

The effect of treating the data as ordinal instead of continuous generally mirrored the effects on sensitivity – just as treating the data as ordinal resulted in higher sensitivity for the BIC and EBIC methods, it also resulted in a higher false positive rate. Once again, this discrepancy was more obvious in the conditions with underlying non-normal distributions and/or extremely asymmetric thresholds, and decreased with an increasing number of categories. The discrepancy still remained even at 7 levels if the underlying distribution was non-normal or the thresholds were extremely asymmetric. However, Figure 7 (which displays the effect of the number of levels on the false positive rate) shows that the reduced gap between the ordinal and continuous methods is due to the improved performance (decreasing false positive rate) of cat-BIC, and some combination of improved performance of cat-EBIC and worsening performance of cont-EBIC. This trend differs from the sensitivity results, where the decrease in the discrepancy was almost solely due to the improved performance of cont-BIC and cont-EBIC. For the MR method, cat-MR tended to have a lower false positive rate (especially at sample sizes below 500, where it often selected near-empty networks) or a similar false positive rate as cont-MR. Once again, the discrepancy between cat-MR and cont-MR was unaffected by the number of levels, as the two lines for these methods are parallel in Figure 7, but the discrepancy did decrease with increasing sample size, and the two performed similarly once sample size was around 500.

Surprisingly, there was very little effect of threshold symmetry and underlying distribution type on false positive rate. The main difference between the underlying distribution being non-normal instead of normal, or the thresholds being extremely asymmetric instead of symmetric, was that cat-BIC and cat-EBIC tended to have much poorer performance in these conditions (due to selecting dense networks), particularly when the number of levels was low, as can be seen in column 2 of Figure 5, or columns 1 and 2 in Figure 6. Once again, this points to instability in these two methods.

Overall, many of the false positive results mirrored those of sensitivity, but the recommendations are slightly different. Although the MR methods were previously recommended due to their high sensitivity and stability across sample sizes, the results here suggest that the cont-BIC method does best at minimizing the false positive rate, even achieving close to no false positives in some conditions.

Relative Bias

Finally, when examining the relative bias for true positives, we considered an edge weight to be substantially biased if the absolute value of the relative bias was greater than 0.1, in line with previous literature (e.g., Flora & Curran, 2004). These results are shown in Figures 8 and 9. All methods tended to produce substantially biased results (with relative bias ranging from -0.8 to 3) at sample sizes less than 1,000. However, going down the columns of Figures 8 and 9 demonstrate that the relative bias of the nonregularized methods decreases as the sample size increases (from around 1 at n = 100 to close to or below 0.1 at n = 3,000), whereas the relative bias of the EBIC method remains relatively constant, or has slight decreases (e.g., the average change in bias from n = 100 to n = 3,000 was only around 0.01) Furthermore, the MR and BIC methods tended to produce edge estimates that overestimated the true edge weight, while the EBIC methods tended to produce edge estimates that underestimated the true edge weight.

Using the polychoric correlation matrix as input resulted in more bias for both nonregularized methods than using the Pearson correlation matrix, while the reverse was true for the EBIC method. Like the sensitivity and false positive results, this discrepancy was more prominent in conditions that had non-normal underlying distributions or extremely asymmetric thresholds, and decreased as the number of levels increased due to the decreasing bias of cat-MR, cat-BIC, and cont-EBIC.

The relative bias tended to be higher when the underlying distribution was non-normal, especially when the thresholds were symmetric, and methods that treated the data as ordinal were more affected than those that treated the data as continuous. However, if the thresholds were extremely asymmetric, there tended to be much less discrepancy between the relative bias of the two different underlying distributions. On the other hand, the effect of threshold symmetry was only present when the underlying distribution was non-normal; in this case, methods that treated the data as ordinal were once again more affected by threshold symmetry, but there was no consistency in whether conditions with extremely asymmetric thresholds resulted in worse bias. Otherwise, if the underlying distribution was normal, the relative bias in conditions with symmetric thresholds was similar to that in conditions with extremely asymmetric thresholds.

Comparing the three methods, the top two rows of Figures 8 and 9 show that at sample sizes less than 500, the EBIC methods produced less biased estimates than either nonregularized method. However, due to the decrease in bias with increasing sample size, the nonregularized methods produced essentially unbiased estimates by n = 3,000, while the EBIC methods were still substantially biased.

The asymptotic results shown in Figure 10 highlight that under the "ideal" scenario (underlying normal distribution, symmetric thresholds, and at least 5 levels), all methods produce essentially unbiased edge weight estimates. When the underlying distribution was normal but the thresholds were extremely asymmetric, as in the 2nd column, estimates from all methods that treated the data as continuous were biased while those from methods that treated the data as ordinal remained unbiased. If the underlying distribution was non-normal, then estimates from all methods were biased, but methods that treated the data as ordinal produced less bias than their continuous counterparts. Interestingly, all methods that used the polychoric correlation matrix produced the same amount of bias as each other, and this bias was less than the bias produced by the methods based on Pearson correlations. Therefore, asymptotically, treating the data as ordinal versus continuous mattered more than the particular estimation method used. This is not unexpected – the polychoric correlation is an estimate of the population correlation between the continuous variables that underlie the ordinal responses, whereas the Pearson correlation is an estimate of the correlation between the observed ordinal responses, and these two population quantities are not the same. Pearson correlations underestimate the correlations among the underlying continuous variables.

In sum, researchers should be aware that all methods will produce substantially biased estimates when the number of categories is small and the underlying distribution is non-normal, especially if the thresholds are extremely asymmetric. These results suggest that, to maximize the accuracy of the estimates, cat-EBIC should be used when estimating networks in samples smaller than 500, as this method produced less biased results than either of the nonregularized methods or cont-EBIC. If the sample size is at least 500, either of the nonregularized methods produce nearly unbiased results, although cat-MR and cat-BIC are preferred in case the underlying thresholds are asymmetric.

Follow-Up Simulation: Random Partial Correlation Matrices

We chose to generate our population networks from empirically-derived partial correlations, as we believed this would be a better representation of psychological networks, and thus our results and recommendations would be more applicable to researchers. Yet to ensure that none of our results were specific to only the population networks we could generate with our empirical dataset, we ran a smaller follow-up simulation for conditions that combined 10 nodes and 50% connectivity with either normal or non-normal (independent generators) underlying distributions, symmetric or extremely asymmetric thresholds, and 3 or 7 levels. Instead of building the population network from the partial correlations of an empirical dataset, the population network was created by transforming a randomly-generated precision matrix into a partial correlation matrix. To generate random precision matrices, we used the *bdgraph.sim* function available in the *BDgraph* package, setting the degrees of freedom for the G-Wishart distribution to the default value of 3 (Mohammadi & Wit, 2019). The degrees of freedom of the G-Wishart distribution control the size of the partial correlation values, with higher degrees of freedom corresponding to smaller partial correlation values (Hsu et al., 2012). A degrees of freedom value of 3 corresponds to a non-informative prior for the precision matrix; therefore, the partial correlation values in the generated population networks could range from -1 to 1 with no restriction, and 90% of the values fell between ±.4. (see Figure 11).

Figures displaying the results for sensitivity, false positive rate, and relative bias are available online. One immediately noticeable difference from the main results presented earlier is that denser networks are being estimated, resulting in both a higher sensitivity and a higher false positive rate. For example, at n = 100, the sensitivity of the methods in the main results were, on average, below 30% while the false positive rate was, on average, below 20%. However, in this follow-up simulation, the average sensitivity at n = 100 was above 60%, while the average false positive rate was above 40%. The higher sensitivity might be partially explained by the wider range of partial correlation values in the population networks - as illustrated in Figure 11, the possible partial correlation values for the empirical dataset ranged only from -0.3 to 0.6, whereas the range for partial correlations generated via *BDgraph* was -1 to 1. In fact, when we recalculated the ability of the estimation methods to accurately detect edges that were within and outside the range of -0.3 to 0.6, that is, the range of edges generated in the main, empirical data simulation, all methods were able to near-perfectly detect edges that were outside this range (unless the thresholds were extremely asymmetric, in which case performance was slightly reduced). The ability of the methods to detect edges within this range was slightly reduced (although still better than that of the main results).

Apart from increased densities leading to a higher sensitivity and false positive rate, the general pattern of results for these two performance measures remained the same. For sensitivity, the improved overall performance meant that almost all methods achieved at least 70% sensitivity by n = 500, instead of n = 3,000, and the false positive rate of the EBIC and cat-BIC methods were much higher than previously reported. However, the effect of treating the data as ordinal versus continuous, the overall effects of the different data characteristics studied, and the recommended methods remained the same.

