The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms

Njål Foldnes and Steffen Grønneberg

BI Norwegian Business School

Structural equation modeling (SEM) of ordinal data is often performed using normal theory maximum likelihood estimation based on the Pearson correlation (cont-ML) or using least squares principles based on the polychoric correlation matrix (cat-LS). While cont-ML ignores the categorical nature of the data, cat-LS assumes underlying multivariate normality. Theoretical results are provided on the validity of treating ordinal data as continuous when the number of categories increases, leading to an adjustment to cont-ML (cont-ML-adj). Previous simulation studies have concluded that cat-LS outperforms cont-ML, and that it is quite robust to violations of underlying normality. However, this conclusion was based on a data simulation methodology equivalent to discretizing exactly normal data. The present study employs a new simulation method for ordinal data to re-investigate whether ordinal SEM is robust to underlying non-normality. In contrast to previous studies, we include a large set of ordinal distributions, and our results indicate that ordinal SEM estimation and inference is highly sensitive to the interaction between underlying non-normality and the ordinal observed distributions. Our results show that cont-ML-adj consistently outperforms cont-ML, and that cat-LS is less biased than cont-ML-adj. The sensitivity of cat-LS to violation of underlying normality necessitates the need for a test of underlying normality. A bootstrap test is found to reliably detect underlying non-normality.

Keywords: Ordinal data, latent variable models, underlying normality, simulation

Variables measured on an ordered categorical scale abound in the psychological and educational sciences. Such ordinal variables often stem from Likert-type scales, and are frequently used as indicators for latent variables that are related according to substantive theory in a structural equation model (SEM). In the following we use the term ordinal SEM to refer to SEM applied to ordinal data. Initially, SEM estimators were developed under the assumption of multivariate normal continuous data (Jöreskog, 1967), with later developments for non-normal continuous data (Browne, 1984; Satorra & Bentler, 1988). In principle, one should not apply estimators based on continuous data to ordinal indicators, since data at the ordinal level of measurement do not support calculation of means and covariances. However, many researchers still model data with ordinal indicators as continuous (e.g., Brennan et al., 2017; Gaspard et al., 2018; Marsh, Abduljabbar, et al., 2013; Marsh, Vallerand, et al., 2013; Neff, Whittaker, & Karl, 2017), using the normal theory maximum likelihood (ML) estimator, with robustified standard errors and model fit statistics. We refer to this con-

tinuous methodology as *cont-ML*. One reason for using cont-ML with ordinal data is the relative ease of handling missing data and large complex models.

A popular approach of modeling the ordinal nature of the data is to add a discretization process based on thresholds to the standard SEM. This approach was first proposed by Christoffersson (1977) for binary data, and subsequently developed to handle observed variables measured with more than two ordered categories (Jöreskog, 1994; Muthén, 1984). First, for any pair of ordinal variables, the so-called *polychoric correlation* is estimated. This is the correlation between two underlying unobserved variables that are posited to have created the observed categorical data through discretization. Since the time of Pearson (1900) it has been a standard assumption that these underlying variables follow a normal distribution. In the second stage, the model parameters are estimated by fitting the model to the matrix of polychoric correlations using least-square principles, and we refer to this methodology as *cat-LS*. In cat-LS formulas for standard errors and model test statistics (Muthén, 1993) are similar to those proposed for non-normal data in the continuous case (Satorra & Bentler, 1988)[1]. Simulation stud-

---

Correspondence concerning this article should be addressed to Njål Foldnes, BI Norwegian Business School. E-mail: njal.foldnes@gmail.com

---

[1]In Mplus (Muthén & Muthén, 2012) cont-ML and cat-LS are available as MLR and WLSMV estimation, respectively. In lavaan (Rosseel, 2012) the cont-ML is obtained with "estima-

ies (e.g., Beauducel & Herzberg, 2006; Li, 2016b; Rhemtulla, Brosseau-Liard, & Savalei, 2012) indicate that cat-LS in general outperforms cont-ML. However, the discrepancy between these two methodologies has been reported to decrease with five or more categories that are symmetrically distributed (e.g., Rhemtulla et al., 2012).

Both cont-ML and cat-LS are frequently used to estimate SEM models with ordinal data, and their properties have been extensively studied in many simulation studies. The consequences of underlying non-normality on estimation bias have been thoroughly investigated (e.g., Flora & Curran, 2004; Li, 2016a; Quiroga, 1994; Rhemtulla et al., 2012), resulting in a general consensus that ordinal SEM is quite robust to moderate violation of underlying normality. However, these studies were conducted using simulation methodology which is equivalent to simulating exactly normal data (Grønneberg & Foldnes, 2019). In the present study, we reassess the robustness issue using a newly developed simulation methodology (Foldnes & Grønneberg, 2019) which ensures proper violation of underlying normality. We also propose and evaluate an alternative to cont-ML which takes advantage of the discretization model employed by cat-LS. We refer to this new method as *cont-ML-adj*. Inference for cont-ML-adj utilizes the same robustification formulas for standard errors and model fit statistics as cont-ML and cat-LS.

Our research questions are as follows. First, how sensitive are cont-ML, cont-ML-adj, and cat-LS to variation in the observed ordinal distributions? That is, we will study the stability, or lack thereof, of correlation and ordinal SEM estimates under a large number of discretizations of the same underlying continuous distribution. Previous studies (e.g., Beauducel & Herzberg, 2006; Li, 2016a, 2016b; Muthén & Kaplan, 1985; Rhemtulla et al., 2012) have included only a small set of threshold configurations to generate ordinal data, typically consisting of a symmetrical distribution and one or two asymmetrical observed ordinal distributions. The focus in these studies has been to compare the performance of different estimators, but using only a handful of threshold configurations. In the present study we utilize 150 threshold configurations in each design condition to assess the extent to which cont-ML, cont-ML-adj, and cat-LS estimation depend on the specific marginal ordinal distribution. To the best of our knowledge, this is a novel perspective leading to insights into the sensitivity of the three estimators to the distributional shapes of the ordinal univariate marginals.

The second research question is: How sensitive are cont-ML, cont-ML-adj, and cat-LS to departures from normality in the underlying dataset? In the case where the observed ordinal data are produced by the discretization of an underlying continuous vector that does not have the multivariate normal distribution, the normal theory polychoric estimator will typically be biased (Foldnes & Grønneberg, 2019b). That is,

the true correlation matrix among the underlying continuous variables will differ from the estimated polychoric correlations, even at the population level. This could have a profound impact on the second stage of cat-LS, with the SEM model estimated from a biased polychoric correlation matrix. Hence, in the case of underlying non-normality, a well-fitting SEM may be estimated with substantial bias in model parameters and fit indices, which invalidates inference. We apply newly developed simulation methodology for ordinal data in the context of a realistic SEM model to investigate the extent to which underlying non-normality produces bias in parameter estimation, confidence interval (CI) coverage rates, and in model fit assessment.

Our third research question relates to the proposed cont-ML-adj estimator. What are its theoretical properties, in relation to both cont-ML and cat-LS? And how does it empirically perform in the setting of a realistic SEM model?

Our fourth and final question is: How well can we detect underlying non-normality in an ordinal dataset, in the context of a medium-sized SEM model? In conditions where underlying non-normality is detrimental to SEM inference, we aim at detecting the underlying non-normality using a statistical test. A bootstrap test for this purpose has recently been proposed (Foldnes & Grønneberg, 2019b), and found to outperform the test of Maydeu-Olivares (2006) in a limited simulation study with small- and medium-sized factor models. In the present article we evaluate the performance of this bootstrap test in a medium-sized SEM setting.

The remainder of the article is structured as follows: We first provide a concise description of cont-ML, cont-ML-adj and cat-LS, and then describe a new simulation method for ordinal data that we employ in the present study. Next, a literature review provides the context of our research questions. This is followed by a discussion of the complexity of interpreting ordinal SEM simulation studies, which provides justification for employing the new simulation approach. Theoretical results on the consistency of our three estimators as the number of categories increase are provided. We report results from three empirical studies on the performance of cont-ML, cont-ML-adj and cat-LS in the context of a medium-sized SEM. We investigate how the shape of ordinal data distributions and underlying non-normality affect parameter bias, confidence interval coverage and model fit testing. The performance of the bootstrap test for underlying non-normality is also evaluated. We conclude with a discussion of the results with recommendations for applied researchers.

Estimation methods

There are two popular limited-information approaches to ordinal SEM. The first approach is cont-ML, which calcu-

---

tor=MLM/MLMV" and cat-LS by specifying the "ordered" option.

lates the correlation or covariance matrix directly on the ordinal observations, which is then provided as input to the standard normal theory maximum likelihood estimator. The second approach is cat-LS, where we first estimate the correlation matrix of an underlying continuous random vector X^* and then uses least squares estimation methods to fit the proposed SEM to the estimated correlation matrix (Muthén, 1984). This approach is implemented in current software packages such as EQS (Bentler, 2006), Mplus (Muthén & Muthén, 2012), LISREL (Jöreskog & Sörbom, 2015), and lavaan (Rosseel, 2012), and is frequently employed by researchers.

cat-LS is based on the following discretization model. We have n independent and identically distributed observations of a random ordinal vector $X = (X_1, X_2, \ldots, X_p)'$. We assume that each X_k may take on discrete values $x_{k,1}, \ldots, x_{k,K}$ for some number K, and that the data stem from discretization of an unobserved continuous p-dimensional random vector $X^* = (X_1^*, X_2^*, \ldots, X_p^*)'$ as follows:

$$X_k = \begin{cases} x_{k,1}, & \text{if} & X_k^* \leq \tau_{k,1} \\ x_{k,2}, & \text{if} & \tau_{k,1} < X_k^* \leq \tau_{k,2} \\ \vdots & & \\ x_{k,K}, & \text{if} & \tau_{k,K-1} < X_k^* \end{cases} \quad (1)$$

for $k = 1, 2, \ldots, p$. We refer to the τ as thresholds which fulfills $\tau_{k,j-1} \leq \tau_{k,j}$ and define $\tau_{k,0} = F_k^{-1}(0), \tau_{k,K} = F_k^{-1}(1)$, where $F_k(x) = P(X_k^* \leq x)$ is the k-th marginal distribution of X^*. For all distributions with infinite support, such as the normal distribution, we have $\tau_{k,0} = -\infty, \tau_{k,K} = \infty$. To identify the thresholds, the marginals of X^* have to be known. Usual practice is to assume that X^* has standardized normal marginals: $X_k^* \sim N(0, 1)$ for $k = 1, 2, \ldots, p$.

The SEM specified by the researcher relates to the continuous X^* vector. That is, model estimation requires the correlation matrix of X^*. This matrix is not identified, and the traditional assumption since Pearson (1900) has therefore been to assume that X^* is multivariate normal. Under this assumption the correlation matrix of X^* may be consistently estimated by the polychoric correlation matrix, which is a limited information maximum likelihood estimator (Olsson, 1979a). The polychoric correlation matrix then forms the basis for estimating the parameters of the SEM, typically using diagonally weighted least squares estimation (DWLS). This means that the discrepancy between model-implied and estimated polychoric correlations is minimized, where the weights are obtained from the diagonal of the asymptotic covariance matrix of the estimated polychoric correlations. Robust standard errors and goodness-of-fit testing are also obtained based on the asymptotic covariance matrix (see Monroe (2018) for further details).

We next propose a third limited-information approach to ordinal SEM, with complete technical descriptions given in a later section. Usually, the values of the coordinates of X are taken as a sequence of consecutive integers $1, 2, \ldots, K$, which is not in general the scale used by X^*. This is the approach of cont-ML. We propose to adjust cont-ML as follows. Instead of integer values, we choose the values of X in such a way that the covariance matrix of X approximates that of X^*. To estimate the parameters of the SEM, the usual normal theory ML procedure is then applied to the appropriately encoded observations. The resulting estimator is called *cont-ML-adj*, and it borrows from cat-LS the discretization model in Equation (1) in order to adjust the cont-ML correlations. While polychoric correlations assume that X^* is normal, cont-ML-adj requires only that the marginal distributions of X^* are known, i.e., we know the functions $P(X_k^* \leq x)$ for $k = 1, \ldots, p$. The marginals do not have to be normal. Since the marginals are known, we will show later in the article that the thresholds $(\tau_{k,j})$ can be estimated from data by estimators, say, $(\hat{\tau}_{k,j})$. Consider the first coordinate X_1. From Equation (1), X_1 takes the value of $x_{1,1}$ when $X_1^* \leq \tau_{1,1}$. We therefore choose $x_{1,1}$ somewhere in the interval $< -\infty, \tau_{1,1}]$. In general, we adjust $x_{k,j}$ to a new value $\hat{x}_{k,j}$ which is placed in the interval $[\tau_{k,j-1}, \tau_{k,j}]$. One way to achieve this is to use $\hat{x}_{k,j} = m_k(\hat{\tau}_{k,j-1}, \hat{\tau}_{k,j})$ where $m_k(x, y) = E\left(X_k^* | x \leq X_k^* \leq y\right)$ is the conditional expectation of X_k^* when the value of X_k^* is compatible with the observed value of X in the original encoding. When X_k^* is assumed to be standard normal so that $F_k = \Phi$ with density $\phi = \Phi'$, we have that $m_k(x, y) = [\phi(x) - \phi(y)]/[\Phi(y) - \Phi(x)]$, so that

$$\hat{x}_{k,j} = [\phi(\hat{\tau}_{k,j-1}) - \phi(\hat{\tau}_{k,j})]/[\Phi(\hat{\tau}_{k,j}) - \Phi(\hat{\tau}_{k,j-1})].$$

Figure 1 shows the probability plot for the original and adjusted values, for a $K = 4$ distribution with floor effects. While the original values are equidistant, the adjusted values are unevenly spaced. The original values range from $X = 1$ to $X = 4$, with probabilities $0.54, 0.36, 0.07$, and 0.03, resulting in adjusted values $\hat{X} = -0.74, 0.60, 1.49$, and 2.21, respectively.
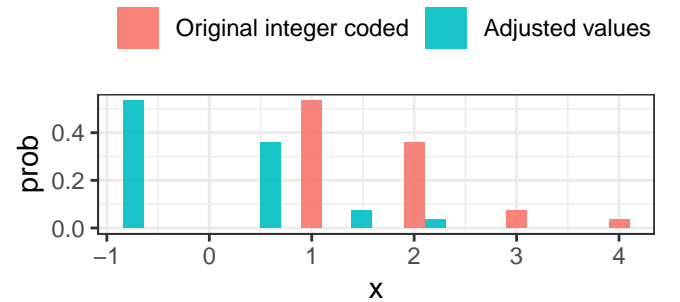

Figure 1. Cont-ML-adj: Adjustment of original values $X = 1, \ldots, 4$.

The following R-code implements the adjustment on the ordinal dataframe ordinal.df:

```
adjust <- function(x){
taus<-c(-Inf,qnorm(cumsum(prop.table(table(x))
    ↪ )))
newvalues <- apply(embed(taus, 2)[, 2:1], 1,
function(x) (dnorm(x[2])-dnorm(x[1]))/(pnorm(x
    ↪ [1])-pnorm(x[2])))
plyr::mapvalues(x, sort(unique(x)), newvalues)
    ↪ }
adjusted.df <- sapply(ordinal.df, adjust)
```

A new simulation technique for non-normal ordinal covariance models

The work of Grønneberg and Foldnes (2019) provided a major impetus for the present study in pointing out that, surprisingly, a large portion of the literature on robustness of ordinal SEM against underlying non-normality has to be re-interpreted. These studies generated ordinal data by discretizing a non-normal continuous vector and the results showed that the effect of non-normality on ordinal SEM was moderate or minor. However, careful examination of the properties of the non-normal continuous vector reveals that the ordinal data produced from it are indistinguishable from ordinal data produced by discretizing a fully normal continuous vector. In brief, thresholds, and marginal distributions are not jointly identified, and the effect of marginals and threshold choice is confounded. These observations have complex implications for the interpretation of simulation studies for ordinal data methods, which we discuss in a separate and upcoming section, see p.7, as well as in the online supplementary material (p.1).

As a response to the shortcoming of the existing simulation methodology, Foldnes and Grønneberg (2019) proposed a new simulation technique for ordinal data in the context of covariance structure analysis, where the continuous marginals are fixed to the standard normal distribution. That is, multivariate ordinal data for covariance structure models are generated by randomly drawing observations from the vector X^* with standard normal marginals, and the ordinal data are obtained through discretizing these observations with user-specified thresholds. Importantly, since we are dealing with factor or structural equation models, the vector X^* must have a user-specified correlation matrix. There are many possible non-normal multivariate distributions for X^* where the univariate marginals are standard normal, and with a given correlation matrix. Foldnes and Grønneberg (2019) proposed to investigate robustness to non-normality by varying the type and degree of non-normality embedded in X^*. Since the marginals are always standard normal, keeping the thresholds fixed is equivalent to keeping the marginal distribution of the observed ordinal variables fixed. That is, we may vary the distribution of X^* from the multivariate normal case to different types of non-normal distributions, without changing either the univariate observed ordinal variables or the correlation matrix of X^*. This approach therefore allows us to disentangle effects of the observed ordinal distribution (level of symmetry, ceiling effects, etc) from effects of the underlying degree and type of non-normality.

To construct non-normal distributions X^* with standard normal marginals and a given correlation matrix, Foldnes and Grønneberg (2019) suggested to use the VITA (VIne-To-Anything) method (Grønneberg & Foldnes, 2017). We here informally present VITA. A more technical introduction to copulas, regular vines and VITA is available in the online supplementary material (p. 6). VITA constructs a multivariate non-normal distribution, with given marginal distributions and correlation matrix, called a regular vine (Bedford & Cooke, 2002). Regular vines is a flexible class of distributions, constructed by assembling bivariate copulas in a hierarchy of tree structures. There are many copulas to choose from, and many possible tree structures. We emphasize that the VITA vectors used in this study all have standard normal marginals, and all match the same fixed target population correlation matrix. To illustrate bivariate copulas, which serve as basic building blocks for VITA distributions, we paired three different bivariate copulas with standard normal marginals and plotted the contours in Figure 2. All three distributions have correlation $\rho = 0.56$ and standard normal marginals. However, the distributions are constructed from different copulas. In Figure 2(a) the copula is the normal copula, which when combined with normal marginals yields the bivariate normal distribution. In Figure 2(b) the copula belongs to the class of Clayton copulas, and the figure presents contours for the distribution obtained by combining this copula with standard normal marginals. Similarly, Figure 2(c) depicts a Joe-type copula paired with standard normal marginals. The two latter distributions differ from the bivariate normal distribution by allowing tail dependencies. As can be seen in Figure 2, relative to the normal copula, the Clayton copula has stronger lower tail dependency, while the Joe copula has stronger upper tail dependency. Also, to illustrate the discretization process, in Figure 2 we have drawn lines representing fixed thresholds of $\tau_{1,1} = -1, \tau_{1,2} = 0$ for X_1^*, and $\tau_{2,1} = -1, \tau_{2,2} = 1$ for X_2^*. The ordinal variables X_1 and X_2 will then have three levels: 1, 2 and 3, for example. The marginal distributions of X_1 and X_2 will be identical across the three distributions in Figure 2. However, the bivariate distributions of the ordinal vector $(X_1, X_2)'$ will not be identical. For instance, the lower left corner of each figure corresponds to the event $X_1 = 1 \cap X_2 = 1$, and inspection of the contour lines suggests that this event is most likely in the Clayton case in Figure 2(b), and least likely in the Joe case in Figure 2(c). In fact, we may calculate these probabilities using the copula package (Hofert, Kojadinovic, Maechler, & Yan, 2013) for R (R Core Team, 2020). The probabilities for $X_1 = 1 \cap X_2 = 1$ are 0.068, 0.095 and 0.046 for the

distributions in Figures 2(a), 2(b) and 2(c), respectively. The normality-based maximum likelihood for the polychoric correlation is maximized at $\rho = 0.56$ only in the normal case in Figure 2(a). For the Clayton distribution in Figure 2(b) the value that maximizes the likelihood is 0.613, while for the Joe distribution in Figure 2(c) the likelihood is maximized at 0.496. This simple illustration suggests that the polychoric estimator is sensitive to the underlying distribution. For the distributions in Figures 2(b) and 2(c), the true correlation is 0.56, but the polychoric estimates are biased, with a relative bias of $+9.5\%$ and -11.4%, respectively.

The full VITA distribution is constructed by specifying bivariate copulas for every pair of variables, where the construction order is dictated by the sequence of trees. Each sequence of trees results in a different distribution, as does each choice of bivariate copulas for the variable pairs, lending the VITA class great flexibility. VITA construction involves numerically solving for a bivariate copula parameter for each pair of variables, rendering VITA calibration more time consuming than other, less flexible methods (e.g., Foldnes & Olsson, 2016; Vale & Maurelli, 1983). VITA calibration is implemented in the R package covsim (Foldnes & Grønneberg, 2020). After VITA calibration, fast simulation is available using the rvinecopulib package (Nagler & Vatter, 2019).

Literature review and the present study

In previous simulation studies only a small number of threshold configurations have been used, i.e., the simulated ordinal data exhibited only a small number of distributional forms. For instance, Rhemtulla et al. (2012) and Li (2016b) employed only three specific threshold configurations (symmetrical, moderately skewed, and severely skewed). That is, for each generated continuous dataset only three marginal distributions of the ordinal datasets were produced. In the present article, our first empirical study reverses this perspective, by generating a large number of ordinal datasets from a limited number of underlying continuous large-sample datasets. We generate a dataset from each of three underlying distributions (multivariate normal and two non-normal distributions) and discretize it using several different thresholds, chosen from a large set of threshold configurations. Then, variation in parameter bias may be studied as a function of variation in threshold sets. Also, in previous studies, with the exception of Rhemtulla et al. (2012), all ordinal variables in the observed ordinal vector shared the same distribution. In other words, ordinal data were generated by applying the same set of thresholds to each continuous variable. Such distributions are rarely encountered in a real-world dataset, where one variable is typically distributed differently than another variable. In the present study no two ordinal variables have the exact same distribution, and thereby better reflecting a real-world dataset. Foldnes and Grøn-
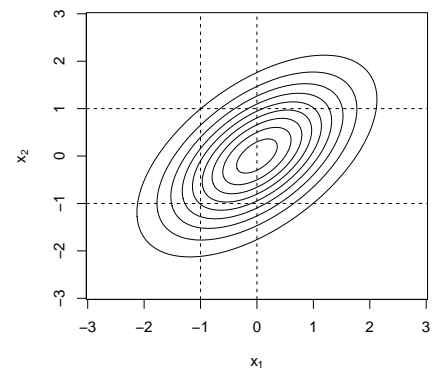
(a) Normal copula
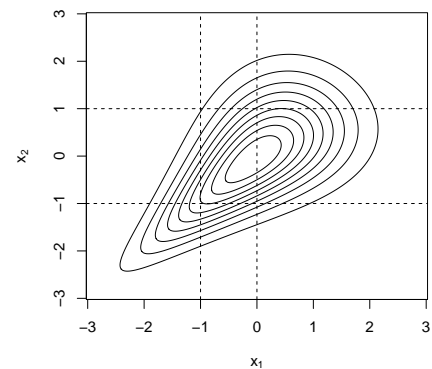
(b) Clayton copula
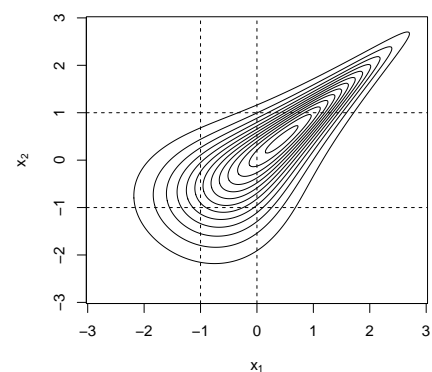
(c) Joe copula

Figure 2. Three bivariate distributions with correlation 0.56 and standard normal marginals.

neberg (2019b) demonstrated, under bivariate non-normality, that the specific choice of thresholds, i.e., of observed ordinal distributions, substantively impacts the bias of the estimated polychoric correlation. In the present study we extend this line of research from the bivariate case of a polychoric correlation to the multivariate case of model parameters, CI coverage, and model fit assessment in a medium-sized SEM. We expect that the variation in bias observed by Foldnes and Grønneberg (2019b) for the polychoric case will propagate into ordinal SEM estimation.

Many simulation studies (e.g., Flora & Curran, 2004; Jin, Luo, & Yang-Wallentin, 2016; Li, 2016a; Moshagen & Musch, 2014; Natesan, 2015; Nestler, 2013; Quiroga, 1994; Rhemtulla et al., 2012; Yang & Green, 2015) have investigated the robustness of ordinal SEM to non-normality. The general consensus formed from these studies is that "estimation of polychoric correlations is robust to modest violations of underlying normality" (Flora & Curran, 2004). For instance, Liu et al. (2017) refer to these studies and suggest that cat-LS factor loading bias may tend to be trivial when the underlying distribution is non-normal (p.501). Recent research (Grønneberg & Foldnes, 2019) identified a problem in the way data were generated in these studies, that weakens and severely complicates claims of robustness of the polychoric estimator. These studies relied on the popular data generation technique of Vale and Maurelli (1983), which is readily available in software packages. Grønneberg and Foldnes (2019) pointed out that, surprisingly, discretizing a non-normal vector obtained using the approach of Vale and Maurelli (VM) is in most cases equivalent to discretizing a normal vector. This is the result of a fundamental lack of identifiability of the underlying continuous vector X^*, combined with the fact that the VM approach in most cases results in data that have a normal copula (Foldnes & Grønneberg, 2015). A careful and detailed discussion of this surprising equivalence is given in the next section.

There are few studies focusing on polychoric correlation estimation bias based on underlying bivariate distributions that are not VM transforms. Monroe (2018) used the bivariate t-distribution, a skew-normal distribution, and a normal mixture and found substantial bias, especially in the two latter distributions. Jin and Yang-Wallentin (2017) used skew-normal and skewed t- distributions and found moderate bias. Likewise, to our knowledge, there is a scarcity of studies on the effect of underlying non-normality on ordinal factor analysis/structural equation modeling (CFA/SEM) based on underlying continuous distributions other than the VM transform. Monroe (2018) employed outliers and the inclusion of a quadratic term to generate multivariate non-normality for a two-factor analytical model using 10 and 20 variables, and reported substantial bias in estimated factor loadings and test statistics when data were generated by using a quadratic term. Maydeu-Olivares (2006) presented a small simulation study using a CFA with 12 variables, and a fixed symmetrical ordinal distribution with 3 levels, where the underlying distribution followed the multivariate t-distribution. Bias in parameter estimates and standard errors was low, with median absolute relative bias at 1% and 4%, respectively, for a sample size of $n = 1000$. Given the scarcity of studies not using the VM method that investigate how ordinal CFA/SEM estimation is affected by underlying non-normality, and the conflicting results among such studies, we deem it important to further investigate this issue.

A majority of simulation studies have employed confirmatory factor models (Hoogland & Boomsma, 1998), and it has therefore been recommended to extend such studies to full SEM models (Beauducel & Herzberg, 2006; Flora & Curran, 2004; Rhemtulla et al., 2012). Hence, in the present study we adopt the full SEM model used by Li (2016b) to assess the performance of various estimators for ordinal SEM under the assumption of underlying normality. This model was constructed from a review of empirical studies using structural equation modeling to be representative from a practical standpoint.

In non-normal conditions where cont-ML, cont-ML-adj, and cat-LS estimation may perform poorly due to estimation bias, there is a need for a test for underlying normality. Before conducting a SEM with ordinal data, a researcher could use such a test to assess whether there is evidence of discretized non-normality in the data. Should the test indicate that the ordinal dataset is unlikely to stem from discretizing a normal vector, cat-LS must be used with caution. Unfortunately, due to the lack of a statistical test for underlying non-normality in popular software packages, the current practice in empirical studies utilizing ordinal SEM is to take the normality of X^* for granted. A true multivariate test for underlying normality was proposed by Maydeu-Olivares (2006), but this test has been largely ignored in the literature and in software packages. Foldnes and Grønneberg (2019b) conducted the first empirical investigation of this test, and also proposed a new parametric bootstrap test. The bootstrap test operates by estimating the polychoric correlation and the thresholds from the original sample. Then, many continuous samples are drawn from a multivariate normal vector whose correlation matrix equals the polychoric correlation matrix. Each continuous sample is further discretized using the estimated thresholds to produce an ordinal data sample. Then, the test statistic of Maydeu-Olivares (2006) is calculated in each ordinal sample. The proportion of times the bootstrapped test statistics exceed the test statistic from the original sample serves as the p-value of the bootstrap test for underlying normality. The test of Maydeu-Olivares (2006) and the bootstrap test were found to maintain Type I error well for dimensionalities less than ten, but only the bootstrap test maintained an adequate Type I error control for larger dimensionalities. The bootstrap test has hitherto been studied only by Foldnes

and Grønneberg (2019b) in a CFA context with at most 15 variables. An aim in the present study was therefore to evaluate the bootstrap test in the context of a medium-sized SEM with 20 variables. It is hoped that such a test may reliably detect conditions where underlying non-normality and the observed ordinal distributional forms render ordinal SEM inference severely biased.

The complications of interpreting ordinal SEM simulation studies

The problem of identification of aspects of the underlying distribution of X^* based on observations from the ordinal-categorical X has recently received attention (Almeida & Mouchart, 2014; Foldnes & Grønneberg, 2019, 2019b; Grønneberg & Foldnes, 2019). Here, we summarize some of the main conclusions from Grønneberg and Foldnes (2019) and Foldnes and Grønneberg (2019), which show that utmost caution must be taken when interpreting simulation studies using the VM simulation method in the context of ordinal variables. We split our discussion into two subsections. First, we outline how the effect of the marginal distribution of X^* and the threshold placement are confounded in simulation designs. Second, we discuss the interpretation of simulation results for cat-LS where the simulation method is to discretize a VM vector. We review the arguments of Grønneberg and Foldnes (2019) which note that in most circumstances, simulating X by discretizing a VM vector X^* is numerically identical to simulating X from an exactly multivariate normal X^*. The only difference is that the thresholds are changed.

Since the ordinal vector X is a censored version of the underlying vector X^*, there will in general be many other vectors, say $\tilde{X}^*$, which when discretized produce exactly the same vector X, or at least a vector with the same distribution. Therefore, one cannot use X to pin-point the exact distribution of X^*. To simulate by discretization of X^* into X therefore also implies that we simulate by discretization of $\tilde{X}^*$ into X. This complicates the interpretation of all simulation studies where ordinal data stem from discretizing continuous variables. The concept of partial identification analysis is briefly introduced as a theoretical solution to this problem in the online supplementary material (p. 1).

Marginal distributions and thresholds are confounded

The marginal distributions of X^* and the thresholds ($\tau_{k,j}$) cannot be separately identified from data. That is, when only observing X, we cannot simultaneously estimate the thresholds and marginal distributions if they are both unknown. Since the values of X^* are not used in generating X except through the thresholds, the marginal scale of X^* is not identified when observing only X. Let $F_k(x) = P(X_k^* \leq x)$ be the k-th marginal distribution of X^*. We assume that F_k is

invertible. Then

$$P(X_k \leq x_{k,j}) = P(X_k^* \leq \tau_{k,j}) = F_k(\tau_{k,j}) \quad (2)$$

Since $P(X_k \leq x_{k,j})$ is a feature of the distribution of X, it can be deduced from data. However, only $F_k(\tau_{k,j})$ is identified, and the marginal F_k and the thresholds $\tau_{k,j}$ are not separately identified. Once the marginal distributions of X^* are specified, we can use this to solve Equation (2) for $\tau_{k,j}$, producing the equation

$$\tau_{k,j} = F_k^{-1}\left(P(X_k \leq x_{k,j})\right). \quad (3)$$

This equation shows that in an empirical setting the thresholds and the marginal distribution are inseparable, unless we know one or the other. Let us further consider the interaction between thresholds and marginal distributions and how this affects simulation when X^* is non-normal. From Equation (1), we have the equivalence

$$X_k = x_{k,j} \iff \tau_{k,j-1} < X_k^* \leq \tau_{k,j}. \quad (4)$$

However, as we will see in examples in the upcoming subsection, there are many random vectors $\tilde{X}^*$ which generate X via Equation (1) using a modified set of thresholds. Indeed, for any sequence $H_1, \ldots, H_p$ of strictly increasing transformations, we have from Equation (4) that

$$X_k = x_{k,j} \iff H_k(\tau_{k,j-1}) < H_k(X_k^*) \leq H_k(\tau_{k,j}).$$

Let $\tilde{\tau}_{k,j} = H_k(\tau_{k,j-1})$ and $\tilde{X}_k^* = H_k(X_k^*)$. Then this equation is

$$X_k = x_{k,j} \iff \tilde{\tau}_{k,j-1} < \tilde{X}_k^* \leq \tilde{\tau}_{k,j}. \quad (5)$$

The distribution of $\tilde{X}^* = (\tilde{X}_1, \ldots, \tilde{X}_k)'$ will be different than the distribution of X^*. Mathematically, the marginals of $\tilde{X}^*$ have changed compared to X^*, but the rest of the distribution, what is formerly known as the copula, has stayed the same. By an appropriate choice of H_k functions, one can show that the marginal distributions of X^* can be transformed into whatever distribution one would like. For example, it is always possible to choose H_k in such a way that $\tilde{X}^*$ is marginally standard normal (Foldnes & Grønneberg, 2019, Proposition 1). Hence, in a simulated sample, there is no single and correct way to define the population thresholds and marginal distributions of X^*. We have seen that simulating X through discretizing X^* yields ordinal observations that are numerically equal to simulating X through discretizing $\tilde{X}^*$. By numerically equal, we mean that the simulated samples of X from X^* and $\tilde{X}^*$ are identical. The simulation of X can therefore be seen as having been generated simultaneously from X^* as well as $\tilde{X}^*$, though with different thresholds.

Let us now return to Equation (3). In our simulation design, outlined in the method section, we vary the marginal distribution of X across other factors, such as the type of non-normality of X^*. Since the marginal distributions of X^* and the threshold values jointly produce the marginal distribution

of X through Equation (3), we decided to fix the marginal distributions of X^* to standard normal, and let the threshold values vary to produce different marginal X distributions. Then the thresholds uniquely define the marginal distribution of X, and vice versa, by

$$P(X_k \leq x_{k,j}) = \Phi(\tau_{k,j}), \qquad \tau_{k,j} = \Phi^{-1}\left[P(X_k \leq x_{k,j})\right].$$

Fixing the marginal distribution of X^* allows us to study the effect of the shape of the ordinal observed marginal distribution of X on ordinal SEM. The choice of standard normal marginals for X^* is natural because this is the scale of the standard methodology: The covariance of a random vector depends on the marginal distributions, yet the marginal distribution of X^* is not identified based on observations from X. Therefore, consistent statistical estimators for $\text{Cov}(X^*)$ must make assumptions on what the marginal distribution of X^* is. The common assumption is that they are standard normal, hence dictating the scale at which the covariance matrix is to be estimated.

The surprising equivalence between ordinal VM simulation and ordinal normal simulation

Here, we consider the consequences of the non-identifiability of the distribution of X^* from the distribution of X in the case of the well-known VM simulation method. Briefly, VM consists of applying third order polynomial transformations to the coordinates of a random vector Z which is multivariate normal with a covariance matrix Σ_Z. In practice, the polynomial transformations are usually either strictly increasing or strictly decreasing. We will limit attention to the strictly increasing case, and refer the reader to Grønneberg and Foldnes (2019) for a full discussion. For each $k = 1, \ldots, p$, VM generates

$$X_k^* = H_k(Z_k), \qquad \text{where } H_K \text{ is strictly increasing.}$$

While the function H_k used in the original VM approach is a third degree polynomial, the upcoming argument rests only on the assumption that H_k is a strictly increasing transformation. The ordinal observations are thereafter generated following Equation (1):

$$X_k = x_{k,j}, \quad \text{when } \tau_{k,j-1} < X_k^* \leq \tau_{k,j} \qquad (6)$$

Since $X_k^* = H_k(Z_k)$ we have that $\tau_{k,j-1} < X_k^* \leq \tau_{k,j}$ is equivalent to

$$\tau_{k,j-1} < H_k(Z_k) \leq \tau_{k,j}. \qquad (7)$$

We next apply the strictly increasing inverse of H_k to the above inequalities, and get

$$X_k = x_{k,j}, \quad \text{when } \tilde{\tau}_{k,j-1} < Z_k \leq \tilde{\tau}_{k,j} \qquad (8)$$

where $\tilde{\tau}_{k,j} = H_k^{-1}(\tau_{k,j})$. This argument can be reversed by starting with discretizing a multivariate normal random vector Z into X, and applying H_k to the inequalities in Equation (8) and arrive at Equation (6).

We conclude that simulating X by discretizing a VM generated X^* is equivalent to simulating X by discretizing an exactly multivariate normal vector Z. The polynomial transformation that the VM method applies to Z is essentially lost under discretization. This has consequences when applying statistical methods to samples from X which assume that X^* is multivariate normal. One such prominent method is the polychoric correlation of Pearson and Pearson (1922) and Olsson (1979a), which is also central to cat-LS. Above we observed that when $X^* = Z$ is exactly multivariate normal, any strictly increasing transformation H_k can be applied to the coordinates of X^*, and the resulting vector, say $\tilde{X}^*$, could have generated X when discretized. The only change is the numerical values of the thresholds. When the transformations are strictly increasing third-order polynomials, the resulting vector $\tilde{X}^*$ has the same distribution as a VM simulated vector. Yet, the polychoric estimator will estimate the correlation matrix of X^* and not $\tilde{X}^*$ which here could have been generated from the VM method. Why is this? The answer lies in the consistency of the polychoric correlation when X^* is normal, which is the case for the fully normal X^* and not for the VM vector $\tilde{X}^*$. The polychoric correlation estimator will estimate the covariance matrix of the version of the underlying vector which is fully normal. Should X be a discretization of a VM vector $\tilde{X}^*$, the polychoric correlation estimator will still implicitly work with the transformed vector Z, because it is exactly normal. Therefore, the polychoric correlation will not estimate the correlation matrix of the simulated X^*, but will instead estimate the correlation matrix of the underlying Z. A more mathematical explanation of this correspondence is given in Grønneberg and Foldnes (2019). Several numerical illustrations of this correspondence, where it is shown that the above observations may be used to exactly predict the results reported in simulation studies on the effect of non-normality on polychoric estimation (e.g., Flora & Curran, 2004), is given in the online appendix of Grønneberg and Foldnes (2019). A particularly unfortunate feature of the VM method is that the covariance matrix of Z and X^* are often very similar, therefore making it more difficult to detect this problem in numerical investigations.