To compute relative bias, we not only restricted our calculations to edges that were present in both the true and estimated network, but also to edges where the true partial correlation had an absolute value greater than .05. We did this because there were a number of very small, nonzero partial correlations present in the population networks that caused the relative bias to explode (i.e., because calculating relative bias requires dividing by the population value), making the averaged results uninterpretable. When we calculated relative bias with these restrictions, we found that although the relative bias was, in general, smaller than in the initial results, the general patterns tended to hold. The biggest differences were that almost all methods were essentially unbiased by n = 250 or n = 500, instead of n = 1,000, and the recommended method changed slightly. Now, for sample sizes below 1,000, either cont-MR or cont-BIC are recommended, instead of cat-EBIC, while for sample sizes of 1,000 or more, cat-MR and cat-BIC remain the recommendation.

Discussion

The aim of the current paper was to evaluate how, when estimating networks on ordinal data, the performance of three estimation methods was affected by various data characteristics and the use of either the polychoric or Pearson correlation in the estimation process. Our results showed that, for structure recovery, the choice of polychoric versus Pearson correlations mattered most for the two methods that were based on the graphical lasso (the BIC and EBIC methods), such that when there were only 3 or 4 categories, polychoric correlations resulted in higher sensitivity, but also a higher false positive rate. In contrast, the MR method produced comparable results regardless of correlation type, except when n was less than 500, where polychoric correlations resulted in both lower sensitivity and a lower false positive rate. In terms of the accuracy of the edge estimates, treating the data as ordinal resulted in greater bias for the MR and BIC methods, and less bias for the EBIC methods. For all three performance measures, the discrepancy between treating the data as ordinal and treating the data as continuous decreased as the number of levels increased, such that the performance tended to be similar at 5 or 7 levels, echoing similar results in the SEM literature (Finney & DiStefano, 2013; Rhemtulla et al., 2012). Having an underlying non-normal distribution and/or extremely asymmetric thresholds generally resulted in worse performance for all three methods, and increased the discrepancy between categorical and continuous estimation methods.

Our results confirm a number of previous notes in the literature about estimating networks with ordinal data. First, Epskamp (2017) found that cat-EBIC would often estimate dense networks when sample size was small in data with five categories, and Epskamp and Fried (2018) pointed out that dense networks were more likely to be estimated when polychoric correlation matrices were used, due to the higher likelihood of not being positive definite. Even after excluding results where the correlation matrix was not positive definite, we saw a similar result, although density of the estimated networks was more influenced by other features of the data (the number of categories, the underlying distribution, and the threshold symmetry) than the sample size. In particular, when the number of categories was low, cat-BIC and cat-EBIC selected dense networks, resulting in sensitivity above 70% but also a false positive rate near or above 30%. The tendency to select dense networks was even greater when the underlying distribution was non-normal and/or the thresholds extremely asymmetric. Second, Epskamp and Fried (2018) pointed out that, with ordinal data, low frequencies in some categories or skewed data might lead to greater bias in the polychoric partial correlations. We confirmed this result here – when thresholds were extremely asymmetric (leading to some response categories having far fewer endorsements than others), and especially when combined with an underlying non-normal distribution (which resulted in some response categories having almost no endorsements), the relative bias tended be higher.

Our paper is unique in that it included two different types of non-normal distributions in the comparisons – one non-normal distribution generated by the Vale-Maurelli method, and the other via the independent generator method. When considering only the non-normal data generated by the Vale-Maurelli method, our results showed that network performance was only modestly impacted by the violation of the normality assumption, even though the skewness and kurtosis values were chosen to represent a moderately non-normal underlying distribution. However, in light of the concerns raised by Grønneberg and Foldnes (2019) and Foldnes and Grønneberg (2020) on the equivalence between discretized data produced by the Vale-Maurelli method and discretized data produced by a multivariate normal distribution, we also included a more “extreme” version of non-normality that avoids this issue. When non-normal data were produced by the independent generator method, we found that network performance significantly worsened, and the unstable performance of some of the estimation methods was made evident. Therefore, our paper showed that the impact of an underlying non-normal distribution depends on the type of underlying non-normality, and is not always modest or ignorable.

Finally, our paper highlighted the importance of simulating data that resembles data in the field. Our main results generated population networks from the set of partial correlations calculated on a psychological dataset, so that the simulated data would more closely mirror characteristics of real psychological data (e.g., most partial correlations have a value below 0.5; Wysocki & Rhemtulla, 2019) than if data had been simulated from a distribution of partial correlation values. Our follow-up simulation showed that if partial correlation values were instead drawn from a G-Wishart distribution with 3 degrees of freedom, then the performance of all methods were improved, which may cause the impression that the methods studied perform better than they would when actually applied to psychological data.

Recommendations

Our paper not only focused on the impact of various data characteristics on network estimation performance, but also on how each of the three methods considered here compared to each other when dealing with ordinal data. Although these three methods are by no means a comprehensive list of the estimation methods available to researchers, we can make some general recommendations based on our results. These recommendations, as always, come with the caveats of being dependent on specific features of the simulated data.

First, researchers should be aware that in certain scenarios, no method can be relied on to produce good results. If the sample size is less than 1,000, then the sensitivity of all methods will be below 70%. Even with a sample size of 3,000, the sensitivity will still be below 70% if the number of categories is low, the underlying distribution non-normal, *and* the thresholds are extremely asymmetric. Similarly, if the number of categories is low or the thresholds are extremely asymmetric (i.e., some response options are used very infrequently), or either of these two conditions are combined with an underlying non-normal distribution, then all methods will produce substantially biased estimates. Finally, if researchers wanted to use either the cat-BIC or cat-EBIC method (since the latter is the current recommendation), then a non-normal underlying distribution or extremely asymmetric thresholds will result in unexpectedly dense networks that can be hard to interpret.

Second, out of the three network estimation methods considered here, no method performed consistently well across conditions or performance measures. Therefore, researchers will have to prioritize a particular performance metric – whether the focus is on obtaining an accurate network structure (and in that case, whether they are more interested in maximizing the number of true positives or minimizing false positives), or on obtaining accurate edge estimates. The strengths and weaknesses of each method are summarized in Table 4. If the focus is on maximizing the sensitivity (i.e., true positive rate), then the MR method (cont-MR for sample sizes below 500, but either cont-MR or cat-MR for larger sample sizes) performs well whilst remaining stable across distribution and symmetry type. Conversely, if the focus is on minimizing the false positive rate (i.e., ensuring that most of the estimated edges are truly present in the population), then cont-BIC tends to have the lowest false positive rate across almost all conditions, and in fact makes almost no false positives when the sample size is large. Finally, if the focus is only on obtaining accurate edge estimates, regardless of the structure of the network, then cat-EBIC is preferred for sample sizes below 500, and cat-MR or cat-BIC for larger samples.

The above recommendations are based on results where population networks were derived from partial correlation values found in a psychological dataset. Although we believe these recommendations are more pertinent to applied researchers, there may be concerns about the generalizability of these recommendations. However, our follow-up simulations demonstrated that most recommendations would carry over to scenarios where the distribution of partial correlations resemble that of a G-Wishart distribution with 3 degrees of freedom, i.e., most partial correlation values fall in the mid-range of [-1, 1], but larger values are possible. The only change is that if researchers believe their data looks more like that of the G-Wishart distribution, and their goal is to obtain accurate edge estimates, then cont-MR or cont-BIC may be preferred over cat-EBIC at smaller sample sizes, as these methods produced less bias in the follow-up simulations (despite being only second-best in the initial simulation).

Limitations

There are a number of limitations to the present study. The first is that our paper only examined three different network estimation techniques. There is a vast array of possible network estimation techniques that could be used when estimating networks on ordinal data. For example, the *mgm* package can estimate networks from data where each variable is either continuous or nominal (thus accounting for the categorical nature of ordinal variables but not the ordering), and the *psychonetrics* package uses the weighted least squares estimator to estimate networks on ordinal data (Epskamp, 2021; Haslbeck & Waldorp, 2020). The three techniques evaluated here were chosen because they are either widely used (e.g., the EBIC and BIC methods) or are theoretically based on common network estimation techniques (e.g., the MR method was based on nodewise regressions).

Second, while we were careful to use network edge values, sample sizes, network sizes, and connectivity estimates derived from empirical work in the field, the simulated data are nonetheless not a perfect representation of real data. For example, all variables in each condition were of the same type with the same underlying distribution and same thresholds. When estimating structural equation models, even if all variables have the same number of levels and underlying distribution, estimation results can change when each variable is allowed to be defined by its own set of threshold values (Foldnes & Grønneberg, 2020). Therefore, having variables in the same network with different numbers of levels or underlying thresholds may similarly affect network estimation performance.

Conclusion

The ubiquity of ordinal data in psychological research and the popularity of the network modeling approach for exploring associations among variables means that, at some point, psychological researchers may need to estimate a network on ordinal data. The present study showed that treating the data as ordinal in the network estimation process does not always result in better estimation performance, and characteristics of the ordinal data – the number of categories, underlying distribution, and symmetry of thresholds – all affect performance, and should be carefully considered when choosing an estimation approach and inspecting the results.