We illustrate the above discussion by considering a typical simulation condition used in ordinal SEM simulation articles. Flora and Curran (2004) included a condition with moderate skewness and moderate kurtosis, where skewness=1.25 and excess kurtosis = 3.75 in all marginal distributions. For the $K = 5$ condition they adopted thresholds from Muthén and Kaplan (1985): $\tau_1 = -1.645, \tau_2 = -0.643, \tau_3 = 0.643$, and $\tau_4 = 1.645$. The VM method first determines the polyno-
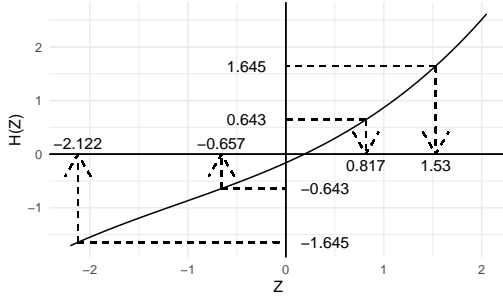
Figure 3. Fleishman polynomial $H(Z) = -0.16 + 0.82Z + 0.16Z^2 + 0.05Z^3$ that yields skewness=1.25 and kurtosis = 3.75. The dashed arrows represent threshold transformations from the Fleishman metric to the standard normal metric.

mial coefficients so that $X_k^* = H(Z_k)$ has the required skewness and kurtosis. Such a polynomial is depicted in Figure 3. Note that the polynomial is strictly increasing, so we may solve for the thresholds, as indicated by dashed arrows. The new thresholds $\tilde{\tau}_1 = -2.122, \tilde{\tau}_2 = -0.657, \tilde{\tau}_3 = 0.817$, and $\tilde{\tau}_4 = 1.53$ are located along the Z axis. This means that discretizing a standard normal variable according to the $\tilde{\tau}_i$, $i = 1, \ldots, 4$ leads to the exact same univariate ordinal distribution as discretizing the Fleishman variable according to the original thresholds, that is, a five-category distribution with proportions $0.017, 0.239, 0.537, 0.144$, and 0.063.

In the simulation design of Flora and Curran (2004) it was required that the correlation between $X_1^* = H(Z_1)$ and $X_2^* = H(Z_2)$ be equal to $\rho_{X^*} = 0.49$. A correlation between Z_1 and Z_2 which makes this possible is $\rho_Z = 0.5084$. We may therefore simulate a VM pair of variables by first simulating from a bivariate normal distribution with correlation $\rho_Z = 0.5084$ and standard marginals, and then apply the polynomial $H(\cdot)$ to both margins. This defines a bivariate distribution whose population (univariate) skewness and kurtosis equals 1.25 and 3.75, respectively, and whose correlation equals $\rho = 0.49$. The crucial insight now is to observe that to discretize a VM dataset using thresholds τ, is exactly the same as discretizing the original data set drawn from (Z_1, Z_2) using the transformed thresholds $\tilde{\tau}$. We exemplify this by simulating a dataset with $n = 500$ using the VM method. This means first drawing a sample with $n = 500$ from a bivariate normal distribution with standard normal marginals and with correlation $\rho_Z = 0.5084$ and then discretizing $X^* = (H_1(Z_1), H_2(Z_2))'$ into X. Figure 4(b) shows a scatter plot of the simulated sample from X^* together with its thresholds. Similarly, Figure 4(a) shows a scatter plot of the simulated sample from the normal data Z, together with the transformed thresholds.

In both plots, we use a color coding for the cells in the resulting 5×5 contingency table and show that the corresponding contingency tables deriving from the data in Fig-
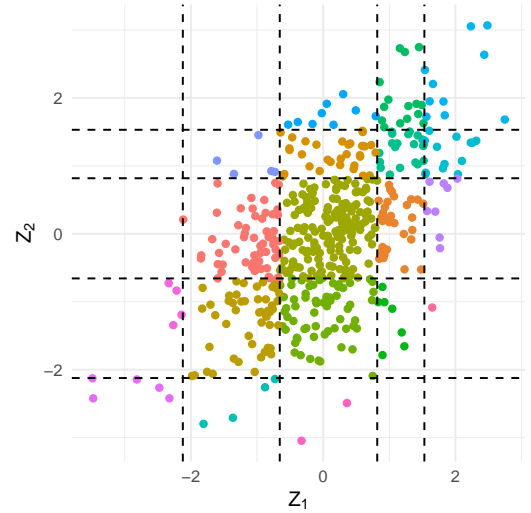
(a) Sample drawn from bivariate normal distribution, with thresholds $\tilde{\tau}_1 = -2.122, \tilde{\tau}_2 = -0.657, \tilde{\tau}_3 = 0.817$, and $\tilde{\tau}_4 = 1.53$.
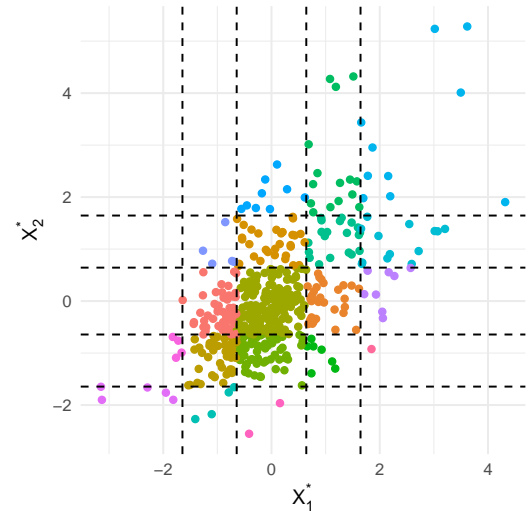
(b) VM transformed sample, with thresholds $\tau_1 = -1.645, \tau_2 = -0.643, \tau_3 = 0.643$, and $\tau_4 = 1.645$.

Figure 4. A $n = 500$ sample drawn from a bivariate normal distribution and then transformed with polynomial $H(Z) = -0.16 + 0.82Z + 0.16Z^2 + 0.05Z^3$.

ures 4(a) and 4(b) are identical. The monotonic VM transformation is such that the discretized data are actually produced by discretizing data drawn from the normal distribution. In large samples drawn from the VM distribution the polychoric estimator will tend to approach 0.5084; refer to the R supplementary material for more details on the computations. The reason for the polychoric correlation not approaching the specified $\rho_X = 0.49$ therefore is not to be found in any underlying non-normality, but is due to the demonstrated equiv-

alence of discretized VM data and discretized normal data. Such an equivalence does not happen when using the VITA simulation method, except in extraordinary cases which may be detected by applying a test for underlying normality to a large VITA sample.

To sum up, studies using the VM method for ordinal data do not, in this sense, violate the normality assumption, but instead do something exactly equivalent to simulating from multivariate normal random vector $\tilde{X}^* = Z$ whose covariance matrix departs slightly from the intended covariance matrix. In many studies, the latent variable model employed was not correctly specified for $\tilde{X}^* = Z$ even when it holds for the VM random vector X^*. This means that several influential simulation studies must be re-interpreted in light of this observation.

Theoretical investigations: What happens when the number of categories increases?

The behavior of ordinal SEM methodology crucially depends on the way correlations are calculated. Here, we develop several fundamental theoretical and mathematical insights into the behavior, as the number of categories increases, of three approaches to calculating correlations in ordinal data. The first two approaches are well-known, namely the Pearson correlation and the polychoric correlation. Our analysis leads us to propose an adjusted version of the Pearson correlation, as well as extending the scope of valid estimation for the polychoric correlation. To the best of our knowledge, all the results presented in this section are new. We give brief numerical illustrations of our results.

We first provide conditions under which the observed vector X is similar to X^* in the sense that $\mathrm{Cov}(X)$ is a good approximation to $\mathrm{Cov}(X^*)$ as $K \to \infty$. This motivates an empirical adjustment to integer-encoded data which approximates the values that lead to consistency. The crucial assumption to achieve this is that the marginal distributions of X^* are known. Since no aspect of the marginal distribution can be estimated through statistical means, this requirement is absolute when aiming at estimating $\mathrm{Cov}(X^*)$ unless we have knowledge of the values of the thresholds.

To our knowledge, our empirical adjustment and its theoretical justification are new, but we emphasize that it is closely related to other methods that convert ordinal variables into continuous variables. These methods use different assumptions and theoretical frameworks, e.g., variants of optimal scaling (De Leeuw, Young, & Takane, 1976), item response theory prediction of summary scores (Harwell & Gatti, 2001), anchoring vignettes (King, Murray, Salomon, & Tandon, 2003), and the Markov Chain Monte Carlo Scaling method of Granberg-Rademacker (2010). The adjustment we suggest has the advantage of being simple and computationally cheap. It does not require numerical optimization and rests only on univariate assumptions on X^*. This is in contrast to more complicated approaches, such as Markov Chain Monte Carlo Scaling, which rests on an assumption of multivariate normality for an underlying variable and is computationally demanding.

Initial numerical experiments indicated that the empirical adjustment method we propose is usually inferior to the polychoric correlation, even when the assumption of full normality of X^* is not true. We found a theoretical answer to why this is the case: the polychoric correlation is consistent as $K \to \infty$, even when the distribution of X^* is non-normal. Again, we must assume that the marginals of X^* are known, though they need not be normal. This observation significantly increases the scope of the polychoric correlation.

Our final contribution in this section is an expression for the population value of the polychoric correlations as $K \to \infty$ if the marginal distributions are supposed to be known, but are misspecified. A consequence of this analysis is the fact that we can consistently estimate the Spearman correlation of X^* as $K \to \infty$, by assuming uniform marginals. We also show that simply applying the regular Spearman correlation to the ordinal data is also consistent as $K \to \infty$.

To make clear that X depends on K, we write $X^{(K)}$, and indicate this relationship in the same manner also for other quantities that depend on K, such as the thresholds.

When is the Pearson correlation consistent with a large number of categories?

Our observations $X^{(K)}$ are not on the interval scale. Their numerical values are therefore strictly speaking arbitrary, as long as the order is properly encoded. The common practice of using cont-ML with integer encoded data is therefore problematic, as the encoding will influence the analysis in an arbitrary manner. As discussed in Bollen (1989, Chapter 4.3), the assumptions underlying SEM are not fulfilled for integer encoded data, and therefore cont-ML need not be consistent, even when the underlying vector X^* is exactly normal.

Instead of encoding data by a sequence of consecutive integers, we propose to encode in such a way that the covariance matrix of $X^{(K)}$ approximates that of X^* as well as possible. We here identify encoding conditions making this possible, which leads us to understand when integer-encoded data can be used as continuous when the number of categories increases to infinity. Also, our analysis suggests a data-driven adjusted encoding which places the ordinal observations on the same scale as the underlying variable X^*

Let us consider why it could be worth studying what happens as $K \to \infty$, although K in practice is typically less than ten. An asymptotic analysis has the advantage of identifying assumptions that have to be fulfilled if consistency is to be reached, even in the best case scenario of having very large K. From the perspective of infinity, it is of little consequence whether K is 7 or 7^{100}, since both numbers are very far from

infinity. The hope in asymptotic analysis is that the resulting approximations are relevant for small K. Whether or not this is the case may be determined by simulations.

Assumption 1. *1. We assume there is an $\epsilon > 0$ such that $\mathrm{E}\,|X_k^*|^{2+\epsilon} < \infty$ for $k = 1, \ldots, p$.*

2. *We have $\tau_{k,j-1}^{(K)} \leq x_{k,j}^{(K)} \leq \tau_{k,j}^{(K)}$ for $k = 1, \ldots, p$ and $j = 1, \ldots, K - 1$.*

3. *We have that $\lim_{K \to \infty} \sup_{1 \leq k \leq p, 2 \leq j \leq K-1} \left[\tau_{k,j}^{(K)} - \tau_{k,j-1}^{(K)} \right] = 0$, and that $\tau_{k,1}^{(K)} \xrightarrow{K \to \infty} F_k^{-1}(0)$ and $\tau_{k,K-1}^{(K)} \xrightarrow{K \to \infty} F_k^{-1}(1)$ for $k = 1, \ldots, p$.*

4. *We have $x_{k,1}^{(K)} F_k(\tau_{k,1}^{(K)}) \xrightarrow{K \to \infty} 0$ and $x_{k,K}^{(K)}[1 - F_k(\tau_{k,K-1}^{(K)})] \xrightarrow{K \to \infty} 0$ for $k = 1, \ldots, p$.*

A technical discussion of these assumptions, a mathematical proof of the following result, and the intuition that led to its formulation is given in the online supplementary material, p. 2.

Proposition 1. *Under Assumption 1 we have $\lim_{K \to \infty} \Sigma_K = \Sigma$, where $\Sigma_K = \mathrm{Cov}(X^{(K)})$ and $\Sigma = \mathrm{Cov}(X^*)$.*

This proposition gives insight into why cont-ML sometimes produces reasonable results. As seen above, the marginals of X^* cannot be deduced from X, but as X^* is usually assumed to be standardized, the scale of the covariance matrix of X is entirely a product of the integer encoding together with the placement of the thresholds, both of which should not affect a statistical analysis. Further, it is clear that as $K \to \infty$, both $|\mathrm{E}\,X_k^{(K)}| \to \infty$ and $\mathrm{Var}\,X_k^{(K)} \to \infty$, hence the covariance matrix of $X^{(K)}$ cannot approximate $\mathrm{Cov}(X)$. But the correlation matrix can approximate $\mathrm{Cov}(X)$, and since it stays invariant under affine transformations (i.e., transformations of the form $x \mapsto ax + b$), cont-ML based on the correlation matrix will be consistent as $K \to \infty$ when the thresholds are evenly spaced. If this is the case, Equation (2) (p. 7) implies that for each k, the CDF of X_k matches that of X_k^*. This means that if the thresholds are evenly spaced, and X_k^* is normal, the bar plot of the ordinal observations will appear normal in shape. The fact that ordinal observations can successfully be treated as continuous when the marginal observations of X are shaped like the normal distribution and X^* in fact is normal, has been observed on several occasions (Babakus, Ferguson, & Jöreskog, 1987; DiStefano, 2002; Johnson & Creech, 1983; Muthén & Kaplan, 1985; Olsson, 1979b; Rhemtulla et al., 2012), but to date, to our knowledge, this seems to have not been explained theoretically. We formalize this into the following corollary.

Assumption 2. *Suppose that for each $1 \leq k \leq p$, the thresholds $\tau_{k,j}^{(K)}$ for $j = 1, 2, \ldots, K - 1$ are evenly spaced and of the form*

$$\tau_{k,j}^{(K)} = z_m^{(K)} + (z_M^{(K)} - z_m^{(K)})(j - 1)/(K - 2)$$

where $z_m^{(K)}$ and $z_M^{(K)}$ are such that Assumption 1 (4) are fulfilled for $x_{k,1}^{(K)} = z_m^{(K)}$ and $x_{k,K}^{(K)} = z_M^{(K)}$.

The proof of the following corollary is given in the online supplementary material (p.4).

Corollary 1. *Suppose Assumption 1 (1), (2) and (3) and Assumption 2 hold. Then, for integer encoded observations, i.e., when $x_{k,j} = j$, we have $\lim_{K \to \infty} \mathrm{Cor}(X_k^{(K)}, X_l^{(K)}) = \mathrm{Cor}(X_k^*, X_l^*)$ for each $1 \leq k, l \leq p$.*

Corollary 1 can in many cases be used to deduce what cont-ML estimates outside Assumption 2. Here, we give an illustration. Suppose there is an invertible function H_k such that

$$\tau_{k,j} = H_k(\tilde{\tau}_{k,j}^{(K)}),$$

where $(\tilde{\tau}_{k,j}^{(K)})$ fulfill Assumption 2. Since

$$\tau_{k,j-1} \leq X_k^* \leq \tau_{k,j}$$
$$\Longleftrightarrow H_k(\tilde{\tau}_{k,j-1}) \leq X_k^* \leq H_k(\tilde{\tau}_{k,j})$$
$$\Longleftrightarrow \tilde{\tau}_{k,j-1} \leq H_k^{-1}(X_k^*) \leq \tilde{\tau}_{k,j}$$

the variable $\tilde{X}^* = (H_k^{-1}(X_1^*), \ldots, H_k^{-1}(X_p^*))'$ is discretize equivalent to X^*, using the thresholds $(\tilde{\tau}_{k,j})$. Therefore, Corollary 1 implies that for integer encoded data, we have

$$\lim_{K \to \infty} \mathrm{Cor}(X_k^{(K)}, X_l^{(K)}) = \mathrm{Cor}(H_k^{-1}(X_k^*), H_l^{-1}(X_l^*)) \quad (9)$$

for each $1 \leq k, l \leq p$. With the exception of rare cases, we have that $\mathrm{Cor}(H_k^{-1}(X_k^*), H_l^{-1}(X_l^*)) \neq \mathrm{Cor}(X_k^*, X_l^*)$, showing that cont-ML will be invalid, even in the best case scenario that $K \to \infty$.

What then, should be done if evenly spaced thresholds are incompatible with the marginal distributions of X^*? We may then use the threshold values to assign an encoding which is not evenly spaced. This leads us to an adjusted Pearson correlation. In practice, the thresholds are unknown, and therefore have to be estimated. A full theoretical analysis of the consequences of empirical threshold estimates is beyond the scope of the present article, but we briefly mention that one may extend Proposition 1 into a generally consistent method based on adjusted observations using an empirical version of Assumption 1 (2) given by

$$\hat{\tau}_{k,j-1} \leq \hat{x}_{k,j} \leq \hat{\tau}_{k,j} \quad (10)$$

where $\hat{\tau}_{k,0} = F_k^{-1}(0)$ which equals $-\infty$ in the normal case, and where $\hat{\tau}_{k,0} = F_k^{-1}(1)$ which equals ∞ in the normal case, and where for $j = 1, \ldots, K - 1$ we have

$$\hat{\tau}_{k,j} = F_k^{-1}(\hat{\pi}_{k,1:j}), \qquad \hat{\pi}_{k,1:j} = \frac{1}{n} \sum_{i=1}^{n} I\{X_{k,i}^{(K)} \leq j\}. \quad (11)$$

Here, $X_{k,i}^{(K)}$ is the k'th marginal of the i'th person in a sample. The estimators $(\hat{\tau}_{k,j})$ are the standard estimator for thresholds, see Pearson and Pearson (1922) and Olsson (1979a).