References

Abacioglu, C., Isvoranu, A.-M., Verkuyten, M., Thijs, J., & Epskamp, S. (2017). Exploring teachers' influence on student motivation in multicultural classrooms: A comparative network analysis. PsyArXiv. https://doi.org/10.31234/osf.io/ck2p5

Bates, D., & Maechler, M. (2019). Matrix: Sparse and dense matrix classes and methods. https://CRAN.R-project.org/package=Matrix

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13(2),* 186-203. https://doi.org/10.1207/s15328007sem1302_2

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91-121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Cain, M. K., Zhang, Z., & Yuan, K.-H. (2016). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence, and estimation. *Behavior Research Methods, 49,* 1716-1735. https://doi.org/10.3758/s13428-016-0814-1

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1(1)*, 16-29.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. J. (2017a). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science, 8(5)*, 528-537. https://doi.org/10.1177/1948550617709827

Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J., & van der Maas, H. J. L. (2017b). Network structure explains the impact of attitudes on voting decisions. *Scientific Reports, 7(4909).* https://doi.org/10.1038/s41598-017-05048-y

Epskamp, S. (2017). Brief Report on Estimating Regularized Gaussian Networks from Continuous and Ordinal Data. PsyArXiv. arXiv:1606.05771v2

Epskamp, S. (2018, May 2). *New features in qgraph 1.5.* The Psychosystems Project. http://psychosystems.org/qgraph_1.5

Epskamp, S. (2021). psychonetrics: Structural equation modeling and confirmatory network analysis. https://CRAN.R-project.org/package=psychonetrics

Epskamp, S., Cramer, A. O. J., Waldorp, L., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualization of relationships in psychometric data. *Journal of Statistical Software, 48(4),* 1-18.

Epskamp, S., & Fried, E. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23(4)*, 617-634. https://doi.org/10.1037/met0000167

Finney, S. J., & DiStefano, C. (2013). *Nonnormal and categorical data in structural equation modeling.* In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. Structural equation modeling: A second course* (p. 439–492). IAP Information Age Publishing.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532. https://doi.org/10.1007/BF02293811

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

Foldnes, N., & Olsson, U. H. (2016). A simple simulation technique for nonnormal data with prespecified skewness, kurtosis, and covariance matrix. *Multivariate Behavioral Research, 51(2-3)*, 207-219. https://doi.org/10.1080/00273171.2015.1133274

Foldnes, N., & Grønneberg, S. (2020). *The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms* [Manuscript submitted for publication]. Department of Economics, BI Norwegian Business School.

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems, 23,* 2020-2028.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse invariance covariance estimation with the graphical lasso. *Biostatistics, 9(3)*, 432-441. https://doi.org/10.1093/biostatistics/kxm045

Friedman, J., Hastie, T., & Tibshirani, R. (2019). glasso: Graphical lasso: Estimation of Gaussian Graphical Models. https://CRAN.R-project.org/package=glasso

Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *Psyhometrika, 82,* 1035-1051. https://doi.org/10.1007/s11336-017-9569-6

Grønneberg, S., & Foldnes, N. (2019). A problem with discretizing Vale-Maurelli in simulation studies. *Psychometrika, 84,* 554-561. https://doi.org/10.1007/s11336-019-09663-8

Haslbeck, J. M. B., & Waldorp, L. J. (2020). mgm: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software, 93(8)*. https:://doi.org/ 10.18637/jss.v093.i08

Higham, N. (2002). Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis, 22,* 329-343.

Hsu, C.-W., Sinay, M. S., & Hsu, J. S. J. (2012). Bayesian estimation of a covariance matrix with flexible prior specification. *Annals of the Institute of Statistical Mathematics, 64(2)*, 319-342. DOI: 10.1007/s10463-010- 0314-5

Johal, S., & Rhemtulla, M. (2021). Comparing estimation methods for psychometric networks with ordinal data. Retrieved from osf.io/t4gb9.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). semTools: Useful tools for structural equation modeling. https://CRAN.R-project.org/package=semTools

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity, 43*, 495-507. https://doi.org/10.1007/s11135-007-9133-z

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48,* 936-949. https://doi.org/10.3758/s13428-015-0619-7

Lorenzo-Seva, U., & Ferrando, P. J. (2020). Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution. *Structural Equation Modeling*. https://doi.org/10.1080/10705511.2020.1735393

Mazumder, R., & Hastie, T. (2012). The graphical lasso: New insights and alternatives. *Electronic Journal of Statistics, 6*, 2125-2149. https:://doi.org/10.1214/12-EJS740

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics, 34(3)*, 1436-1462. https://doi.org/10.1214/009053606000000281

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105(1)*, 156-166. https://doi.org/10.1037/0033-2909.105.1.156

Mohammadi, R., & Wit, E. C. (2019). “BDgraph: An R package for Bayesian structure learning in graphical models.” *Journal of Statistical Software, 89(3)*, 1-30. https://doi.org/10.18637/jss.v089.i03

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes [Unpublished manuscript].

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44(4)*, 443-460. https://doi.org/10.1007/BF02296207

Olvera Astivia, O. L. (2020). Issues, problems, and potential solutions when simulating contnuous, non-normal data in the social sciences. *Meta-Psychology, 4*. https://doi.org/10.15626/MP.2019.2117

Revelle, W. (2019). psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 1.9.12

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software, 48(2)*, 1-36.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, 58(1),* 267-288.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 465-471. https://doi.org/10.1007/BF02293687

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Lynn, B., Schoevers, R. A., &

Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports, 4(5918)*. https://doi.org/10.1038/srep05918

van der Maas, H. L. J., Kan, K.-J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence, 5(2)*. https://doi.org/10.3390/jintelligence5020016

Williams, D. R., Briganti, G., Linkowski, P., Mulder, J., & Rhemtulla, M. (2020). *On accepting the null hypothesis of conditinal independence in partial correlation networks: A Bayesian analysis* [Manuscript in preparation]. Department of Psychology, University of California, Davis.

Williams, D. R., & Rast, P. (2019). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology, 73*, 187-212. https://doi.org/10.1111/bmsp.12173

Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On Nonregularized Estimation of Psychological Networks. *Multivariate Behavioral Research. 54(5)*, 719-750. https://doi.org/10.1080/00273171.2019.1575716

Wysocki, A. C., & Rhemtulla, M. (2019). On penalty parameter selection for estimating network models. *Multivariate Behavioral Research.* https://doi.org/10.1080/00273171.2019.1672516

Tables and Figures

Table 1

Summary of Network Estimation Approaches Used

Method	Input	Abbreviation	What is being estimated?	Zeroes determined by:	Non-zero elements represent:
EBIC graphical lasso	Pearson correlation	cont-EBIC	Precision matrix (K)	l_1-regularization (setting $\gamma=0.5$¿	Shrunken partial correlations
	Polychoric correlation	cat-EBIC			
Nonregularized Bayesian Information Criterion	Pearson correlation	cont-BIC	Precision matrix (K)	l_1-regularization (setting $\gamma=0$¿ + iterative changes	Nonregularized partial correlations
	Polychoric correlation	cat-BIC			
Nonregularized multiple regression	Pearson correlation	cont-MR	Regression coefficients (β)	p-value < .05, using AND rule	Nonregularized partial correlations
	Polychoric correlation	cat-MR			

Table 2

Threshold Values

		Category Thresholds					
Threshold Condition	No. of Levels	h_1	h_2	h_3	h_4	h_5	h_6
Symmetric	3	-0.83	0.83				
	4	-1.25	0.00	1.25			
	5	-1.50	-0.50	0.50	1.50		
	7	-1.79	-1.07	-0.36	0.36	1.07	1.79
Moderately Asymmetric	3	-0.50	0.76				
	4	-0.31	0.79	1.66			
	5	-0.70	0.39	1.16	2.05		
	7	-1.43	-0.43	0.38	0.94	1.44	2.54
Extremely Asymmetric	3	0.58	1.13				
	4	0.28	0.71	1.23			
	5	0.05	0.44	0.84	1.34		
	7	-0.25	0.13	0.47	0.81	1.18	1.64

Note: h_i represents the category threshold value, which relates the underlying continuous value,

$Y_j^{¿}$, to the observed ordinal response, Y_j, such that $Y_j = \begin{cases} 1\, Y_j^{¿} \le h_1 \\ 2\, h_1 < Y_j^{¿} \le h_2 \\ \vdots\ \vdots \\ 7\, Y_j^{¿} \ge h_6 \end{cases}$

Table 3

Convergence Failures for the BIC and MR Methods

		BIC Method Total Failures: 27,948 (2.6%)		MR Method Total Failures: 18,971 (1.8%)	
		cat-BIC	cont-BIC	cat-MR	cont-MR
Number of Nodes					
	10	0.13%	< .01%	< .01%	< .01%
	20	2.46%	< .001%	0.04%	1.71%
Connectivity					
	0.3	0.54%	< .01%	< .01%	0.39%
	0.5	0.78%	< .01%	< .01%	0.54%
	0.7	1.26%	< .01%	< .01%	0.79%
Sample Size					
	100	2.25%	< .01%	0.03%	0.50%
	250	0.19%	< .01%	< .01%	0.34%
	500	0.06%	< .01%	< .01%	0.31%
	1,000	0.05%	< .01%	< .01%	0.28%
	3,000	0.04%	< .01%	< .01%	0.29%
Distribution Type					
	Normal	0.46%	< .01%	0.02%	0.61%
	Non-Normal (VM)	0.59%	0%	< .01%	0.66%
	Non-Normal	1.63%	< .01%	0.01%	0.44%

	(IG)				
Threshold Symmetry					
	Symmetric	0.71%	< .01%	0.02%	0.41%
	Moderately Asymmetric	0.32%	0%	0.01%	0.42%
	Extremely Asymmetric	1.56%	< .01%	0.01%	0.89%
Levels					
	3	1.47%	0%	0.02%	0.14%
	4	0.70%	< .01%	0.01%	0.17%
	5	0.34%	0%	0.01%	0.36%
	7	0.09%	< .01%	< .01%	1.05%

Note. Total value in the second row represents the total number of convergence failures out of 1,080,000 iterations. Cell values represent the percentage of failures (out of 1,080,000 iterations) for that factor, collapsed across all other factors.

Table 4

Strengths and Weaknesses of Estimation Methods

Estimation Method	Strengths	Weaknesses
EBIC	• No convergence failures • High sensitivity when n > 1,000 • Less bias than other methods when n < 500	• Low sensitivity when n < 1,000 (selects empty networks) • Unstable structure recovery performance when treating data as ordinal • False positive rate increases with sample size • Bias does not decrease with increasing sample size, and greater bias than other methods at $n \geq 500$
BIC	• When treating data as continuous, minimizes false positives more than other methods • Bias decreases with increasing sample size until essentially	• Highest number of convergence failures, especially with polychoric correlations • Unstable structure recovery performance when treating

	unbiased when n = 3,000	data as ordinal • Substantially biased estimates when n < 500
MR	• Not as affected by treating the data as ordinal versus continuous • Stable sensitivity performance across all conditions • Bias decreases with increasing sample size until essentially unbiased by n = 3,000	• Slight increase in false positive rate with sample size • Substantially biased estimates when n < 500

Figure 1

Diagrammatic Representation of the Nonregularized Bayesian Information Criterion Method

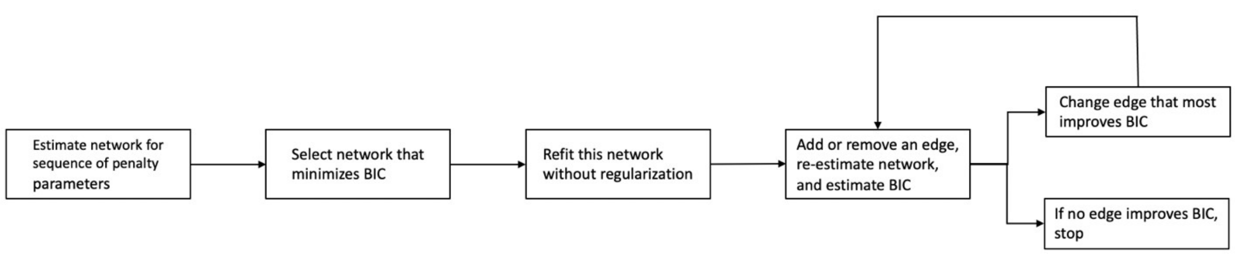
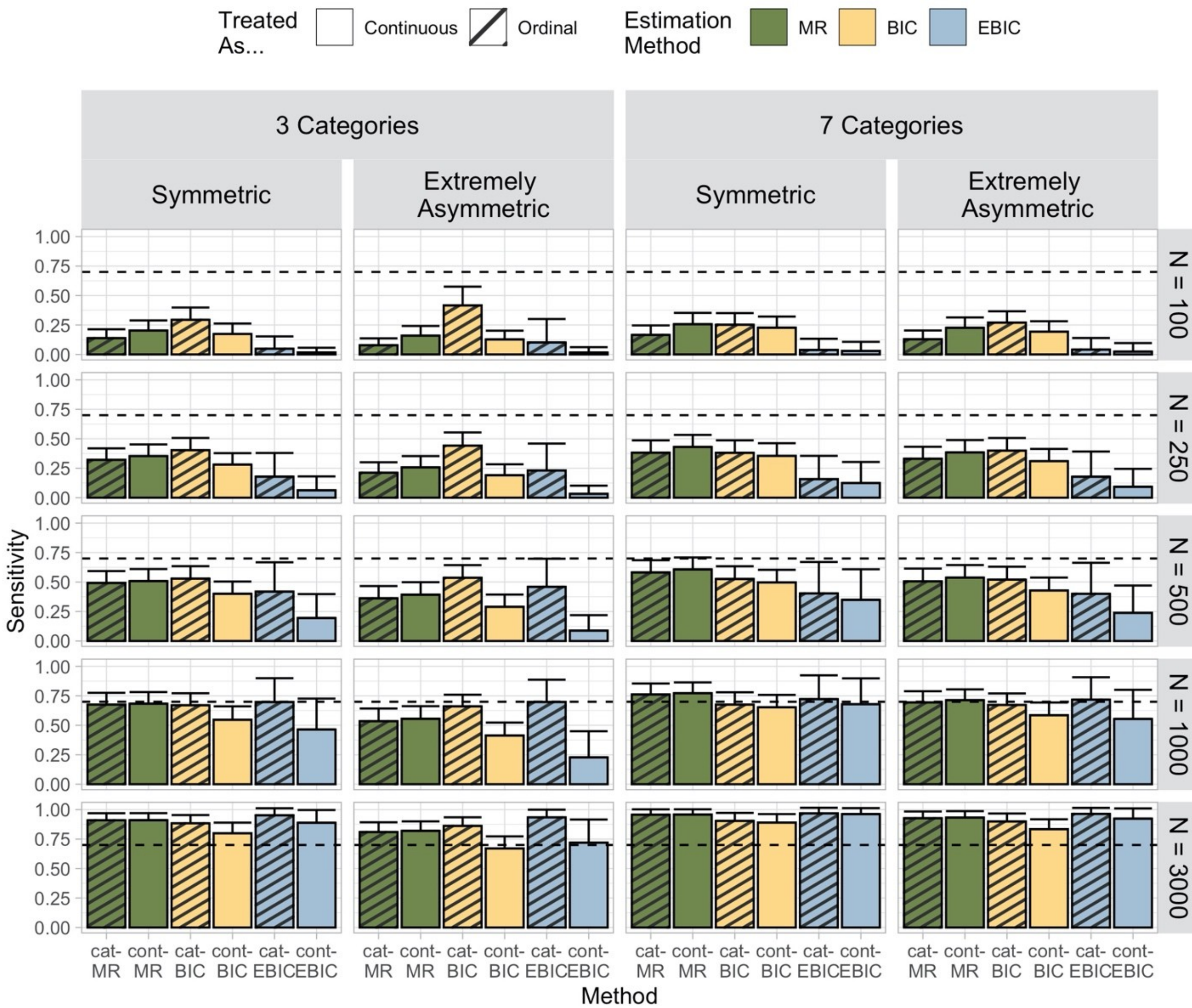