Ordinal SEM based on such empirically adjusted observations is referred to as cont-ML-adj. The requirement of Equation (10) leaves the concrete choice of $(\hat{x}_{k,j})$ open. A reasonable suggestion is for $k = 1, \ldots, p$ and $j = 1, \ldots, K$ to let

$$\hat{x}_{k,j} = m_k(\hat{\tau}_{k,j-1}, \hat{\tau}_{k,j})$$

where

$$m_k(x, y) = \mathrm{E}\left(X_k^* \middle| x \le X_k^* \le y\right).$$

Note that m_k is a known function since the marginal distributions (F_k) are assumed known. Ignoring sampling error, the value $\hat{x}_{k,j}$ is the conditional expectation of X_k^* when we know that $\hat{\tau}_{k,j-1} \le X_k^* \le \hat{\tau}_{k,j}$. When X_k^* is standard normal so that $F_k = \Phi$ with density $\phi = \Phi'$, we have that

$$m_k(x, y) = [\phi(x) - \phi(y)]/[\Phi(y) - \Phi(x)]$$

(Kotz, Balakrishnan, & Johnson, 2004, Chapter 13, section 10.1). The suggested empirical adjustment procedure is therefore to encode the ordinal observations according to

$$\hat{x}_{k,j} = [\phi(\hat{\tau}_{k,j-1}) - \phi(\hat{\tau}_{k,j})]/[\Phi(\hat{\tau}_{k,j}) - \Phi(\hat{\tau}_{k,j-1})].$$

cat-LS is consistent when the number of categories increases and marginal distributions are correctly specified

In our simulations (Study 1, p.16), it is seen that cont-ML-adj performs overall better than the unadjusted cont-ML, although the improvement is generally moderate. In some of our simulation conditions, cont-ML-adj outperforms cat-LS, and this motivates a more general comparison as a topic for further research. However, in most simulation conditions, cat-LS based on the polychoric estimator of Olsson (1979a); Pearson and Pearson (1922) is the best approach, despite cont-ML-adj being valid even when X^* is not multivariate normal, as $K \to \infty$.

A theoretical explanation for the relatively high performance of the polychoric correlation is that as $K \to \infty$, cat-LS is actually also a consistent method as long as the marginals of X^* are known and standardized. This is suprising, since the polychoric estimator was developed under the assumption that X^* is fully normal. This new observation significantly increases the scope of the polychoric correlation, and may justify its use in empirical studies where full underlying normality is unknown. The following argument can be made mathematically rigorous, but the approximations which underly its formal proof are involved, and we consider it outside the scope of the present article. We will however numerically illustrate its validity.

We consider the thresholds known, as they can always be consistently estimated. This can be seen from Equation (11): Since $\hat{\pi}_{k,1:j}$ consistently estimates $\pi_{k,1:j} = P(X_k^{(K)} \le j) = P(X_k^* \le \tau_{k,j}) = F_k(\tau_{k,j})$ the continuous mapping theorem

implies that $\hat{\tau}_{k,j}$ is a consistent estimator for any set of (continuous) marginals F_k. Therefore,

$$\hat{\tau}_{k,j} = F_k^{-1}(\hat{\pi}_{k,1:j}) \xrightarrow[n\to\infty]{P} F_k^{-1}(\pi_{k,1:j})$$
$$= F_k^{-1}[F_k(\tau_{k,j})] = \tau_{k,j} \qquad (12)$$

where $\xrightarrow[n\to\infty]{P}$ means convergence in probability as the sample size n increases to infinity. Now with a finer and finer net of thresholds, the polychoric correlation matrix, which is the normal theory MLE (maximum likelihood estimator) based on progressively less censored versions of X^* as $K \to \infty$, will approach the normal theory pairwise MLE of the underlying X^* as $K \to \infty$.

The normal theory pairwise MLE based on samples from the underlying X^* when we only estimate the correlation (and use the knowledge that X^* has standardized marginals) will, as the sample size increases, converge towards

$$\rho^\circ(F_{k,l}) = \arg\max_{-1 \le r \le 1} \mathrm{E}_{F_{k,l}} \log \phi_2(X_k^*, X_l^*; r),$$

where $s^\circ = \arg\max_s H(s)$ means the arguments that maximize $H(s)$ are contained in s°, i.e., the set (or singleton when the maximum is unique) such that $s \in s^\circ$ implies $H(s) \ge H(\tilde{s})$ for all $\tilde{s}$. Also, $F_{k,l}$ is the bivariate CDF of $(X_k^*, X_l^*)'$, and $\phi_2(\cdot, \cdot; r)$ is the density of a bivariate normal Z with standardized marginals and correlation r. The proof of the following lemma is given in the online supplementary material (p.4).

Lemma 1. *If F_k, F_l are standardized, then $\rho^\circ(F_{k,l}) = \mathrm{Cor}(X_k^*, X_l^*)$.*

Therefore, the polychoric correlation is consistent when $K \to \infty$ as long as the standardized marginals are correctly specified.

Suppose now that the true marginal distributions are not known, but are erroneously assumed to equal F_k for $k = 1, \ldots p$, when in fact they are equal to G_k for $k = 1, \ldots, p$. Suppose F_k, G_k are continuous and strictly increasing for $k = 1, 2, \ldots, d$, and that F_k is standardized. Following the argument that leads to Equation (12), we see that we then have $\hat{\tau}_{k,j} \xrightarrow[n\to\infty]{P} F_k^{-1} G_k(\tau_{k,j})$. Since the marginals are arbitrary, we can choose them in a manner explained in the online supplementary material (p. 5) so that the thresholds in fact are consistent. We are then estimating a covariance matrix with elements $\tilde{\sigma}_{k,l} = \mathrm{Cor}(F_k^{-1} G_k(X_k^*), F_l^{-1} G_l(X_l^*))$ as $K \to \infty$.

As an example application, let us consider now how to use this to estimate the Spearman correlation $\rho_{k,l}^S$ of X_k^*, X_l^*, which is the Pearson correlation of the copula of X_k^*, X_l^* (Nelsen, 2007, Theorem 5.1.6). That is, $\rho_{k,l}^S = \mathrm{Cor}(F_k(X_k^*), F_l(X_l^*))$. In order to reach $\rho_{k,l}^S$ by using polychoric correlations, the following considerations are necessary. Since $(U_k, U_l)' = (F_k(X_k^*), F_l(X_l^*))'$ has uniform marginals on $[0, 1]$, which are not standardized, we deduce the standardized versions of

these distributions. Since $E\,U_k = 1/2$ and $\sqrt{\text{Var}\,U_k} = 1/\sqrt{12} = 1/(2\sqrt{3})$, we have that $\tilde{U}_k = (U_k - 1/2)(2\sqrt{3})$ is standardized, and uniform on $[-\sqrt{3}, \sqrt{3}]$. Therefore, in order to reach the Spearman correlation, we estimate the thresholds using the assumption that the marginal distributions of X_k^* are uniform on $[-\sqrt{3}, \sqrt{3}]$, so that their CDF is

$$F_k(x) = \left(\frac{x-a}{b-a}\right)I\{a \le x \le b\} + I\{x > b\}, \quad a = -b, b = \sqrt{3}.$$

Numerical illustrations of this technique are found after the next sub-section. Note that a simple extension of cont-ML-adj is to use standardized uniform marginals, which leads to another estimator of Spearman's correlation.

The Spearman correlation is consistent as the number of categories increases

The empirical Spearman correlation based on the ordinal observations is also consistently estimating the Spearman correlation of X^* as $K \to \infty$. Since this is far simpler than computing the polychoric correlation with standardized uniform marginals, the above discussion is merely an illustration of our arguments. Note that applying both the empirical Spearman correlation and the polychoric correlation based on standardized uniform marginals will consistently reach the Spearman correlation of X^* as the sample size and K increase, and also note that this is achieved without assumptions on the marginal distribution or any other feature of the distribution of X^*. This is because the Spearman correlation is defined as the correlation of the distribution of X^* when the marginals are transformed to a specific distribution, namely the uniform distribution. See the online supplementary material (p. 4) for a proof of the following proposition.

Proposition 2. *Suppose Assumption 1 holds. Assume in addition that X^* has continuous marginal distributions. Then the population Spearman's correlation based on $X^{(K)}$ consistently estimates the population Spearman's correlation based on X^*.*

Numerical illustrations

Here, we provide four illustrations of consistency, or lack thereof, as the number of categories increases. Reproduction files are found in the supplementary material of this article.

Illustration 1 demonstrates the result in Corollary 1 that the Pearson correlation is consistent as K increases, provided the thresholds are evenly distributed. A large sample with $n = 10^7$ was generated from a distribution with a Clayton copula (C) and normal marginals, and from a fully normal distribution (N), both with a correlation of $\rho = 0.7$. The samples are then discretized according to Equation (1) with evenly spaced thresholds varying between -3.2 and 3.2, the number of which is equal to $K - 1$ (producing K categories). For $K = 1$ we used $\tau_{k,1} = 0$. For $K = 2$ we used

$\tau_{k,1} = -3.2/2, \tau_{k,2} = 3.2/2$. For the remaining values of K, we used $\tau_{k,j} = z_m^{(K)} + (z_M^{(K)} - z_m^{(K)})(j-1)/(K-2)$ with $z_m = -3.2, z_M = 3.2$. Based on the discretized datasets, we calculated the Pearson correlation for $X^{(K)}$ for each $K = 2, \ldots, 19$. Since n is large, the computed values are close to the population values. The resulting values are plotted in Figure 5. We clearly see the consistency as K increases, and the approximation is very good around $K = 16$. Whether X^* has a normal or Clayton copula is of little consequence in this illustration.
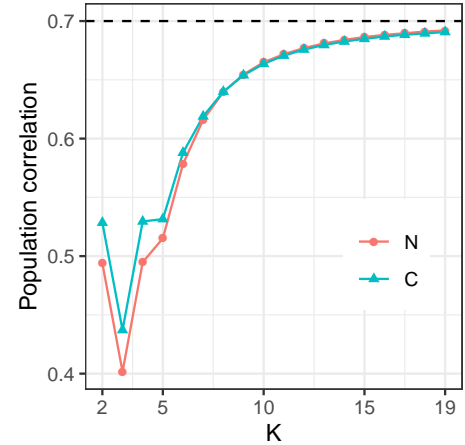
Figure 5. Population values for the Pearson correlation of integer-coded data from evenly spaced thresholds. Dashed horizontal line at 0.7 represents true correlation of X^*. $K=$ number of categories. Normal=underlying bivariate normal distribution. Clayton=Underlying distribution with normal marginals and Clayton copula.

Illustration 2 demonstrates Equation (9), for the same two large datasets N and C, this time discretized only once, with unevenly spaced thresholds. We let $K = 51$ and use the following thresholds: $\tau_{k,j} = \tilde{\tau}_{k,j}^3$ where $\tilde{\tau}_{k,j} = x_m^{(K)} + (x_M^{(K)} - x_m^{(K)})(j-1)/(K-2)$ for $x_m = -(3.2)^{1/3}, x_m = (3.2)^{1/3}$. In Table 1 (p.13), we see that the Pearson correlation does not reach $\rho = 0.7$ even for $K = 51$. As dictated from Equation (9), it instead reaches $\text{Cor}((X_1^*)^{1/3}, (X_2^*)^{1/3})$. In contrast, both the Pearson-adjusted correlation and the polychoric correlation are consistent as $K \to \infty$.

Table 1
The inconsistency of the Pearson correlation with unevenly spaced thresholds.

	True ρ	Pearson	Pearson-adj	polychoric	$\text{Cor}\left((X_1^*)^{1/3}, (X_2^*)^{1/3}\right)$
Normal	0.70	0.63	0.70	0.70	0.63
Clayton	0.70	0.65	0.70	0.70	0.65

Illustration 3 also demonstrates the consistency of the adjusted Pearson and the polychoric correlations with unevenly spaced threshold generated in a similar way as was done in

Illustration 2. This time we investigate the rate of convergence, as K increases, see Figure 6. Under normality, it is seen that the polychoric correlation is unbiased for all K. Under the Clayton copula, the polychoric correlation is biased for low values of K, but approaches the true underlying correlation rather quickly, being virtually unbiased for $K = 10$. We also see the consistency of the adjusted Pearson correlation for both underlying distributions, as K increases. In contrast, the standard Pearson correlation converges at a fast rate to a value not close to $\rho = 0.7$, but converges instead to $\mathrm{Cor}\left((X_1^*)^{1/3}, (X_2^*)^{1/3}\right)$ as dictated by theory.

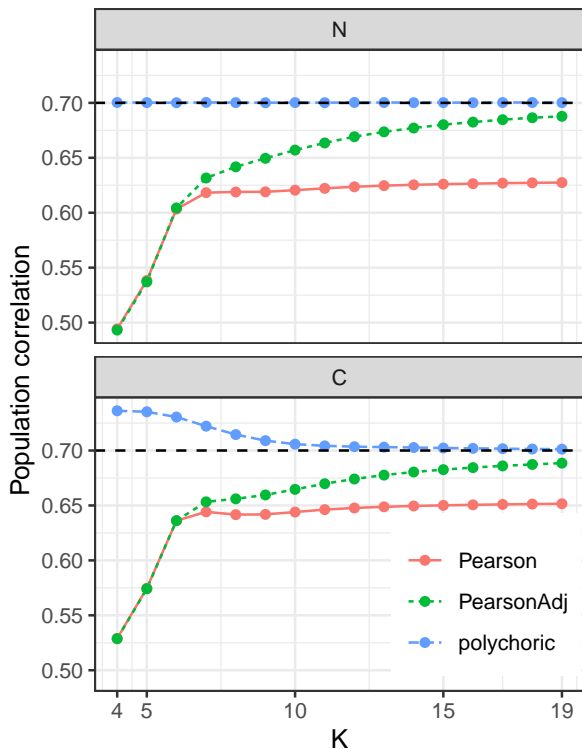*Figure 6.* Population correlation values for data stemming from unevenly spaced thresholds. Dashed horizontal line at 0.7 represents true correlation of X^*. Pearson = Pearson correlation. Pearson-adj= The Pearson correlation of adjusted values. polychoric= polychoric correlation. K= number of categories. N and C refer to underlying distributions with normal and clayton copula, respectively.

Illustration 4 demonstrates techniques developed for the polychoric correlation as outlined after Lemma 1. We identify a setting where the Pearson correlation and Spearman correlation differ considerably: $X_1^* = Z, X_2^* = Z^3 + \epsilon$ where Z, ϵ are independent, and both are standard normal. We simulated a sample from (X_1^*, X_2^*) of size $n = 10^7$. Since this is a large sample, the estimates based on this sample are close to population quantities. We calculated $(U_1, U_2)' = (F_1(X_1^*), F_2(X_2^*))'$ where F_1, F_2 are the CDFs

of X_1^*, X_2^* respectively. The correlation of $(U_1, U_2)'$ is the Spearman correlation of $(X_1^*, X_2^*)'$ and equals 0.714. We then discretized $(U_1, U_2)'$ with 60 evenly spaced thresholds on the interval $[0.001, 0.999]$ and calculated the standard polychoric correlation, assuming normal marginals, which equaled 0.804. This polychoric correlation does not estimate $\mathrm{Cor}(U_1, U_2)$ nor $\mathrm{Cor}(X_1^*, X_2^*)$ but instead it estimates $\mathrm{Cor}(\Phi^{-1}(U_1), \Phi^{-1}(U_2))$ which we calculated as equal to 0.807. We then calculated the polychoric correlation using standardized uniform marginals, which equaled 0.714, which consistently approximates $\mathrm{Cor}(U_1, U_2)$. Finally, we calculated the Pearson correlation directly on the integer-encoded data, which according to theory is consistent. Indeed, we get that this estimate equals 0.713, close to the true 0.714. Note that $\mathrm{Cor}(X_1^*, X_2^*)$, which equaled 0.750, is not estimated by any of the procedures we considered.

Finally we mention that in the supplementary material we also provide R code that illustrates the general consistency of the Spearman correlation. For both the evenly and non-evenly spaced threshold configurations, Spearman's correlation based on the ordinal observations is demonstrated to be close to the true population correlation of X^*.

Method

Each of the studies presented here is based on the same SEM model and the same underlying multivariate continuous distributions. The threshold sets are of three types, according to whether they produce distributions with floor effects, ceiling effects or with symmetry. In each type, fifty threshold sets were generated. The large amount of threshold sets guarantees that our findings are not overly dependent upon specific ordinal distributions, and therefore allows us to interpret results as smoothed across fifty thresholds all sharing the same basic type, e.g., across fifty distributions all having floor effects.

Study 1 is conducted at the population level, where we discretize the three distributions using a large number of threshold sets. The outcome variables are the relative bias of Pearson, Pearson adjusted, and polychoric correlations and cont-ML, cont-ML-adj, and cat-LS SEM model estimates across threshold sets and underlying distributions.

Study 2 is a simulation study where we draw and discretize medium-sized samples from the three underlying distributions, using the same threshold sets as in Study 1. The outcome variables are CI coverage rates for model parameters and rejection rates for a test of correct model specification.

Study 3 evaluates the performance of the bootstrap test at small to medium sample sizes, and is limited to nine threshold sets.

Population model

We follow Li (2016b), who conducted a thorough analysis of 36 empirical studies that used ordinal structural equation modeling, and found that the median number of factors was five and that the median number of variables was 18. Li (2016b) therefore emulated a realistic empirical setting by constructing a five-factor structural equation model. This model was deemed representative of medium-sized SEM model specifications encountered in applications, and we present its structural part in Figure 7. Not depicted in the path diagram are the observed ordinal variables, four for each factor, so that the simulated data had $p = 20$ dimensions. The factor loadings for the continuous X^* variables were set to $0.8, 0.7, 0.6$ and 0.5 within each factor, and the corresponding residual variances were chosen so that each X_k^*, $k = 1, \ldots, 20$, had unit variance. The interfactor correlation was $\phi = 0.3$ for the two exogenous factors, each of which had unit variance. The residuals ζ_1, ζ_2 and ζ_3 had variances so that the path values in Figure 7 were standardized regression coefficients. The model-implied correlation matrix contains mostly moderate correlations, see Table 2.

Figure 7. Population model. Only the structural model among five latent variables is shown, and for clarity indicator variables are not depicted.
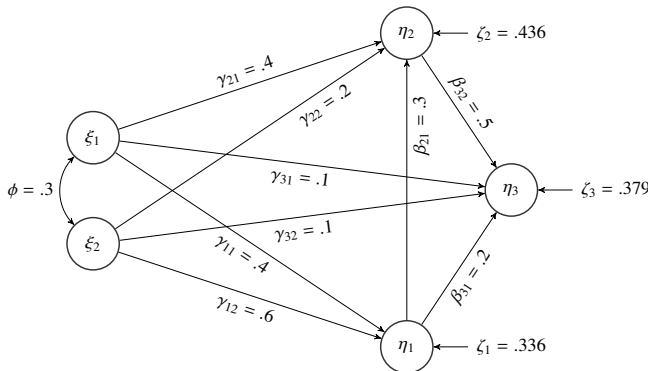
Distributions for the underlying continuous vector

We considered three types of underlying continuous distributions: the multivariate normal (N) and two non-normal VITA distributions. In the first VITA distribution (C) we used Clayton bivariate copulas to construct the regular vine, while in the second distribution (J) we used Joe bivariate copulas. The tree structure was identical in both VITA distributions. The N, C, and J distributions were constructed in a way to have the correlation structure given in Table 2. That is, the SEM model is correctly specified with respect to all three distributions. Also, each of the N, C, and J distributions has standard normal univariate marginal distributions. Hence, it

is the multivariate dependence structures inherent in the distributions that differ from one another.