Figure 2

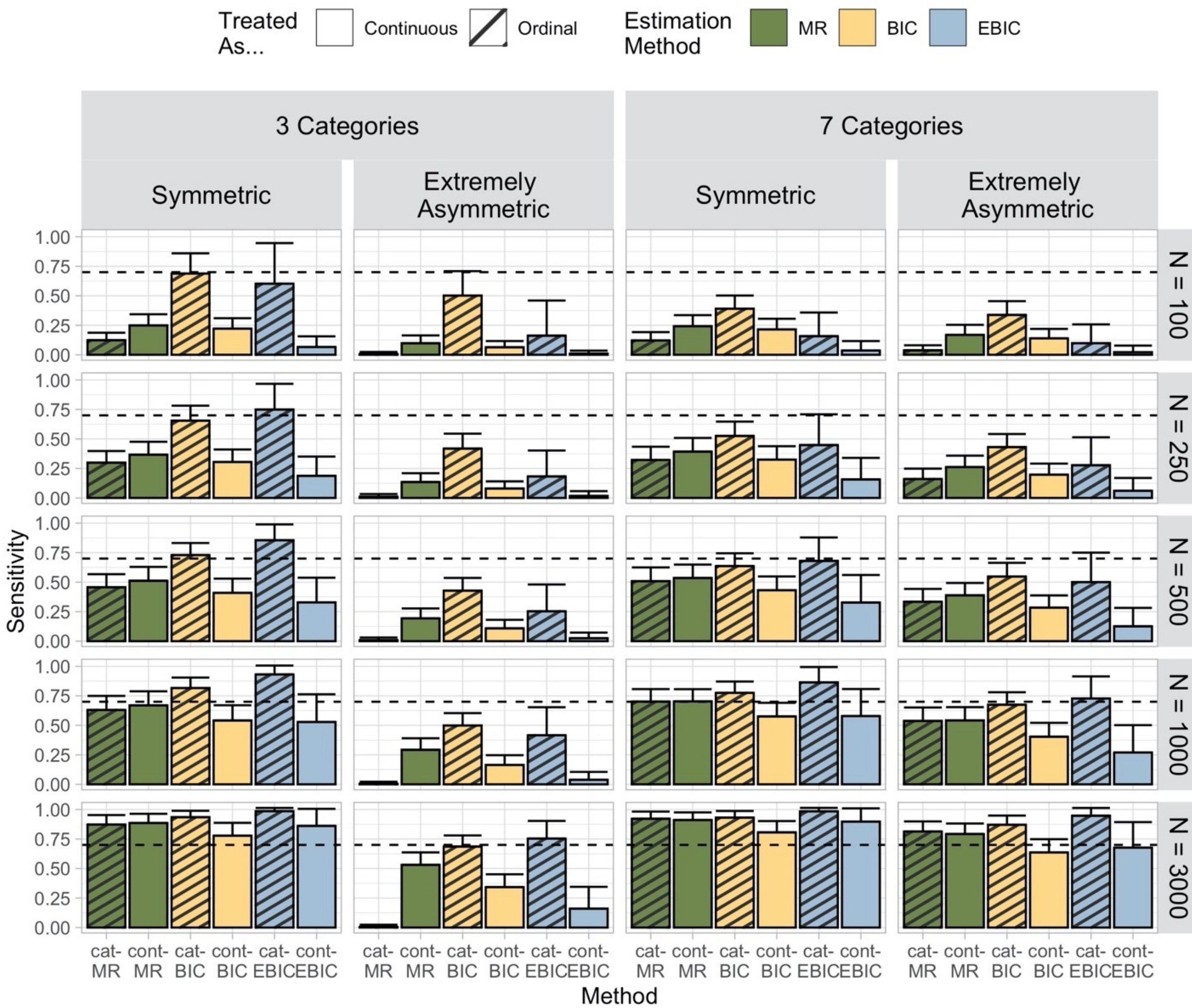
Sensitivity for Networks Estimated from Underlying Normal Distribution



Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed horizontal line represents 70% sensitivity, and the error bars represent 1 standard deviation.

Figure 3

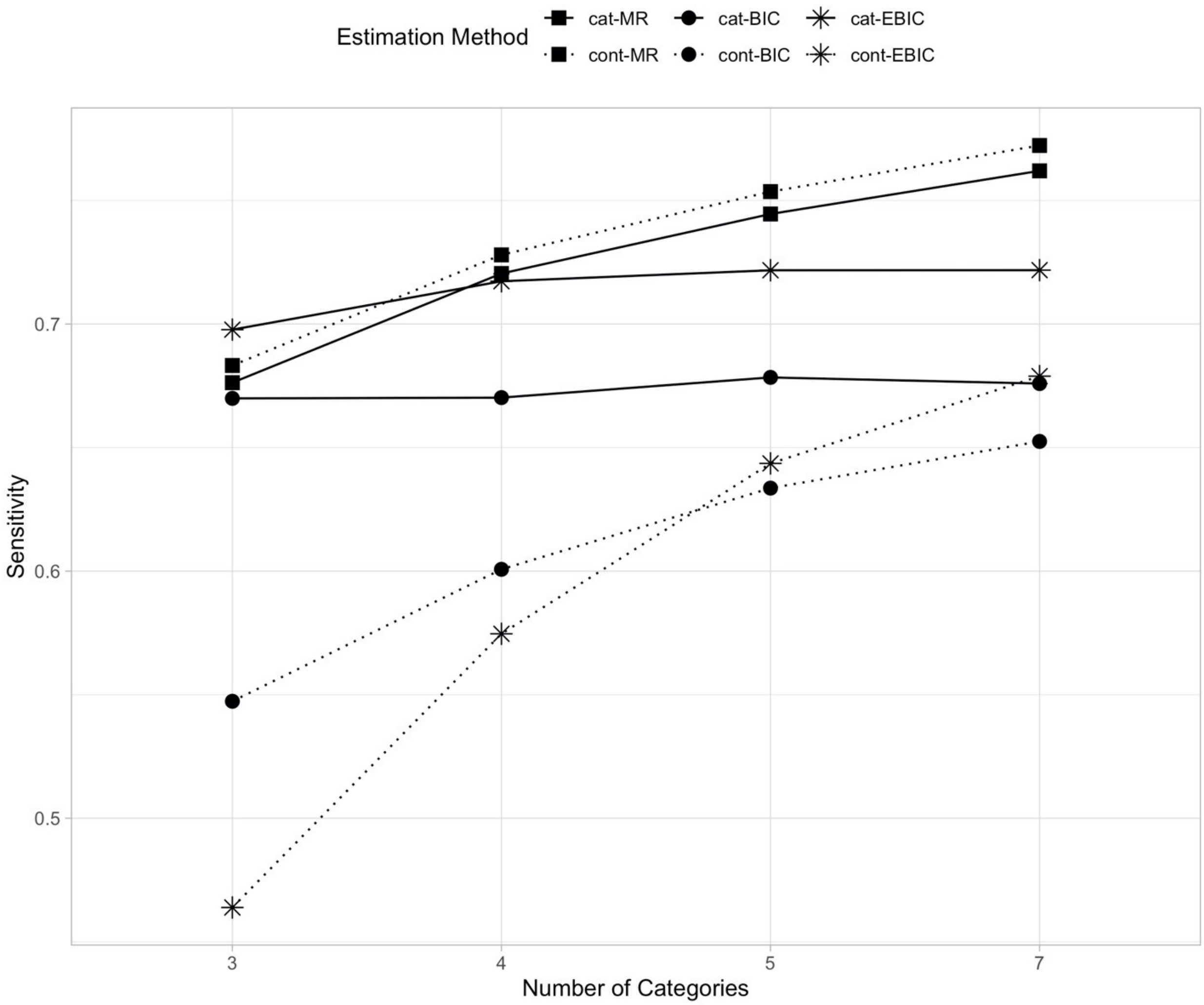
Sensitivity for Networks Estimated from Underlying Non-Normal Distribution



Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed line represents 70% sensitivity, and the error bars represent 1 standard deviation.

Figure 4

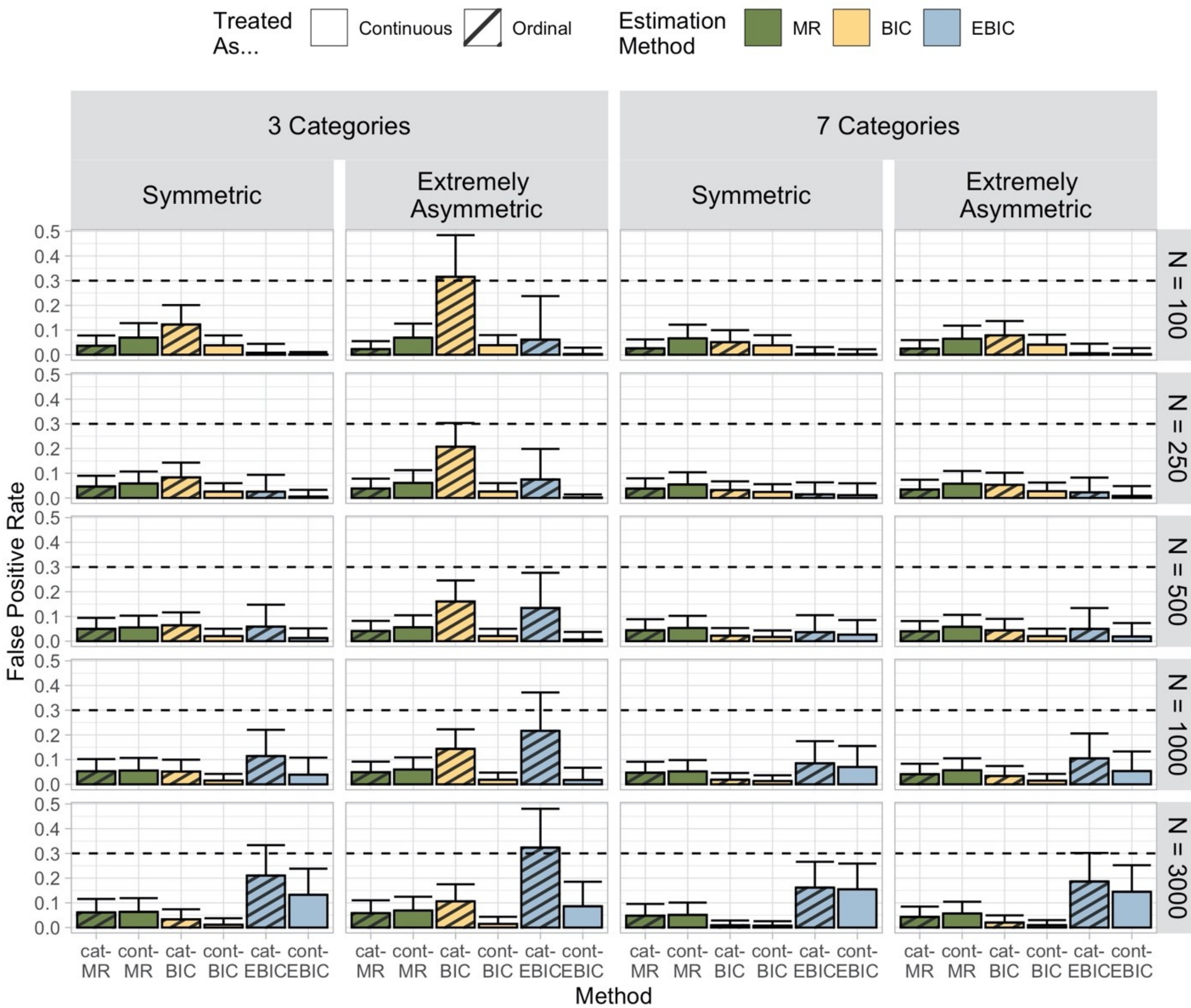
Sensitivity as a Function of the Number of Levels