Sample size

In Study 1 we generated three large ($n = 10^5$) samples from distributions N, J, and C. The large sample size reduces sampling fluctuations so that this study is conducted close to the population level. Understanding parameter bias at the population level is a prerequisite before investigating sampling properties. Given that for fixed K, cont-ML and cont-ML-adj are generally not consistent, and cat-LS is not consistent under the VITA distributions C and J, population-level parameter bias does not disappear with increasing sample size. Using a large dataset allows us to calculate the bias with virtually no confounding due to sampling fluctuation. In Study 2 we included medium to large sample sizes of $n = 500$ and $n = 1000$, while in Study 3 a range from small to medium sample sizes, $n = 100, 300, 500$, was employed.

Distributions of the observed variables

Li (2016b) reported that in empirical research the most common number of levels in ordinal data are 4, 5 and 7, and we included these levels in our study. For each K, we considered three general types of distributions for the ordinal marginal distributions in X, which we refer to as Symmetrical, Floor, and Ceiling. Symmetrical implies that the probability distribution of X is symmetric with highest values in the middle of the range $1, \ldots, K$. Floor refers to a floor effect, that is, the highest probability occurs for the lowest levels of X and decreases with increasing levels. Ceiling refers to a ceiling effect, that is, the lowest probability occurs for the lowest level of X and increases with increasing levels. We note that distributions with floor or ceiling effects are routinely encountered in the social sciences.

A single univariate ordinal distribution is defined by a set of thresholds. For each of the twenty variables, we employed a different set of thresholds, so that the observed ordinal vector had marginals with different distributions, as we believe this is representative of real-world datasets. We refer to the collection of twenty different threshold sets as a threshold *configuration*. To avoid our results being contingent upon a small number of specific threshold values, we randomly generated fifty threshold configurations in each class of symmetrical, floor, and ceiling distributional forms. Our results concerning these three distributional forms may therefore be seen as smoothed over fifty specific distributions belonging to each form. The generation of threshold configurations using R is available in the supplementary material. In Figure 8 we have plotted, for $K = 4$, the ordinal distributions associated with three threshold configurations. The upper, middle, and lower panels in Figure 8 refer to ceiling, floor, and symmetrical configuration, respectively. The columns in the figure correspond to the twenty ordinal variables.

Table 2

Correlation matrix Σ implied by the population model.

	X_1^*	X_2^*	X_3^*	X_4^*	X_5^*	X_6^*	X_7^*	X_8^*	Y_1^*	Y_2^*	Y_3^*	Y_4^*	Y_5^*	Y_6^*	Y_7^*	Y_8^*	Y_9^*	Y_{10}^*	Y_{11}^*	Y_{12}^*
X_1^*	1.00																			
X_2^*	0.56	1.00																		
X_3^*	0.48	0.42	1.00																	
X_4^*	0.40	0.35	0.30	1.00																
X_5^*	0.19	0.17	0.14	0.12	1.00															
X_6^*	0.17	0.15	0.13	0.10	0.56	1.00														
X_7^*	0.14	0.13	0.11	0.09	0.48	0.42	1.00													
X_8^*	0.12	0.10	0.09	0.07	0.40	0.35	0.30	1.00												
Y_1^*	0.37	0.32	0.28	0.23	0.46	0.40	0.35	0.29	1.00											
Y_2^*	0.32	0.28	0.24	0.20	0.40	0.35	0.30	0.25	0.56	1.00										
Y_3^*	0.28	0.24	0.21	0.17	0.35	0.30	0.26	0.22	0.48	0.42	1.00									
Y_4^*	0.23	0.20	0.17	0.15	0.29	0.25	0.22	0.18	0.40	0.35	0.30	1.00								
Y_5^*	0.41	0.36	0.30	0.25	0.34	0.30	0.26	0.21	0.43	0.38	0.32	0.27	1.00							
Y_6^*	0.36	0.31	0.27	0.22	0.30	0.26	0.23	0.19	0.38	0.33	0.28	0.24	0.56	1.00						
Y_7^*	0.30	0.27	0.23	0.19	0.26	0.23	0.19	0.16	0.32	0.28	0.24	0.20	0.48	0.42	1.00					
Y_8^*	0.25	0.22	0.19	0.16	0.21	0.19	0.16	0.13	0.27	0.24	0.20	0.17	0.40	0.35	0.30	1.00				
Y_9^*	0.36	0.32	0.27	0.23	0.35	0.30	0.26	0.22	0.43	0.37	0.32	0.27	0.48	0.42	0.36	0.30	1.00			
Y_{10}^*	0.32	0.28	0.24	0.20	0.30	0.27	0.23	0.19	0.37	0.33	0.28	0.23	0.42	0.37	0.32	0.26	0.56	1.00		
Y_{11}^*	0.27	0.24	0.20	0.17	0.26	0.23	0.20	0.16	0.32	0.28	0.24	0.20	0.36	0.32	0.27	0.23	0.48	0.42	1.00	
Y_{12}^*	0.23	0.20	0.17	0.14	0.22	0.19	0.16	0.14	0.27	0.23	0.20	0.17	0.30	0.26	0.23	0.19	0.40	0.35	0.30	1.00

Note. Values are rounded to two decimal places. $X_1^* - X_4^*$ are indicators for ξ_1. $X_5^* - X_8^*$ are indicators for ξ_2. $Y_1^* - Y_4^*$ are indicators for η_1. $Y_5^* - Y_8^*$ are indicators for η_2. $Y_9^* - Y_{12}^*$ are indicators for η_3.

Data generation and model estimation

In each of the finite sample conditions in Studies 2 and 3 we generated 1000 datasets. In Study 3, we used 1000 bootstrap replications for each sample. The simulations in Studies 2 and 3 were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway. The samples were generated in the R computing environment (R Core Team, 2020), with the help of the VineCopula package (Schepsmeier et al., 2018). cont-ML and cont-ML-adj were computed based on the covariance matrix using fully standardized solutions. Polychoric correlation estimation was performed with the sirt package (Robitzsch, 2019). Model estimation and testing was achieved with lavaan (Rosseel, 2012), as detailed in the supplementary material. The models were identified by setting the variance of the latent variables to unity. The model fit statistic for all three estimators was the scaled-and-shifted statistic of Asparouhov and Muthén (2010). The bootstrap test of underlying normality was implemented using the R package discnorm (Foldnes & Grønneberg, 2020b). Computer code is provided in the supplementary material.

Evaluation criteria

In Studies 1 and 2 the outcome variable for a statistic $\hat{\theta}$ was relative bias (RB), defined as the mean bias across all replications in the design condition: $RB = 100 \cdot \overline{(\hat{\theta} - \theta_0)}/\theta_0$, where θ_0 denotes the true population value of the parameter. Note that while in Study 2 the replications are random samples from multivariate distributions, the replications in Study 1 consists of 50 configuration sets. To avoid canceling of negative and positive bias within a condition, the mean absolute relative bias (MARB) was also employed. CI coverage is defined as the proportion of times the 95% CI contained the true population value. For the statistical tests in Study 2 and Study 3 the evaluation criterion was the rejection rate of the test at the 5% level of significance.

Performance criteria for estimation and inference outcomes are as follows. We deem acceptable relative bias smaller than 5%, 95 % CI coverage rates between 92% and 98%, and rejection rates between 2.5% and 7.5% (Bradley, 1978).

Results

Study 1: Bias in parameter estimates at the population level

Relative bias at the population level was calculated for the 190 correlations. Our design crosses three underlying distributions with three levels of *K* and three ordinal distributional forms, yielding 27 conditions. In each of these conditions, MARB was calculated from 190 relative bias values across 50 threshold configurations, yielding Table 3.

Compared to the standard Pearson correlation, the adjusted Pearson correlation managed to reduce bias in 25 of a total of 27 conditions. The overall mean MARB reduction of the adjusted Pearson correlation relative to the original
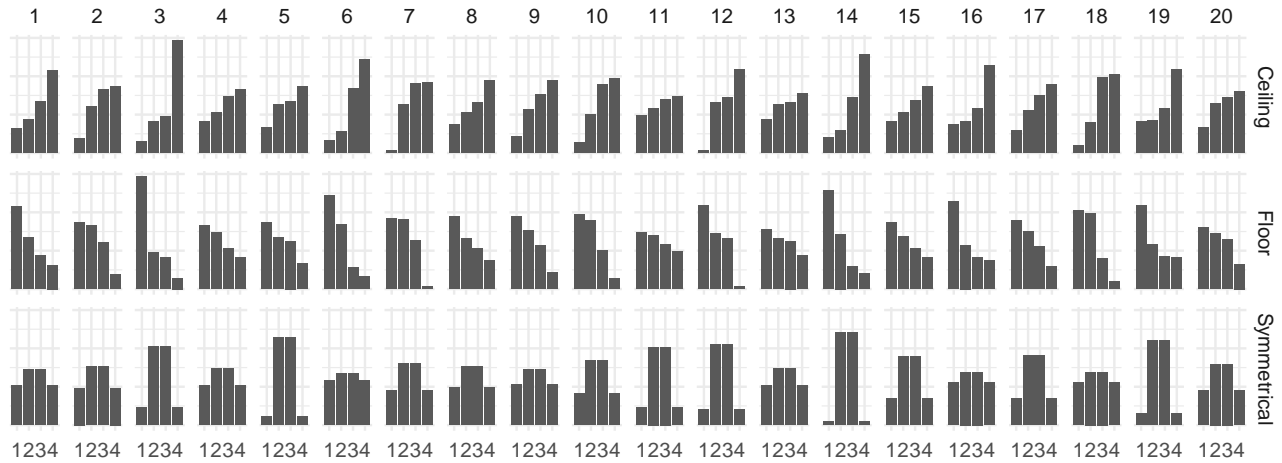
Figure 8. Three configurations of thresholds for $K = 4$. The upper, middle, and lower panels contain distributions of twenty ordinal variables with ceiling, floor, and symmetrical distributions, respectively.

Pearson correlation was 14%. The improvement offered by the adjusted Pearson correlation was most notable for $K = 7$: The overall percentage improvement for $K = 4, 5$, and 7 was, respectively, 0%, 12%, and 31%. Given that the adjustment to the Pearson correlation was found overall to reduce bias relative to the original Pearson correlation, we do not comment further on the latter.

The adjusted Pearson and the polychoric correlation achieved acceptable bias in 6 and 14 of the 27 conditions, respectively. The overall MARB for the adjusted Pearson and the polychoric correlation was 14.1% and 7.8%, respectively. Overall, the polychoric correlation was therefore less biased at the population level than the adjusted Pearson correlation. However, in conditions with underlying non-normality and asymmetrical ordinal distributions, the polychoric correlation did not always outperform the adjusted correlation. Specifically, for Joe VITA discretized into floor distributions, and for Clayton VITA discretized into ceiling distributions, the adjusted Pearson correlation was less biased than the polychoric correlation.

Let us now turn to parameter bias in the population for the SEM model depicted in Figure 7, which we first illustrate for two parameters in the model. Similar analyses for all model parameters are found in the online supplementary material. Figure 9 contains boxplots for bias in the structural regression coefficient γ_{21}, relating ξ_1 to η_2. With underlying normality, cont-ML relative bias was acceptable in 61% of the 450 (three levels of K, three distributional forms and 50 threshold configurations) ordinal distributions. For cont-ML-adj and cat-LS the corresponding rate of acceptable bias were 63% and 99%, respectively. Under the Joe VITA the acceptance rates were 55%, 65%, and 60% for cont-ML, cont-ML-adj, and cat-LS, respectively, while the respective per-

centages for the Clayton VITA were 60%, 67%, and 71%. Under non-normality, the distributional shape of the ordinal data greatly affects the estimators. Categorization of the Joe VITA into ceiling effects produces severe negative bias in all three estimators. The Joe VITA discretized into floor effect distributions on the other hand does not produce bias in the majority of threshold configurations. For the Clayton VITA; it is seen that floor effect distributions lead to pronounced negative bias for all three estimators. There is also considerable bias variation within each distributional form, especially at $K = 4$. For instance, for the 50 $K = 4$ threshold configurations with floor effects, cat-LS relative bias ranged from +10.4% to -7.4%. Variation across threshold configurations decreased with increasing K, with the mean range in cont-ML relative bias decreasing from 35% to 27% for $K = 4$ to $K = 7$, respectively. For cont-ML-adj and cat-LS, relative bias range for $K = 4$ were 31% and 33%, respectively, decreasing to 17% and 16% for $K = 7$, respectively. Also, as K increases, relative bias tends to decrease. Overall, for $K = 4$ and $K = 7$, MARB values for cont-ML, cont-ML-adj and cat-LS were (6.3%, 4.6%), (5.8%, 3.3%), and (4.6%, 2.8%), respectively.

Figure 10 presents relative bias for the factor loading λ_{11}, relating X_1^* to ξ_1. As expected, cat-LS under the ideal condition of multivariate normality is unbiased. However, in striking contrast to the above findings for structural coefficient γ_{21}, cont-ML and cont-ML-adj are biased under normality, even with symmetrical ordinal distributions. The bias in λ_{11} was acceptable in only 43% of the 450 discretizations obtained by aggregating over K and ordinal distributions. In comparison, for cont-ML-adj and cat-LS the percentage of acceptable relative bias was 46% and 100%, respectively. In other words, under normality only cat-LS was unbiased.
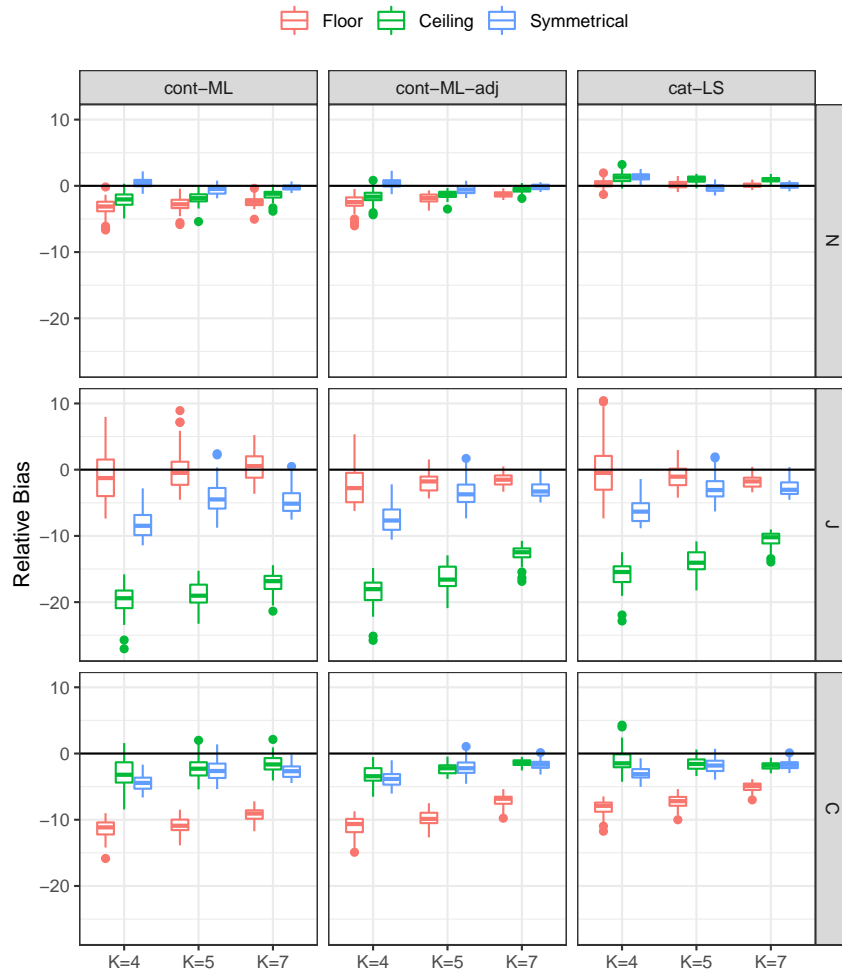
Figure 9. Study 1. Relative bias for regression parameter γ_{21}. Each boxplot is based on 50 threshold configurations. Floor, Ceiling, and Symmetrical refer to ordinal distributions with floor effect, ceiling effect, and symmetry, respectively. cont-ML, cont-ML-adj, and cat-LS refer to estimation based on Pearson, Pearson adjusted, and polychoric correlations, respectively. K is the number of categories. N, J, and C, refer to normal, Joe VITA, and Clayton VITA underlying distributions.

With underlying non-normality all three estimators exhibited bias. Under the Joe VITA, cont-ML, cont-ML-adj and cat-LS bias was acceptable in 54%, 63%, and 66%, respectively, of the 450 threshold configurations. For the Clayton VITA, the corresponding acceptance rates were 55%, 59%, and 68%. We note that there is one specific symmetrical configuration for $K = 5$ where the bias is about -34% for cont-ML and cont-ML-adj. Similar to findings for γ_{21}, we observed that the extent and direction of bias was dictated by the interaction of underlying distribution and ordinal distributional form. cat-LS with underlying Joe VITA distributions is positively biased, often above +10%, when the ordinal variables exhibit floor effects. In the same condition, bias become negative for ceiling effects. For underlying Clayton VITA, the situation is reversed, and ordinal distributions with ceiling effects yield positive bias, while ordinal distributions with

floor effects is associated with negative bias. Overall, bias is reduced with increasing number of categories.

We now summarize bias across groups of model parameters. It is natural to assemble the SEM parameters into four groups: The first group consists of the 8 factor loadings for the exogenous variables ξ_1 and ξ_2, and the second group consists of the twelve loadings for the endogenous variables η_1, η_2, and η_3. The third group consists of the six structural coefficients γ_{ij} relating the η-variables with the ξ-variables. Finally, the fourth group consists of the three structural coefficients relating the η variables. We refer to these four groups respectively as $\lambda_{ex}, \lambda_{en}, \gamma$, and β. We next report the mean absolute relative bias calculated across parameters in each of the four parameter groups. First, let us investigate the effect of K, the number of categories. Table 4 shows that estimation bias consistently decreases with increasing number
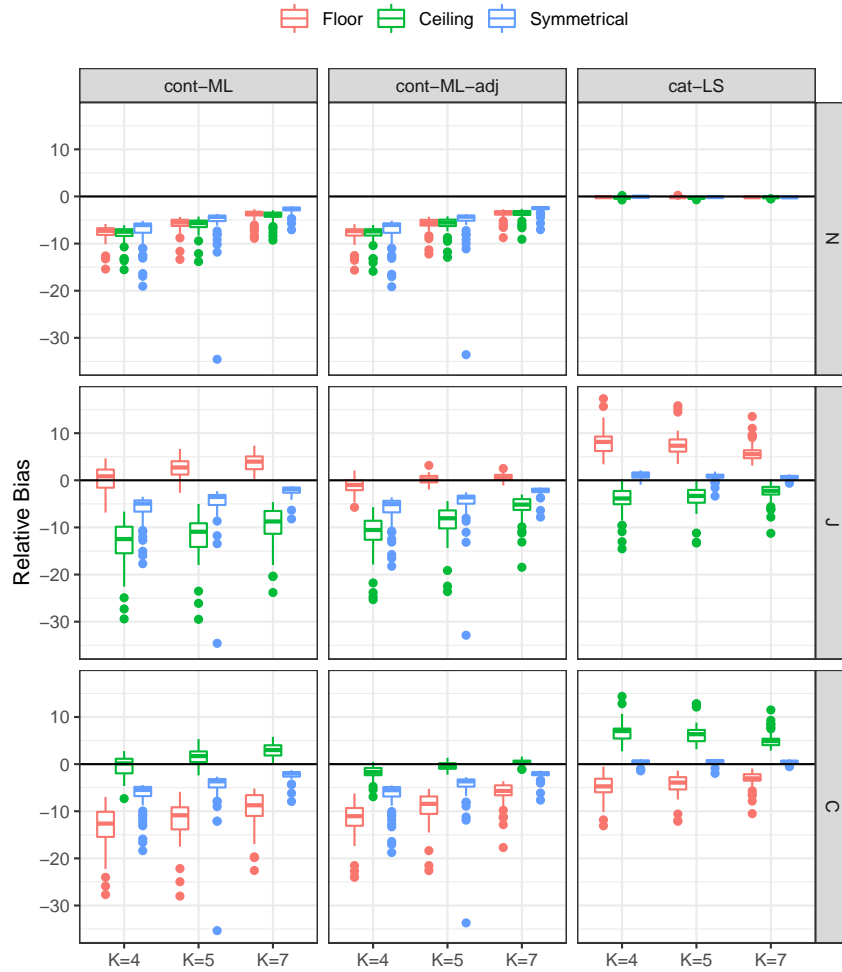
Figure 10. Study 1. Relative bias for factor loading associated with X_1^* and ξ_1. Each boxplot is based on 50 threshold configurations. Floor, Ceiling, and Symmetrical refer to ordinal distributions with floor effect, ceiling effect, and symmetry, respectively. cont-ML, cont-ML-adj, and cat-LS refer to estimation based on Pearson, Pearson adjusted, and polychoric correlations, respectively. K is the number of categories. N, J, and C refer to normal, Joe VITA, and Clayton VITA underlying distributions.

of categories. This holds across model estimation methods and parameter groups. Also, it is seen that across all conditions, cat-LS (MARB 3.1%) is less biased than cont-ML-adj (MARB 4.7%), which in turn is less biased than cont-ML (MARB 5.5%). In none of the 12 conditions (4 parameter groups vs. 3 levels of K) did cont-ML or cont-ML-adj outperform cat-LS. However, in all nine conditions cont-ML-adj outperformed cont-ML, so the proposed adjustment to Pearson correlation represents an improvement, although not enough to outperform cat-LS.

Further study of Table 4 reveals that the distribution of bias among the parameter groups differs across estimation methods. For cont-ML and cont-ML-adj the most biased parameters are the factor loadings for exogenous variables, while for cat-LS the most biased parameters are the struc-

tural regression coefficients from the exogenous to the endogenous latent variables. Nevertheless, in this group of parameters cat-LS still has less bias than cont-ML and cont-ML-adj.

We next investigate another characteristic that determines ordinal distributions, namely the distributional shapes, and how they interact with the underlying continuous distribution, see Table 5. The underlying continuous distribution affects bias for all three estimation methods. Inspection of the last column in Table 5 reveals that cont-ML and cont-ML-adj are particularly biased when the underlying Joe VITA distribution is discretized so as to produce ceiling effects. With this combination, bias is above 13% for both estimators for parameter groups λ_{en} and γ, that is, for factor loadings and structural paths emanating from the endogenous latent vari-

ables. The other condition where cont-ML and cont-ML-adj are particularly biased is when the Clayton VITA is discretized to produce floor effects, with bias at or well above 10% for groups λ_{en} and γ.

In contrast to cont-ML and cont-ML-adj, cat-LS expectedly exhibits negligible bias with underlying normality. Even with underlying non-normality, cat-LS bias is still small, provided the ordinal distribution is symmetrical. Floor and ceiling effects in the ordinal distributions inflate cat-LS bias under non-normality to around 5%-6%. In particular, there are two such conditions (of a total of nine conditions obtained from crossing the three underlying and the three ordinal distributions) where cont-ML-adj is less biased than cat-LS. In the first condition the Joe VITA is discretized to produce floor effects (cat-LS: 5.9% vs. cont-ML-adj: 2.1%), and the second condition is discretizing the Clayton VITA to produce ceiling effects (cat-LS: 5.1% vs. cont-ML-adj: 1.8%). Detailed inspection of Table 5 reveals that in these two conditions bias in cat-LS is located mainly in the factor loadings. Given that structural parameters, being estimates of association among latent constructs, in many cases are of more substantive interest than factor loadings, we calculated bias for the combined (γ and β) group of structural coefficients. The results indicate that in the two conditions where cat-LS is more biased than cont-ML-adj, the structural biases were comparable. More precisely, underlying Joe VITA combined with floor effects yielded a combined structural parameter bias of 3.2% and 3.1% for cat-LS and cont-ML-adj, respectively. Similarly, for underlying Clayton VITA combined with ceiling effects, combined structural parameter bias was 2.6% and 2.9% for cat-LS and cont-ML-adj, respectively. Therefore, in terms of structural parameter bias, cont-ML-adj did not clearly outperform cat-LS in any condition. On the other hand, for factor loading bias cont-ML-adj outperformed cat-LS in 2 of 9 conditions.

To sum up, at the population level we found that cat-LS overall was less biased than cont-ML-adj, which again was less biased than cont-ML. Hence it is worthwhile to replace the Pearson correlations with adjusted Pearson correlations before doing the SEM analysis based on Pearson correlations, although this methodology was found to perform considerably worse than cat-LS. We found a strong interaction between the underlying continuous distribution and the ordinal observed distribution on parameter bias. Especially two conditions, namely Joe VITA combined with floor effects, and Clayton VITA combined with ceiling effects, produced poor performance of cat-LS relative to cont-ML-adj. In these conditions the bias was mostly related to factor loadings, and structural parameter bias in cat-LS was comparable to cont-ML-adj.

Table 3

Study 1: Mean absolute relative bias of estimated correlations.

Underlying distribution	Ordinal distribution	K	Pearson	Pearson-adj	polychoric
N	Floor	4	17.8	17.4	0.7
		5	13.7	12.9	0.6
		7	9.6	8.1	0.4
	Ceiling	4	18.0	17.5	0.7
		5	13.8	12.9	0.6
		7	9.8	8.1	0.4
	Symm.	4	14.1	13.9	0.6
		5	11.0	10.7	0.5
		7	6.1	5.6	0.4
J	Floor	4	5.5	5.5	16.1
		5	5.6	3.3	13.1
		7	7.4	1.8	9.6
	Ceiling	4	43.1	39.6	26.3
		5	38.6	32.9	22.4
		7	33.9	24.6	17.3
	Symm.	4	19.3	18.7	5.1
		5	14.3	13.4	3.3
		7	9.5	7.9	2.1
C	Floor	4	37.0	33.8	19.4
		5	32.5	27.6	16.4
		7	28.0	20.0	12.5
	Ceiling	4	4.1	5.7	14.6
		5	3.4	2.9	12.1
		7	5.3	1.2	9.0
	Symm.	4	16.6	16.3	2.5
		5	12.7	12.1	1.7
		7	7.8	6.8	1.1
Total			16.2	14.1	7.8

Note. Floor, Ceiling, and Symmetrical refer to ordinal distributions with floor effect, ceiling effect, and symmetry, respectively. Pearson, Pearson-adj, and polychoric refer to estimation based on Pearson, adjusted Pearson, and polychoric correlations, respectively. N, J, and C refer to underlying distributions of type normal, Joe VITA, and Clayton VITA, respectively.

Study 2: CI coverage and model fit rejection rates

Across all conditions, cont-ML, cont-ML-adj and cat-LS always produced properly converged solutions in model estimation. Reported here are results for sample size $n = 500$. Findings for $n = 1000$ were not qualitatively different, and are presented in the online supplementary material.

Coverage rates at the 95% confidence level for the regression coefficient γ_{21} and for the factor loading coefficient λ_{11} are presented in Figures 11 and 12. With underlying normality, coverage rates for the structural parameter γ_{11} are deemed acceptable for all estimators. With underlying Joe or Clayton VITA, performance deteriorates, especially for cat-LS. Overall, for the Joe VITA, cont-ML, cont-ML-adj and cat-LS have acceptable coverage in 96.5%, 96.7%,

Table 4
Study 1: Mean absolute relative bias of model parameters.

Estimator	Parameter group	K			Group mean	Overall mean
		4	5	7		
cont-ML	λ_{ex}	6.0	4.9	3.5	4.8	
	λ_{en}	9.0	7.3	5.6	7.3	
	γ	6.9	5.9	5.1	5.9	5.5
	β	3.4	2.9	2.4	2.9	
cont-ML-adj	λ_{ex}	5.6	4.2	2.4	4.1	
	λ_{en}	8.5	6.3	3.9	6.3	
	γ	6.3	5.1	3.7	5.0	4.7
	β	3.0	2.3	1.5	2.3	
cat-LS	λ_{ex}	3.5	2.9	2.3	2.9	
	λ_{en}	3.8	3.1	2.4	3.1	
	γ	4.9	4.1	3.1	4.0	3.1
	β	2.6	2.1	1.5	2.0	
Overall mean		5.6	4.5	3.2		

Note. cont-ML, cont-ML-adj, and cat-LS refer to model estimator. $\lambda_{ex}, \lambda_{en}, \gamma$, and β refer to groups of model parameters, containing, respectively, factor loadings for exogenous and endogenous latent variables, regression coefficients among exogenous and endogenous latent variables, and among endogenous latent variables. $K=$ number of categories.

Table 5
Study 1: Mean absolute relative bias of model parameters.

Estimator	Underlying distribution	Ordinal distribution	Parameter group				Overall
			λ_{ex}	λ_{en}	γ	β	
cont-ML	N	Floor	5.2	6.6	3.6	2.2	5.0
		Ceiling	5.2	6.6	2.9	1.4	4.7
		Symmetrical	5.0	5.3	1.3	1.2	3.9
	J	Floor	3.4	3.1	4.7	3.7	3.6
		Ceiling	7.8	16.3	15.8	4.4	11.5
		Symmetrical	2.9	5.8	6.4	3.0	4.4
	C	Floor	7.9	14.6	10.7	4.1	9.9
		Ceiling	2.6	2.3	3.6	3.9	2.8
		Symmetrical	3.3	5.5	4.3	2.3	4.0
cont-ML-adj	N	Floor	5.3	6.2	2.8	1.7	4.7
		Ceiling	5.2	6.2	2.3	1.2	4.5
		Symmetrical	4.9	5.2	1.2	1.1	3.8
	J	Floor	1.7	1.5	3.6	2.2	2.1
		Ceiling	5.9	13.0	14.0	3.9	9.3
		Symmetrical	3.0	5.5	5.3	2.5	4.1
	C	Floor	6.0	11.8	9.6	3.4	8.1
		Ceiling	1.2	1.4	3.0	2.5	1.8
		Symmetrical	3.5	5.3	3.6	1.9	3.9
cat-LS	N	Floor	0.3	0.3	1.1	1.1	0.5
		Ceiling	0.3	0.3	1.3	1.1	0.6
		Symmetrical	0.3	0.2	1.2	1.2	0.6
	J	Floor	7.5	6.7	3.5	2.5	5.9
		Ceiling	4.4	7.0	11.8	3.4	6.6
		Symmetrical	2.4	1.0	4.7	2.3	2.5
	C	Floor	2.8	5.8	7.2	2.5	4.5
		Ceiling	6.4	5.9	2.5	2.6	5.1
		Symmetrical	1.5	0.6	3.0	1.6	1.6

Note. cont-ML, cont-ML-adj, and cat-LS refer to model estimator. $\lambda_{ex}, \lambda_{en}, \gamma$, and β refer to groups of model parameters, containing, respectively, factor loadings for exogenous and endogenous latent variables, regression coefficients among exogenous and endogenous latent variables, and among endogenous latent variables. N, J, and C refer to underlying distributions of type normal, Joe VITA, and Clayton VITA, respectively.

and 51.1% of the conditions, respectively. Similarly, for the Clayton VITA, acceptable coverage was achieved in 100%, 100%, and 75.1% of the conditions. Therefore, with underlying non-normality, cat-LS coverage for this structural parameter was far less acceptable than the coverage attained by cont-ML and cont-ML-adj.

Let us next investigate coverage for the factor loading coefficient λ_{11} which relates X_1^* with ξ_1, see Figure 12. Under normality cat-LS performs well, while cont-ML and cont-ML-adj perform poorly. In fact, cont-ML and cont-ML-adj had acceptable coverage in 0.7% and in 1.1% of the conditions with underlying normality, respectively, while cat-LS had acceptable coverage in 100% of the conditions. With underlying non-normality, performance deteriorates markedly for all three estimators. Overall, for the Joe VITA, cont-ML, cont-ML-adj, and cat-LS have acceptable coverage in 15.1%, 30.9%, and 9.3% of the conditions, respectively. With the Clayton VITA, acceptable coverage was achieved in 17.8%, 38.7%, and 25.1% of the conditions. Therefore, cat-LS did not perform as well as cont-ML and cont-ML-adj with underlying non-normality. The results are highly dependent upon the specific thresholds. Consider, for instance, cont-ML with five categories and underlying Joe VITA distribution. Across the fifty ceiling distributions coverage then ranges from 0% to 83.6%, with median coverage 36.8%.

There is a notable difference between the coverage of γ_{21} and λ_{11}. The former parameter is well estimated, while the latter has very poor coverage for all estimators in the majority of conditions. It is only cat-LS with underlying normality that yields valid inference for the factor loading λ_{11}. To sys-

tematically inquire into the differential coverage among different types of model parameters, aggregated coverage rates for parameter groups $\lambda_{ex}, \lambda_{en}, \gamma$, and β, are presented in Table 6. For the factor loading parameters in λ_{ex} and λ_{en}, coverage is generally lower than coverage associated with the structural parameters γ and β. It is especially under cont-ML and cont-ML-adj that factor loadings have unacceptably low coverage rates. Coverage is slightly, but consistently, higher for the loadings of exogenous factors ξ_1 and ξ_2, compared to coverage associated with endogenous factors η_1, η_2, and η_3. Coverage for structural parameters is acceptable for all estimation methods. Overall, we see that coverage improves with the number of categories. The overall coverage rate is highest for cat-LS and lowest for cont-ML.

To further inquire into the effect of observed ordinal distributions, we next look into how the observed distributional forms interact with the type of underlying distribution, see Table 7. cat-LS achieves acceptable coverage under multi-

variate normality regardless of ordinal distributional shape, and under Clayton VITA with symmetrical ordinal distributions. cont-ML does not achieve acceptable coverage in any of the nine conditions. cont-ML-adj has acceptable coverage when the Joe and Clayton VITA distributions are discretized into ordinal data with floor and ceiling effects, respectively. With underlying normality, cont-ML and cont-ML-adj coverage is only slightly better under symmetrical ordinal distributions compared to ordinal distributions with floor or ceiling effects. cat-LS is expectedly performing very well in terms of coverage under multivariate normality, for all distributional forms. With underlying non-normal distributions, all three estimators have coverage that is dependent upon the ordinal distributional form. For cont-ML and cont-ML-adj, coverage is particularly poor when the Joe VITA is discretized into ordinal data with ceiling effects, and when the Clayton VITA is discretized into data with floor effects. For cat-LS poorest coverage appears when the Joe and Clayton VITA distributions are discretized into floor and ceiling effects, respectively. To sum up, in those conditions where cont-ML and cont-ML-adj coverage is poorest, cat-LS offers improved coverage, and vice versa, in poor coverage conditions for cat-LS, cont-ML, and cont-ML-adj offer improved coverage.

Table 6
Study 2: 95% CI coverage rates.

Estimator	Parameter group	K			Group mean	Overall coverage
		4	5	7		
cont-ML	λ_{ex}	68.2	74.5	80.6	74.5	
	λ_{en}	65.5	72.3	78.2	72.0	79.7
	γ	94.6	94.6	94.6	94.6	
	β	94.9	94.9	94.9	94.9	
cont-ML-adj	λ_{ex}	70.0	78.6	87.6	78.7	
	λ_{en}	67.3	76.5	86.1	76.6	82.8
	γ	94.7	94.7	94.8	94.7	
	β	95.0	94.9	95.0	95.0	
cat-LS	λ_{ex}	84.8	85.8	87.3	86.0	
	λ_{en}	83.0	84.1	86.0	84.4	87.7
	γ	94.0	93.6	93.1	93.6	
	β	94.4	93.9	93.4	93.9	
Overall coverage		79.6	83.4	87.3		

Note. cont-ML, cont-ML-adj, and cat-LS refer to model estimator. $\lambda_{ex}, \lambda_{en}, \gamma$, and β refer to groups of model parameters, containing, respectively, factor loadings for exogenous and endogenous latent variables, regression coefficients among exogenous and endogenous latent variables, and among endogenous latent variables. $K=$ number of categories.

The scaled-and-shifted test of correct model specification was evaluated by calculating rejection rates at the 5% level of significance. In Figure 13 are depicted boxplots of rejection rates. Each of the 81 boxplots represent 50 threshold configurations in which rejection rate was calculated. cont-ML-adj attains acceptable performance in all 4050 conditions, while cont-ML has acceptable performance in 99.3%,

Table 7
Study 2: 95% CI coverage rates according to ordinal distributional form.

Estimator	Underlying distribution	Ordinal distribution	Parameter group				Overall
			λ_{ex}	λ_{en}	γ	β	
cont-ML	N	Floor	79.0	76.9	94.6	95.1	83.0
		Ceiling	79.6	76.4	95.0	95.5	83.1
		Symmetrical	82.5	80.7	94.8	95.4	85.6
	J	Floor	86.1	85.0	94.2	94.3	88.2
		Ceiling	39.7	34.8	93.8	94.2	54.5
		Symmetrical	84.1	83.2	94.5	94.7	87.0
	C	Floor	44.8	39.2	94.5	94.9	58.0
		Ceiling	90.1	89.4	94.9	94.6	91.3
		Symmetrical	84.1	82.4	95.0	95.0	86.8
cont-ML-adj	N	Floor	79.6	77.5	94.7	95.3	83.5
		Ceiling	80.0	77.1	94.9	95.5	83.5
		Symmetrical	82.7	81.0	94.8	95.4	85.8
	J	Floor	93.1	93.0	94.5	94.5	93.5
		Ceiling	53.3	48.8	94.0	94.3	64.1
		Symmetrical	84.8	84.2	94.6	94.8	87.6
	C	Floor	56.7	51.4	94.6	94.9	66.3
		Ceiling	93.8	93.5	95.1	94.8	94.1
		Symmetrical	84.6	83.0	95.1	95.1	87.2
cat-LS	N	Floor	94.3	94.0	94.4	95.2	94.3
		Ceiling	94.1	94.0	94.7	95.4	94.3
		Symmetrical	94.2	94.0	94.6	95.3	94.3
	J	Floor	66.4	62.5	91.6	91.8	72.6
		Ceiling	82.9	79.6	93.6	93.7	84.9
		Symmetrical	89.8	89.9	91.8	92.2	90.5
	C	Floor	86.8	83.7	94.3	94.6	87.9
		Ceiling	73.0	69.7	93.6	93.4	78.0
		Symmetrical	92.2	91.7	93.5	93.6	92.4

Note. cont-ML, cont-ML-adj, and cat-LS refer to model estimator. $\lambda_{ex}, \lambda_{en}, \gamma$, and β refer to groups of model parameters, containing, respectively, factor loadings for exogenous and endogenous latent variables, regression coefficients among exogenous and endogenous latent variables, and among endogenous latent variables. N, J, and C refer to the underlying continuous distribution being normal, Joe VITA, and Clayton VITA, respectively. Floor, Ceiling, and Symmetrical refer to the ordinal distributional shape.