Note. Sensitivity is plotted as a function of the number of levels for n = 1,000, with an underlying normal distribution and symmetric thresholds. However, this pattern tends to hold across different sample sizes, underlying distributions, and threshold symmetries. Dotted lines represent data that were treated as continuous, while solid lines represent data that were treated as ordinal.

Figure 5

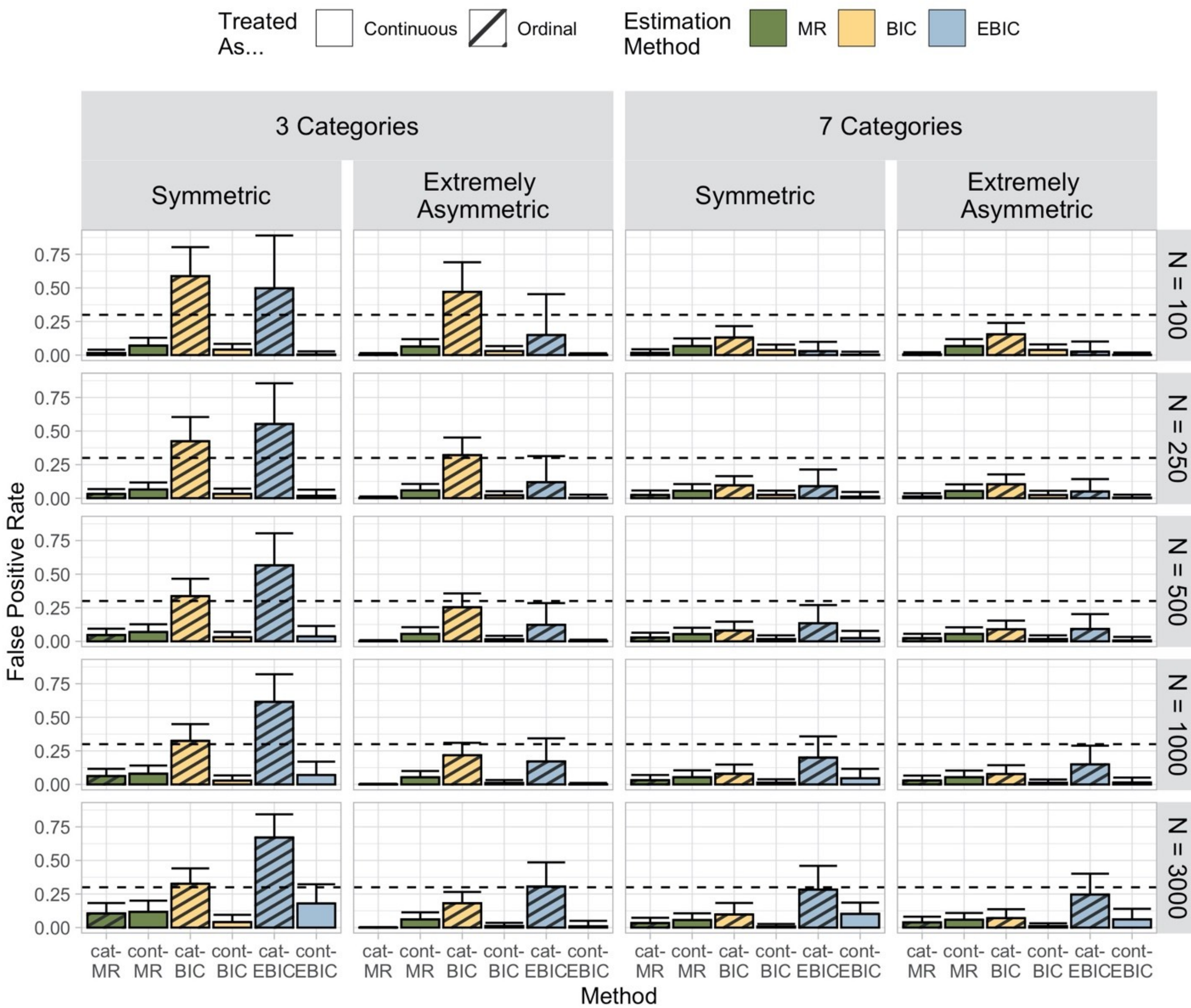
False Positive Rate for Networks Estimated from Underlying Normal Distribution



Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed line represents a 30% false positive rate, and the error bars represent 1 standard deviation.

Figure 6

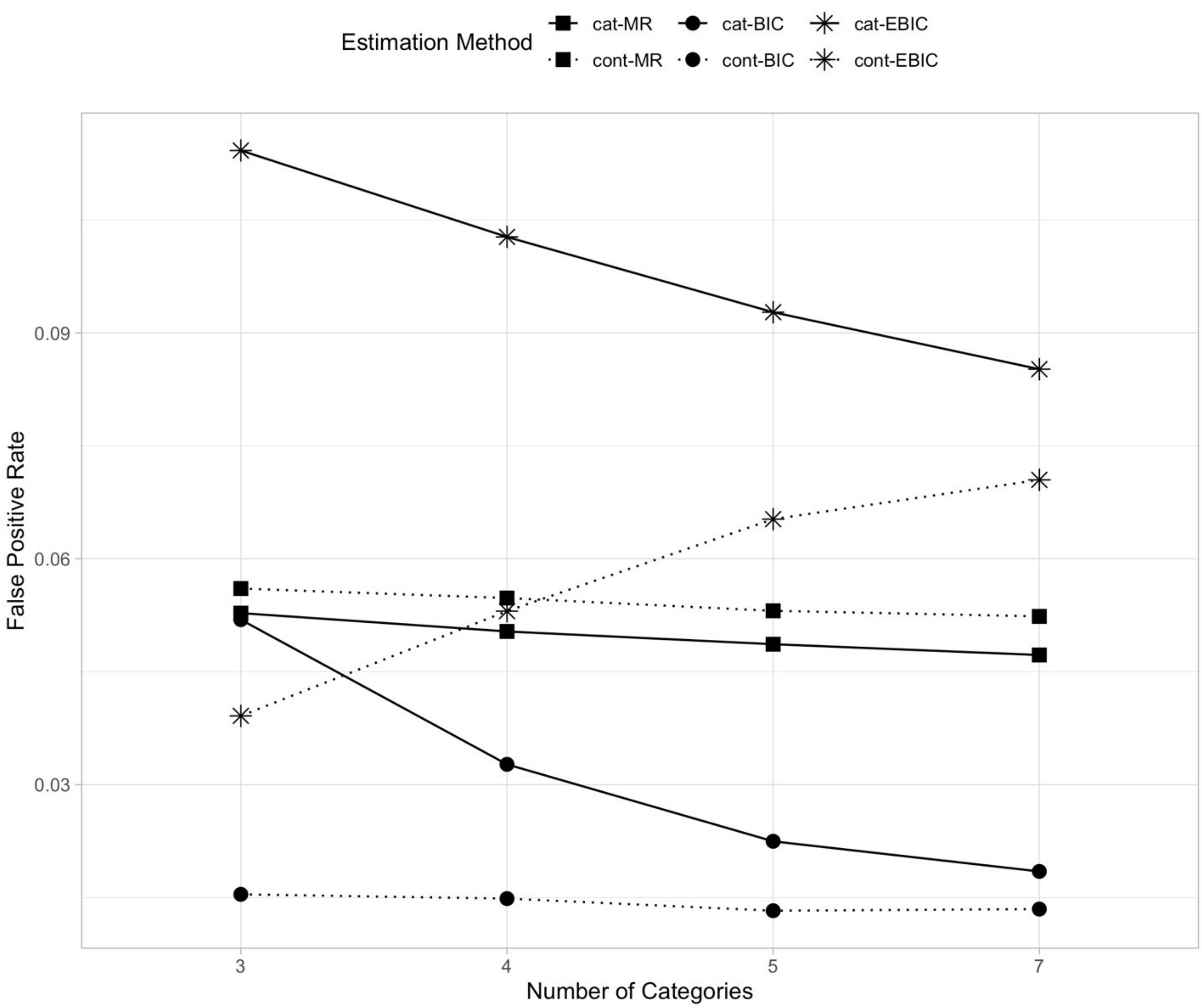
False Positive Rate for Networks Estimated from Underlying Non-Normal Distribution



Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed line represents a 30% false positive rate, and the error bars represent 1 standard deviation.

Figure 7

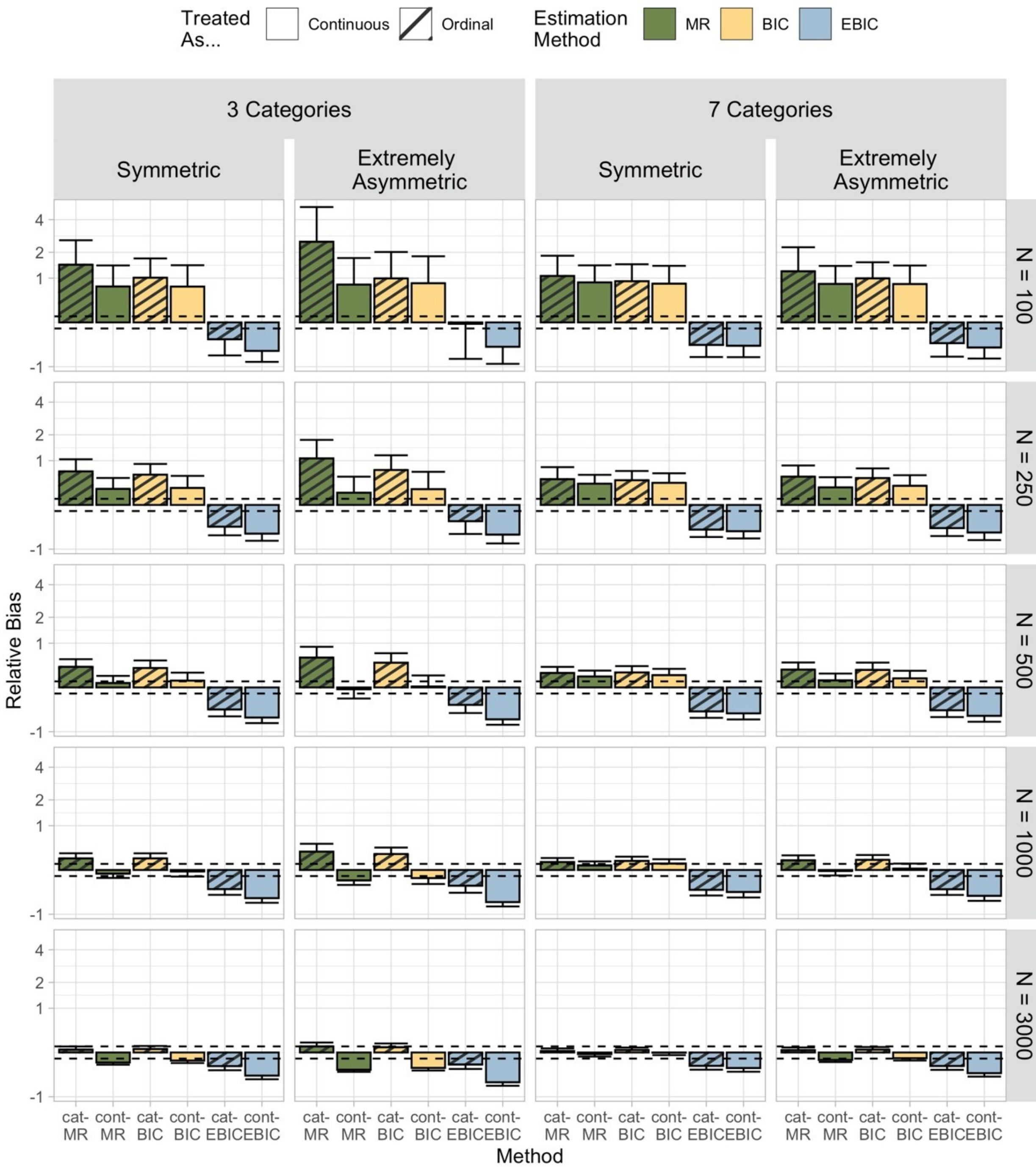
False Positive Rate as a Function of the Number of Levels



Note. False positive rate is plotted as a function of the number of levels for n = 1,000, with an underlying normal distribution and symmetric thresholds. Dotted lines represent data that were treated as continuous, while solid lines represent data that were treated as ordinal.

Figure 8

Relative Bias for True Positive Edges in Networks Estimated from Underlying Normal Distribution

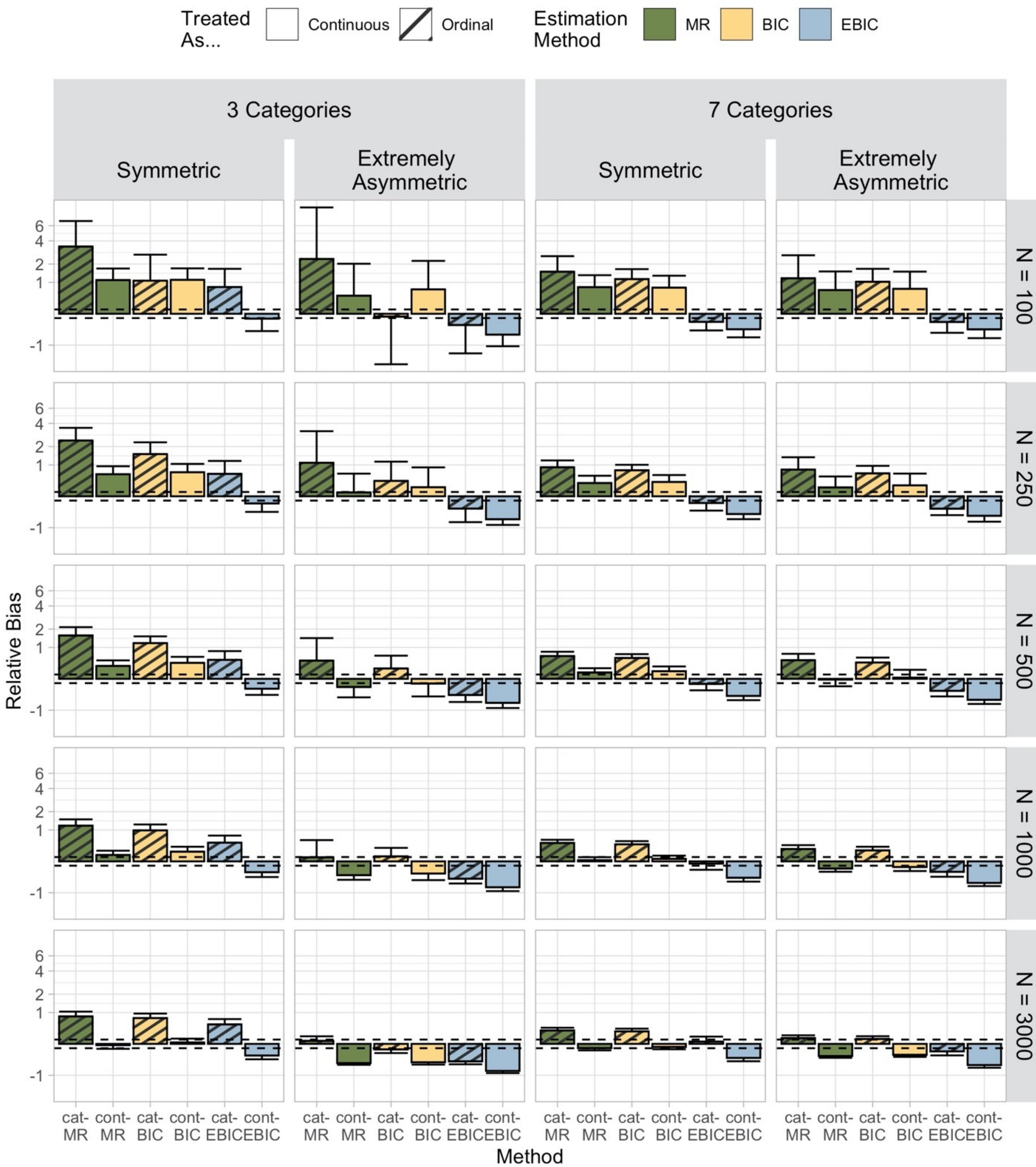


Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed lines correspond to a relative bias of -0.1 or 0.1 to aid in detecting estimates that are substantially biased, and the error bars represent 1 standard deviation. To facilitate comparisons across graphs, the plotted values represent $\log(|Relative\ Bias|+1)$ although the y-axis is in the scale of the original relative bias.