98% and 100% of conditions, when underlying distribution is normal, Joe VITA, and Clayton VITA, respectively. For cat-LS, 100% of the conditions involving underlying normality had acceptable rejection rates. However, under non-normality cat-LS performs poorly in terms of Type I error control: Only 3.8% and 12% of conditions involving Joe and Clayton underlying VITA distributions, respectively, produced acceptable rejection rates. In these conditions it is seen that the type of underlying distribution interacts with ordinal observed distributional form when it comes to cat-LS rejection rate. Also, with an increasing number of categories, it is more likely to reject the null hypothesis of correct model specification.

It is surprising that cont-ML and cont-ML-adj model fit testing achieved acceptable Type I error control across a large majority of conditions, and utmost caution must be exercised
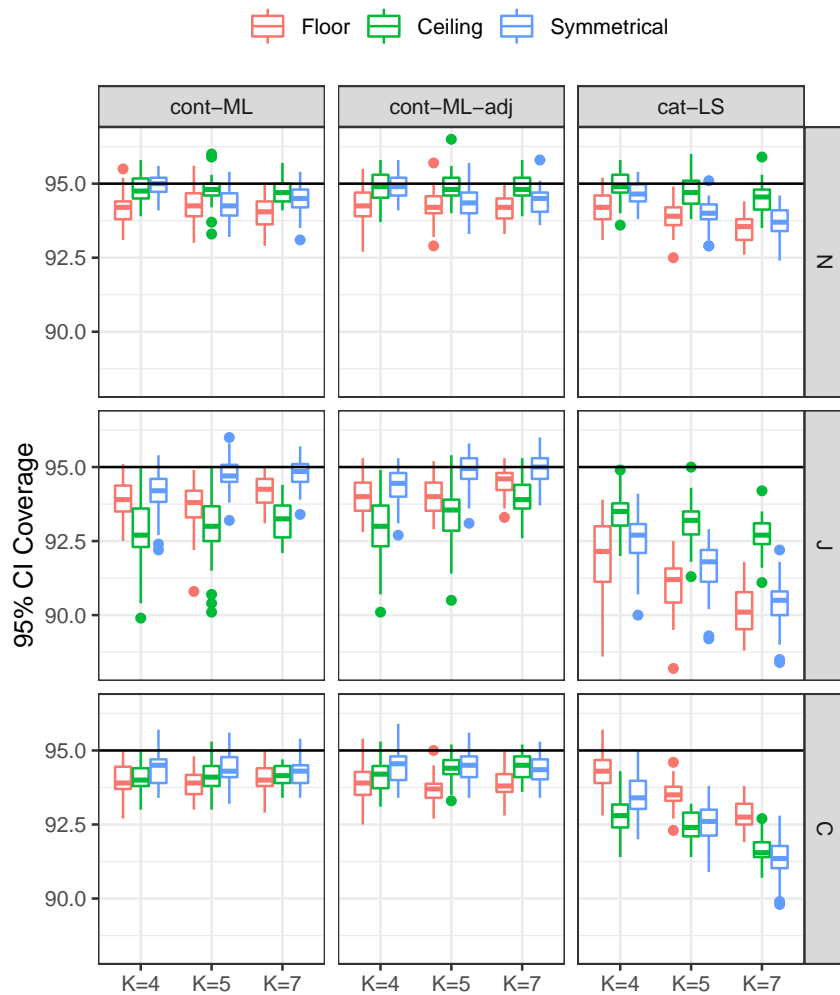
Figure 11. Study 2: Coverage rates for regression parameter γ_{21} at sample size $n = 500$. Each boxplot is based on 50 threshold configurations. Floor, Ceiling and Symmetrical refer to ordinal distributions with floor effect, ceiling effect and symmetry, respectively. cont-ML, cont-ML-adj and cat-LS refer to estimation based on Pearson, adjusted Pearson and polychoric correlations, respectively K is the number of categories. N, J, and C refer to normal, Joe VITA, and Clayton VITA underlying distributions.

when interpreting these rejection rates, as cont-ML, cont-ML-adj and cat-LS are all inconsistent, with the exception of cat-LS under normality. Study 1 demonstrated that the Pearson (adjusted) correlations are biased (see Table 3) in the population, so that the SEM model would seem misspecified at the population level. Study 1 indeed showed that model parameters obtained with cont-ML and cont-ML-adj were biased (Tables 4 and 5). However, the specific SEM model utilized in this study encompasses a large set of correlation structures, including those calculated in cont-ML and cont-ML-adj estimation. In the upcoming discussion section we discuss this topic further, and demonstrate that imposing a few correctly specified constraints in the original SEM model produces unacceptable Type I error control for the cont-ML

and cont-ML-adj model test.

Study 3: The bootstrap test for underlying non-normality

The bootstrap test for underlying normality was evaluated using nine of the 150 threshold configurations employed in Studies 1 and 2. Type I error control (underlying distribution N) was acceptable in 17%, 78%, and 89% of the conditions for sample sizes $n = 100, 300$, and 500, respectively. At sample size $n = 500$, Type I error control was acceptable in all conditions, except under two of the three symmetrical configurations.

In Table 8 are given the rejection rates at the 5% level of significance. Within each distributional form (floor effects, ceiling effects and symmetrical) there was little variation in
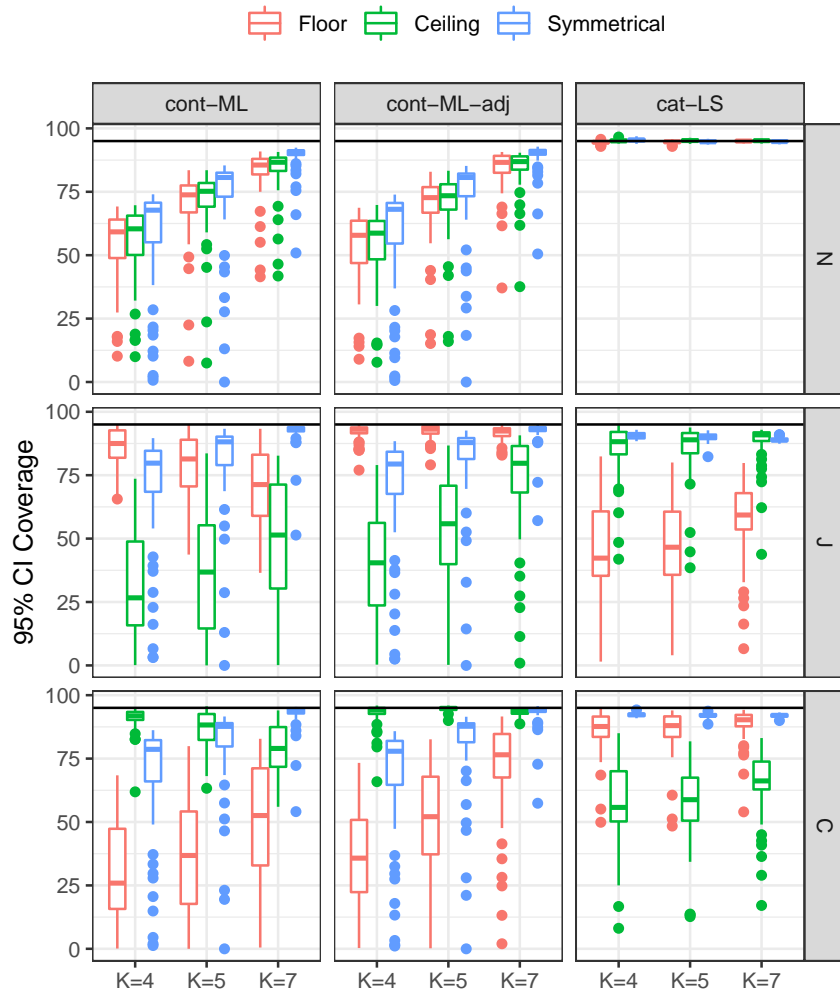
Figure 12. Study 2: Coverage rates for factor loading relating X_1^* to ξ_1. Each boxplot is based on 50 threshold configurations. Floor, Ceiling and Symmetrical refer to ordinal distributions with floor effect, ceiling effect and symmetry, respectively. cont-ML, cont-ML-adj and cat-LS refer to estimation based on Pearson, adjusted Pearson and polychoric correlations, respectively K is the number of categories. N, J, and C refer to normal, Joe VITA, and Clayton VITA underlying distributions.

rejection rates across the three threshold configurations, so we decided to collapse results across these configurations. The full table with rejection rates for each threshold configuration is presented in the supplementary material.

The power of the bootstrap test (underlying distribution J and C) is generally high. For sample sizes $n = 300$ and $n = 500$ the bootstrap test always rejected the null hypothesis of underlying normality, for both types of underlying normality, both levels of K, and all ordinal distributional forms. At the smallest sample size, $n = 100$, violation of the normality assumption was not always detected, especially under the distribution C. However, the bootstrap test exhibited power in the ranges $0.6 - 0.8$ and $0.7 - 0.9$ for $K = 4$ and $K = 7$, respectively.

Discussion

In the present article we carefully analyzed the discretization model which is frequently used to analyze ordinal data in SEM. Identification issues were discussed, and it was argued that to investigate the effect of underlying non-normality, a new simulation method based on VITA distributions is well-suited. Also, we presented theoretical results on the consistency of three kinds of correlation matrices as the number of categories increases. First, we proposed a new encoding of ordinal data that ensures that the Pearson correlation matrix of the adjusted data will approach the true underlying correlation matrix as K increases. Second, we showed that this convergence also holds for the polychoric correlation matrix, provided correct marginal distributions are specified for the underlying vector. Thirdly, consistency of the Spearman cor-
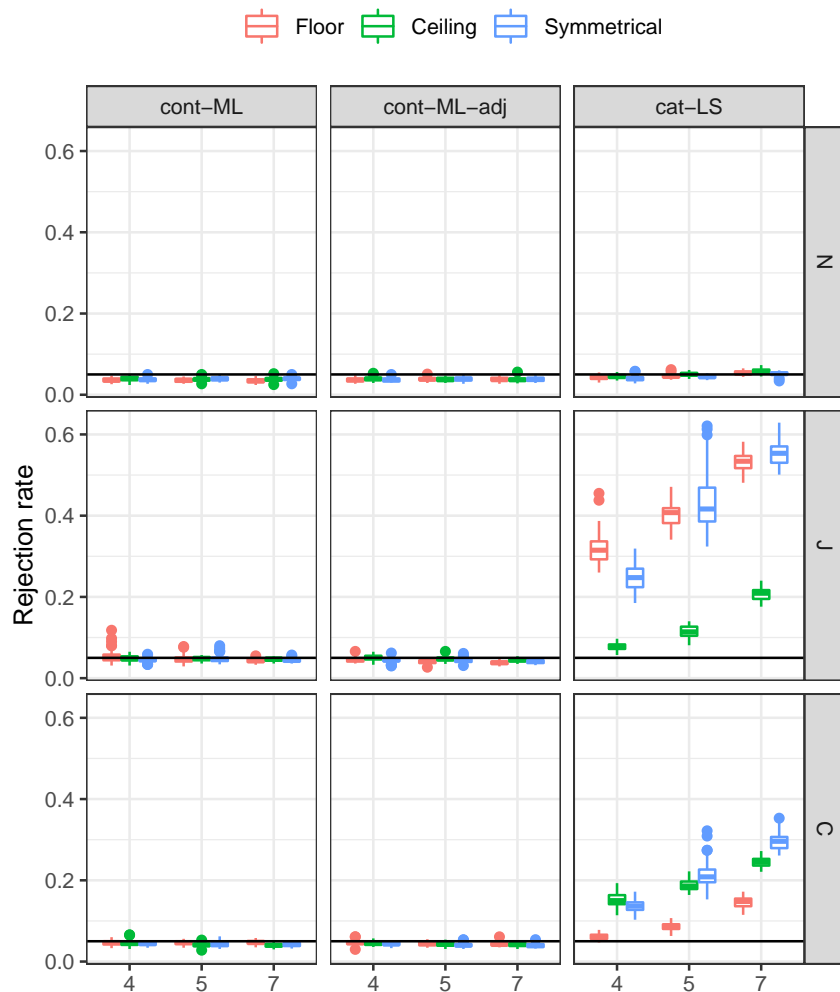
Figure 13. Study 2: Rejection rates at the 5% level of significance of the scaled-and-shifted test statistic. Each boxplot is based on 50 threshold configurations. Floor, Ceiling, and Symmetrical refer to ordinal distributions with floor effect, ceiling effect, and symmetry, respectively. K is the number of categories. N, J, and C refer to normal, Joe VITA, and Clayton VITA underlying distributions. The horizontal line corresponds to 5% rejection rate.

relation based on ordinal data towards the population Spearman correlation of the underlying variable was shown under weak assumptions.

Next, we investigated in the context of a specific medium-sized SEM model, the extent to which underlying non-normality and variation in ordinal distributional forms affects ordinal SEM estimation and inference. We also evaluated for the first time the performance of a bootstrap test for underlying non-normality in the context of a medium-sized SEM model.

Our first research question asked whether variation in ordinal distributional forms affects ordinal SEM estimation and inference. Study 1 showed that population-level correlations and SEM model parameters for all three estimators to a large degree depended on the specific distributional form.

We found systematic variation in estimates as a function of whether the ordinal variables had floor effects, were symmetrical, or had ceiling effects. There was also significant variation within these classes of distributional forms. That is, discretizing a continuous vector in two ways, both of which produce floor effects in all twenty variables, may lead to very different model parameter estimates.

Our second research question asked whether cont-ML, cont-ML-adj, and cat-LS are robust to underlying non-normality. We found that all three estimators are substantially biased when the underlying normality assumption is violated. In the present study cat-LS with underlying non-normality performed much worse than previously reported (e.g., Flora & Curran, 2004; Li, 2016a; Rhemtulla et al., 2012). The sensitivity of cat-LS to underlying non-normality

Table 8

Study 3: Rejection rates at the $\alpha = 0.05$ level of significance of the bootstrap test for underlying normality.

Underlying distribution	K	n	Floor	Ceiling	Symm.
N		100	0.08	0.08	0.11
	4	300	0.05	0.05	0.07
		500	0.05	0.05	0.06
		100	0.18	0.13	0.29
	7	300	0.07	0.06	0.08
		500	0.05	0.05	0.08
J		100	0.92	0.89	0.96
	4	300	1.00	1.00	1.00
		500	1.00	1.00	1.00
		100	0.85	0.99	0.98
	7	300	1.00	1.00	1.00
		500	1.00	1.00	1.00
C		100	0.76	0.66	0.80
	4	300	1.00	1.00	1.00
		500	1.00	1.00	1.00
		100	0.96	0.68	0.88
	7	300	1.00	1.00	1.00
		500	1.00	1.00	1.00

Note. N, J, and C refer to the underlying continuous distribution being normal, Joe VITA, and Clayton VITA, respectively. Floor, Ceiling, and Symmetrical refer to the ordinal distributional shape. $K=$ number of categories. $n=$ sample size.

should be taken into account by researchers. For instance, with underlying non-normality model fit assessment in terms of the chi-square statistic will tend to indicate that the model has poorer fit than is really the case. We also found that for all three estimators the effect of underlying non-normality was strongly moderated by the distributional form of the ordinal variables. For instance, in Figure 9, cont-ML mean relative bias for the structural coefficient γ_{21} across 50 ordinal ceiling distributions with $K = 4$ was -20% under the Joe VITA. The same ceiling ordinal distributions produced a mean relative bias of only -3% under the Clayton VITA. For the 50 ordinal $K = 4$ distributions with floor effects, the situation is reversed, with a mean relative bias of -1% for the Joe VITA, compared to -11% for the Clayton VITA. This pattern may allow us to speculate that upper (lower) tail dependency combined with ordinal ceiling (floor) effects makes it especially hard to recover the true underlying correlation.

The third topic on our research agenda was the theoretical and empirical assessment of the new cont-ML-adj estimator, which may be seen as a hybrid of cont-ML and cat-LS, since it incorporates the discretization model but still treats data as continuous in the sense that Pearson correlations are used.

In contrast to cont-ML, we proved that cont-ML-adj is consistent as K increases. Across all conditions in our simulation design cont-ML-adj outperformed cont-ML on all evaluation criteria. However, cat-LS outperformed cont-ML-adj in most conditions. Interestingly, the structural regression coefficients did not suffer from poor CI coverage in any of the three estimators, while the factor loading parameters had unacceptable coverage, especially under cont-ML and cont-ML-adj. Similar results were also reported in Li (2016a) and Li (2016b). The acceptable performance of structural parameter coverage may well be, however, a by-product of the specified model in this and related studies.

We urge caution on this point because, as seen in Study 1, the parameter estimates are biased for all three estimators under non-normality. We have observed in numerical experiments that the bias is concentrated mainly in either the factor loadings or in the regression coefficients, depending upon how we identify the SEM. If the model is identified by setting unit variance for the latent variables, which we do in the present study, we impose a correct assumption on the latent variables, which reduces bias in the structural part of the model. However, when identifying the model by fixing one factor loading per latent variable to its true value, bias is reduced in the factor loadings, but increases in the structural part of the model. This means that the high quality of structural coefficient inference observed in Study 2 may be of limited external validity.

Similarly, caution is also warranted when interpreting the even more striking Type I error control of the test of model fit demonstrated by cont-ML and cont-ML-adj in Study 2. We know from Study 1 that these estimators are biased at the population level, so we would not in general expect the model fit test to perform well. We discuss this issue further at the end of this section.

We have seen that cont-ML, cont-ML-adj, and cat-LS need not give valid results when underlying normality is violated. It is therefore important to evaluate methods of detecting underlying non-normality. To answer our fourth research question, we evaluated in Study 3 the performance of the newly proposed bootstrap test (Foldnes & Grønneberg, 2019b) of underlying normality in ordinal data. The test was found to adequately maintain Type I error control for sample sizes of $n = 300$ or more. Also, the bootstrap test exhibited high power to detect underlying non-normality.

We next discuss some limitations in the present study, which may lead to follow-up studies extending the present work. Our results were obtained in the context of a single medium-sized SEM whose population values were chosen to be representative of real-world applications (Li, 2016b). The correlations implied by this model are moderate, with a maximum value of 0.56. We conjecture that bias becomes even more pronounced under stronger correlations up to a certain point, as observed by Foldnes and Grønneberg

(2019b). Robustness of ordinal SEM to underlying non-normality, and its dependence on threshold configurations in other application settings (e.g., multilevel or multi-group analyses) should be studied in future work. Our implementation of cat-LS was based on DWLS estimation. An alternative would be unweighted least squares (ULS) estimation. We believe however that our results will replicate under ULS estimation, given that ULS and DWLS have been reported to yield similar results (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Li, 2016b). As for the proposed new cont-ML-adj estimator, we considered only one simple way of obtaining adjusted values. There may be additional improvements to be gained from identifying better placements of the numerical values of the ordinal observations.