Figure 9

Relative Bias for True Positive Edges in Networks Estimated from Underlying Non-Normal

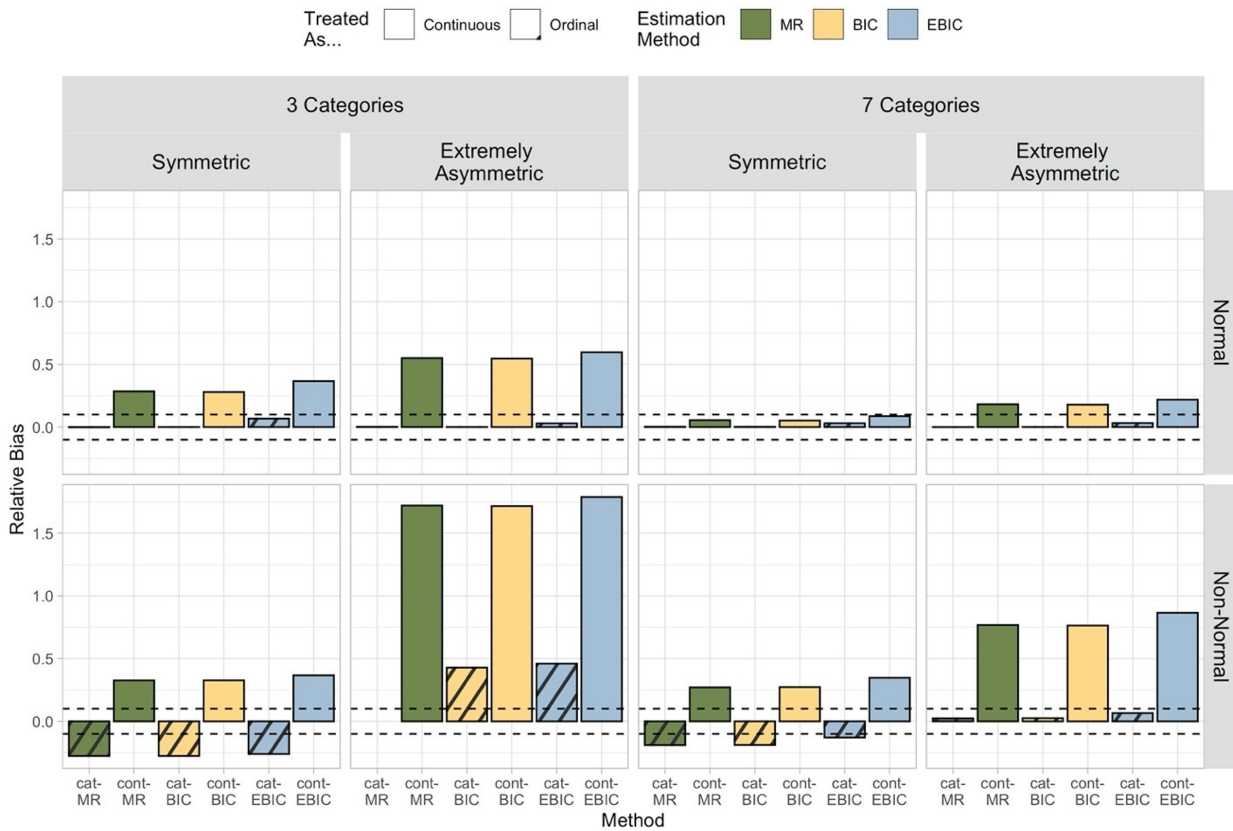
Distribution



Note. Different sample sizes are shown in the rows (increasing from top to bottom), while results for conditions with 3 levels are in the left 2 columns, and 7 levels are in the right 2 columns. Within each pair of columns, conditions with symmetric thresholds are on the left and those with extremely asymmetric thresholds are on the right. The dashed lines correspond to a relative bias of -0.1 or 0.1 to aid in detecting estimates that are substantially biased, and the error bars represent 1 standard deviation. To facilitate comparisons across graphs, the plotted values represent $\log\left(|Relative\ Bias|+1\right)$ although the y-axis is in the scale of the original relative bias.

Figure 10

Relative Bias for True Positive Edges in Networks Estimated with a Sample Size of 1,000,000

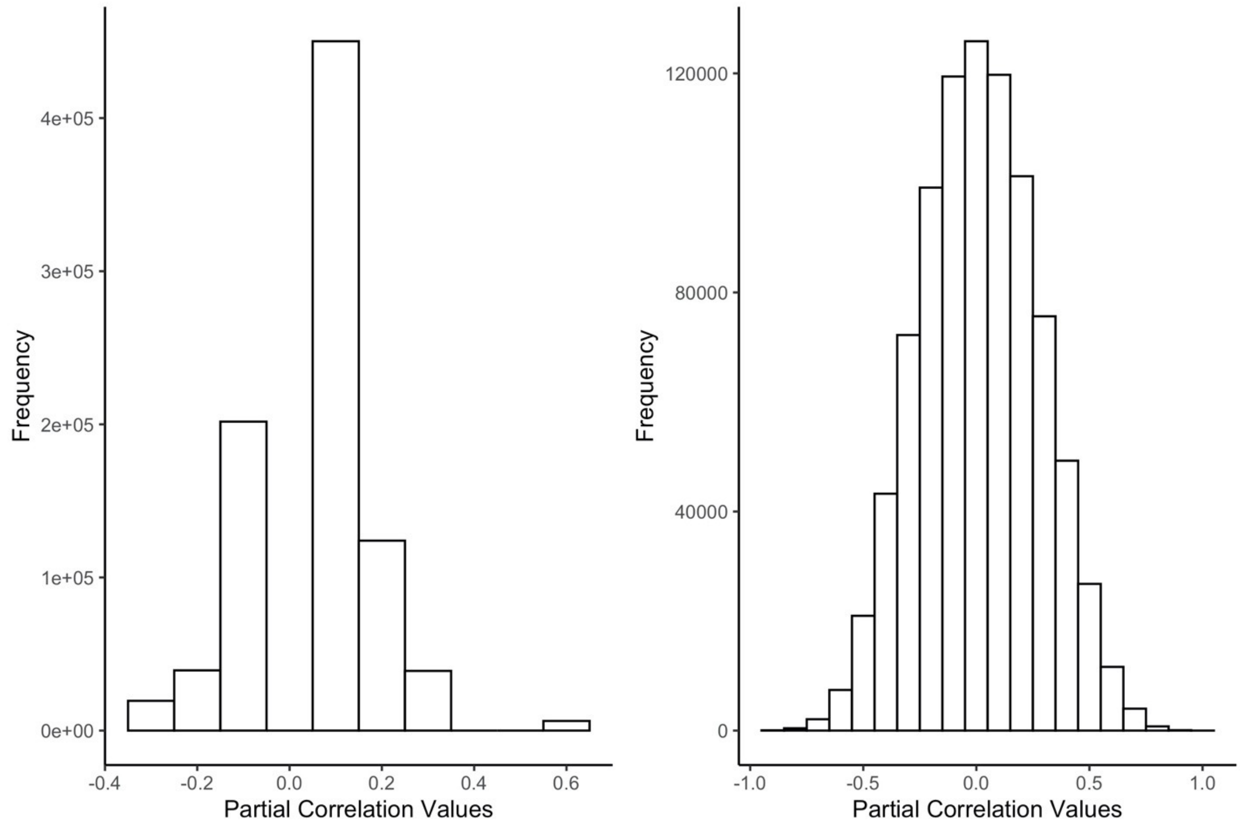


Note. The top row shows relative bias for networks estimated under an underlying normal distribution, while the bottom row shows relative bias for networks estimated under an underlying non-normal distribution. Within each row, results for conditions with 3 levels are in the leftmost columns, and conditions with 7 levels in the rightmost. Finally, within each pair of columns, the left column represents conditions with symmetric threshold, and the right extremely asymmetric thresholds. There are no results for cat-MR when the data have an underlying non-normal distribution and thresholds are extremely asymmetric due to a convergence failure.

Figure 11

Histograms Displaying Ranges of Population Partial Correlations

(a) (b)



Note. Histograms illustrating the range of population partial correlation values for (a) the empirical dataset used in the main simulation and (b) the randomly-generated values used in the follow-up simulation.