The external validity of our simulation study is also limited by our choice of underlying and ordinal observed distributional forms. First, cat-LS assumes that the underlying variable is normal, and while we have studied its performance under non-normality, we have kept the underlying marginals exactly normal. This means that we have studied a situation where the marginals are correctly specified. Second, we have studied only two non-normal multivariate distributions. Using the VITA simulation technique of Grønneberg and Foldnes (2017) and implemented in the R-package covsim (Foldnes & Grønneberg, 2020), one may easily extend our study to a much broader range of dependencies. Third, although we considered many threshold configurations, the ordinal variables in each condition all belonged to the same type (ceiling, floor or symmetrical) of distributions. In real-world settings some observed variables may have symmetrical distributions, some may have ceiling effects, and others may have floor effects. An extension of our work would therefore be to discretize the continuous vector X^* into variables whose distributions are mixed among ceiling, floor, and symmetrical observed distributions. Fourth, we have simulated data only with a correctly specified underlying moment model. Therefore, our discussion of the performance of model fit tests only concerns Type I error control.

Our simulation study was based on the VITA simulation methodology of Grønneberg and Foldnes (2017). The two advantages of using VITA to simulate X^* in contrast to using the VM method are: First, VITA will in most cases not be discretized equivalent to an exactly multivariate normal random vector. Second, we are able to fix marginals and control other aspects of the distribution of X^* including $\text{Cov}(X^*)$, which means that we can isolate the effect of the marginal distribution of X as explained above. The VITA method is still just one possible member of the space of distributions that when discretized into an ordinal vector $\tilde{X}$ has exactly the same distribution as X. Unless a partial identification analysis is conducted, as discussed in the online supplementary material (p. 1), simulation studies have to be interpreted with caution.

Interpreting cont-ML and cont-ML-adj model fit rejection rates

In Study 2 (see Figure 13) we reported that cont-ML and cont-ML-adj maintained Type I error rates surprisingly well for the the test of correct model specification. When dealing with consistent estimation methodology, observing a rejection rate which is close to nominal in a simulation study is in favor of the estimation methodology. However, Study 1 demonstrated that cont-ML and cont-ML-adj estimates are biased in the population (see Table 3). This should be kept in mind when interpreting the rejection rates in Study 2. The model fit test assesses whether the population covariance matrix of X is expressible by the considered covariance model: $\Sigma(\theta_0) = \text{Cov}(X)$ for some vector θ_0 of model parameters. The alternative hypothesis is that no such θ_0 exists. The model fit test is consistent, in other words, the rejection rate will converge to 100% under the alternative hypothesis. Our simulations therefore indicate that the null hypothesis is true when using cont-ML and cont-ML-adj. The null hypothesis only specifies that the covariance model is fulfilled, and it may therefore be true also when the population parameters of the underlying model for X^* are not reached. Our design ensures that the model holds for X^*, and there is therefore a θ° such that $\Sigma(\theta^\circ) = \text{Cov}(X^*)$. From Study 1 we know that $\theta^\circ \neq \theta_0$. This means that the covariance of the ordinal observations X are compatible with the covariance model $\theta \mapsto \Sigma(\theta)$. The specific SEM model utilized in this study encompasses a large set of correlation structures, coincidentally also including those calculated in cont-ML and cont-ML-adj estimation. The almost nominal rejection rates associated with cont-ML and cont-ML-adj therefore occur as a by-product of the model we set up in Figure 7, and should count in favor of cont-ML and cont-ML-adj only with these important caveats in mind. Why then, do not cat-LS rejection rates approximate the nominal 5% level under non-normality? It so happens that the parameters towards which cat-LS converges are not included in the covariance model $\theta \mapsto \Sigma(\theta)$. If cat-LS estimates were inconsistent under non-normality yet converged to a parameter configuration compatible with the covariance model, this would not by itself be in favor of cat-LS. This argument is not a defense of the high rejection rate of the goodness of fit test with cat-LS. Instead, it is a cautionary note on interpreting rejection rates for the goodness of fit test with inconsistent estimators in general.

We confirm this observation by restricting the covariance model through imposing four correctly specified constraints on our model. We impose equality constraints on all five factor loadings that take the value 0.7 in the population. We simulated one thousand datasets, each of size $n = 1000$ from a multivariate normally distributed X^*, which were discretized into ordinal datasets with $K = 4$ and ceiling effects. Three

model fit statistics were computed: cont-ML without and with constraints, and cat-LS with constraints. The rejection rates were 4.9%, 99.9%, and 4.7% respectively. Compatible with the simulations from Study 2, cont-ML for the original unrestricted model has a rejection rate close to nominal. This is the same for cat-LS with the constraints, which is expected since we simulate under normality. In contrast, cont-ML with the (true) constraints has a rejection rate close to 1, illustrating that cont-ML does not in general reach the covariance model of X^*. This supports our claim that the approximately nominal rejection rates observed for cont-ML and cont-ML-adj observed in Study 2 were due to the fact that the unconstrained SEM model considered encompasses the covariance matrices of X. Details of this numerical illustration are given in an R-script in the supplementary material.

Conclusions

Several well-cited simulation studies have concluded that ordinal SEM estimation (e.g., the WLSMV option in Mplus) is fairly robust to violation of underlying normality (e.g., Flora & Curran, 2004; Quiroga, 1994; Rhemtulla et al., 2012), and consequently researchers rarely test for underlying normality in ordinal data. The present article describes the first comprehensive simulation study where the simulated data are not compatible with underlying normality. The results show that ordinal SEM is sensitive to underlying non-normality, resulting in parameter bias at the population level and unacceptable CI coverage rates at the finite-sample level. We also introduced and analyzed cont-ML-adj, which theoretically and empirically outperforms the classical cont-ML approach to treating ordinal data as continuous.

The conclusion based on our simulation study is that standard methods, as well as the new cont-ML-adj, are all highly sensitive to underlying non-normality. It is therefore important for researchers to carefully consider whether underlying normality is a reasonable assumption, and to empirically assess whether the data contain evidence to the contrary. A bootstrap test to this purpose is available in the R package discnorm (Foldnes & Grønneberg, 2020b). In the present study the bootstrap test adequately controlled the Type I error for a sample size of $n = 300$. Also, the test was able to reliably detect underlying non-normality at this sample size. We recommend that applied researchers use this test in applied work.

From our theoretical study of the problem, we also conclude that all three estimators will work better as the number of categories K increases. That is, when K increases, statistical analysis of ordinal data becomes more well behaved. Therefore, applied researchers ought to aim at increasing K when designing studies. Whether or not this is feasible in a practical situation involves many considerations that are outside the scope of the present work.

So which estimator do we recommend? We must be care-ful to point out that the external validity of our study is limited, as it is for all simulation studies. However, based on our study we recommend cat-LS over cont-ML-adj. If underlying normality is seen as reasonable, and the bootstrap test of underlying normality is not rejected, cat-LS estimation and inference are valid procedures. Should the bootstrap test indicate that we are handling data stemming from underlying non-normality, cat-LS still is preferable to cont-ML-adj. In terms of population bias, Study 1 found that overall, the mean relative model parameter bias under non-normality was 0% and -4% for cat-LS and cont-ML-adj, respectively, and the corresponding values for the mean absolute relative bias were 4.4% and 4.9%, respectively. In addition, Study 2 revealed that CI coverage rates under non-normality are overall higher with cat-LS (84%), compared to cont-ML-adj (82%) (see Table 7). However, with underlying non-normality, we must interpret cat-LS results with caution. Estimates may be biased, and model fit assessment may be biased towards poor fit. In brief, applied researchers should be aware that cat-LS in the presence of underlying non-normality can not be considered a robust methodology.

Acknowledgments

References

Almeida, C., & Mouchart, M. (2014). Testing normality of latent variables in the polychoric correlation. *Statistica*, *74*(1), 3–25.

Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. *Mplus technical appendix*, 1–8.

Babakus, E., Ferguson, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of marketing research*, *24*(2), 222–228.

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: A Multidisciplinary Journal*, *13*(2), 186-203.

Bedford, T., & Cooke, R. M. (2002). Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, *30*(4), 1031–1068.

Bentler, P. (2006). *Eqs 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. doi: 10.1002/9781118619179

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152.

Brennan, S. E., McKenzie, J. E., Turner, T., Redman, S., Makkar, S., Williamson, A., . . . Green, S. E. (2017). Development and validation of seer (seeking, engaging with and evaluating research): a measure of policymakers' capacity to engage with and use research. *Health research policy and systems*, *15*(1), 1.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62–83.

Christoffersson, A. (1977). Two-step weighted least squares factor analysis of dichotomized variables. *Psychometrika*, *42*(3), 433–438.

De Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, *41*(4), 471–503.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural equation modeling*, *9*(3), 327–346.

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.

Foldnes, N., & Grønneberg, S. (2020a). covsim: Simulate from distributions with given covariance matrix and marginal information [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=covsim (R package version 0.1.0)

Foldnes, N., & Grønneberg, S. (2020b). discnorm: Test for discretized normality in ordinal data [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=discnorm (R package version 0.1.0)

Foldnes, N., & Grønneberg, S. (2015). How general is the Vale–Maurelli simulation approach? *Psychometrika*, *80*(4), 1066–1083.

Foldnes, N., & Grønneberg, S. (2019a). On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika*, *84*(4), 1000–1017.

Foldnes, N., & Grønneberg, S. (2019b). Pernicious polychorics: The impact and detection of underlying non-normality. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–19.

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.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A monte carlo study comparing dwls and uls estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641.

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, *52*, 1–14.

Granberg-Rademacker, J. S. (2010). An algorithm for converting ordinal scale measurement data to interval/ratio scale. *Educational and Psychological Measurement*, *70*(1), 74–90.

Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *Psychometrika*, *82*(4), 1035–1051.

Grønneberg, S., & Foldnes, N. (2019). A problem with discretizing Vale-Maurelli in simulation studies. *Psychometrika*, *84*, 554–561.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, *71*(1), 105–131.

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2013). copula: Multivariate dependence with copulas [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=copula (R package version 0.999-7.)

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*(3), 329–367.

Jin, S., Luo, H., & Yang-Wallentin, F. (2016). A simulation study of polychoric instrumental variable estimation in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 680–694.

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, *82*(1), 67–85.

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*(3), 398–407.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482.

Jöreskog, K. G. (1994). Structural equation modeling with ordinal variables. *Lecture Notes-Monograph Series*, 297–310.

Jöreskog, K. G., & Sörbom, D. (2015). Lisrel 9.20 for windows [computer software]. *Skokie, IL: Scientific Software International, Inc.*.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2003). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American political science review*, *97*(4), 567–583.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2004). *Continuous multivariate distributions, volume 1: Models and applications* (Vol. 1). John Wiley & Sons.

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949.

Li, C.-H. (2016b). The performance of ml, dwls, and uls estimation with robust corrections in structural equation models with ordinal variables. *Psychological methods*, *21*(3), 369.

Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological methods*, *22*(3), 486.

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., ... Parker, P. (2013). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of arab and anglo-saxon countries. *Journal of Educational Psychology*, *105*(1), 108.

Marsh, H. W., Vallerand, R. J., Lafrenière, M.-A. K., Parker, P., Morin, A. J., Carbonneau, N., ... and others (2013). Passion: Does one scale fit all? construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychological Assessment*, *25*(3), 796.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, *71*(1), 57–77.

Monroe, S. (2018). Contributions to estimation of polychoric correlations. *Multivariate behavioral research*, *53*(2), 247–266.

Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, *10*(2), 60-70.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*(2), 171–189.

Muthén, B., & Muthén, L. (2012). Mplus version 7: User's guide. *Los Angeles, CA: Muthén & Muthén*.

Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. *SAGE Focus Editions*, *154*, 205–205.

Nagler, T., & Vatter, T. (2019). rvinecopulib: High performance algorithms for vine copula modeling [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rvinecopulib (R package version 0.5.1.1.0)

Natesan, P. (2015). Comparing interval estimates for small sample ordinal CFA models. *Frontiers in psychology*, *6*, 1599. doi: 10.3389/fpsyg.2015.01599

Neff, K. D., Whittaker, T. A., & Karl, A. (2017). Examining the factor structure of the self-compassion scale in four distinct populations: Is the use of a total scale score justified? *Journal of Personality Assessment*, *99*(6), 596–607.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nestler, S. (2013). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 127-143. doi: 10.1111/j.2044-8317.2012 .02044.x

Olsson, U. (1979a). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460.

Olsson, U. (1979b). On the robustness of factor analysis against crude classification of the observations. *Multivariate behavioral research*, *14*(4), 485–500.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. vii. on the correlation of characters not quantitatively measurable. *Philos. Trans. R. Soc. SA*, *196*, 1–47.

Pearson, K., & Pearson, E. S. (1922). On polychoric coefficients of correlation. *Biometrika*, *14*(1), 127–156.

Quiroga, A. M. (1994). *Studies of the polychoric correlation and other correlation measures for ordinal variables*. (Unpublished doctoral dissertation). Uppsala University.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

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.

Robitzsch, A. (2019). sirt: Supplementary item response theory models [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=sirt (R package version 3.4-64)

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Satorra, A., & Bentler, P. (1988). Scaling corrections for statistics in covariance structure analysis (UCLA statistics series 2). *Los Angeles: University of California at Los Angeles, Department of Psychology*.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., & Erhardt, T. (2018). Vinecopula: Statistical inference of vine copulas [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=VineCopula (R package version 2.1.8)

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(3), 465–471.

Yang, Y., & Green, S. B. (2015). Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology*, *11*(1), 23-34.

Online Supplementary Material to Article
*The sensitivity of structural equation modeling
with ordinal data to underlying non-normality
and observed distributional forms*

Njål Foldnes & Steffen Grønneberg

On Partial identification

Here, we consider the interpretation of simulation studies with discretized ordinal vectors when considering more general rather than normal theory methods such as the polychoric estimator. Interpretation then becomes more complex. For normal theory methods, we were able to pin-point an underlying single random vector X^* which the polychoric estimator targeted. In general, such a simple answer will not hold, as the distribution of X^* is not identified from the distribution of X. If we are to study the effect of non-normality of X^* without further strong distributional restrictions that lead to the identification of the distribution of X^* based on X, we must exercise utmost caution, since we are dealing with a non-identified statistical model where unexpected and unusual phenomena may happen.

We here review and extend some of the discussion in Foldnes and Grønneberg (2019). Let us consider the interpretation of using the VM simulation technique to generate X^* that is then discretized into X, which for the sake of argument is encoded using integers from 1 to K. That is, in Equation (1) we have $x_{k,j} = j$ for $1 \le k \le p$ and $1 \le j \le K$. Based on a simulated sample from X, we compare the known Pearson correlation matrix of X^* to that of the empirical correlation matrix of the sample from X. For large samples, this will estimate $\mathrm{Cov}(X)$. What the interpretation and justification of this comparison?

As we have seen, when simulating X^* which is then discretized into X using Equation (1) and some threshold set, there is a large class of random vectors $\tilde{X}^*$ that when discretized according to Equation (1) with adjusted thresholds, yields the same distribution for X. Formally, Foldnes and Grønneberg (2019) calls X^* and $\tilde{X}^*$ *discretize equivalent* if they can be discretized into vectors X and $\tilde{X}$ where X and $\tilde{X}$ have the same distribution. Recall that if two vectors have the same distribution, they are statistically identical. As discussed in Foldnes and Grønneberg (2019), it is not just the marginals of X^* that may vary, but also features of the copula of X^*. When simulating X through discretizing X^* via Equation (1), we are therefore implicitly simulating X *also* through discretizing $\tilde{X}^*$ for all $\tilde{X}^*$ which are discretize equivalent to X^*. In general, the covariance matrix of discretize equivalent vectors X^* and $\tilde{X}^*$ will differ, and may even differ greatly. The comparison of $\mathrm{Cov}(X)$ and $\mathrm{Cov}(X^*)$ is therefore somewhat arbitrary: Why compare $\mathrm{Cov}(X)$ with $\mathrm{Cov}(X^*)$ and not to $\mathrm{Cov}(\tilde{X}^*)$? After all, $\tilde{X}^*$ can be said to generate X just as much as X^*.

This observation rests on an ontological question: When we have instructed the computer to first generate X^*, which is then discretized into X, why is it not clear that our X^* is the "true" underlying vector? The reason that this may be a problematic interpretation is that simulation methods only deals with distribution functions. In a real data-set, assuming we are taking the discretization procedure described by the model in Equation (1) very seriously, it seems unproblematic to think that there *exists* a real X^*, and refer to its distribution and properties in the singular. But in simulation studies, we are only dealing with vectors fulfilling certain distributional laws. Everything is essentially a matter of combining the same source of pseudo random variables into something with a pre-defined distribution, and generating numbers with the pre-defined distribution is the only success criteria.

From this perspective, when simulating X by discretizing X^* using some thresholds, we claim that we are simultaneously also simulating X by discretizing any $\tilde{X}^*$ (with possibly different thresholds) where $\tilde{X}^*$ is discretize equivalent to X^*. Since this point may strike the reader as controversial, let us consider in more detail what makes a simulation method valid. Let us suppose that we have two independent random variables U_1, U_2, both uniformly distributed on $[0, 1]$. A standard way to simulate two independent standard normal random variables is then

$$Z_1 = \Phi^{-1}(U_1), Z_2 = \Phi^{-1}(U_2).$$

This method is valid, since the joint distribution of Z_1, Z_2 is as claimed. Similarly, the Box-Muller method of generating two independent normal random variables is given by

$$Z_3 = \sqrt{-2 \log U_1} \cos(2\pi U_2), Z_4 = \sqrt{-2 \log U_2} \cos(2\pi U_1).$$

This method is equally valid, as one can show that the joint distribution of Z_3, Z_4 is also that of a bivariate independent standard normal distribution. From a simulation point of view, it is irrelevant how the simulated random variables came about. Only the resulting distribution is relevant. Therefore, when simulating X by discretizing X^* using thresholds $(\tau_{k,j})$ is also a perfectly valid method for simulating X by discretizing a discretize equivalent $\tilde{X}^*$ using thresholds $(\tilde{\tau}_{k,j})$ as long as the resulting variable X has the same distribution.

In order to quantify the types of variables $\tilde{X}^*$ that we are implicitly also simulating from when simulating X through discretizing the simulated X^*, let us write $X^* \equiv \tilde{X}^*$ as a short hand for the statement that X^* is discretize equivalent to $\tilde{X}^*$. Let us write the distribution of a random vector Y as P_Y, and consider a set of distributions $\mathcal{P}$ which contains P_{X^*}. This set may for example be the set of all distributions whose marginal distributions are all standard normal. The space of possible covariance matrices compatible with generating X is the set

$$\mathcal{S}_K = \{\mathrm{Cov}(\tilde{X}^*) : \tilde{X}^* \equiv X^*, P_{\tilde{X}^*} \in \mathcal{P}\}.$$

The calculation of such sets is called partial identification analysis, see e.g., Manski (2003). Partial identification analysis is an important theme in modern econometrics, but to our knowledge has hitherto played a minor role in psychometrics. A recent article (Grønneberg, Moss, & Foldnes, 2020) studies $\mathcal{S}_K$ when $K = 2$, and further limits attention to each element of the covariance matrix separately. Grønneberg et al. (2020) is able to quantify the amount of knowledge of the distribution of X^* that is needed to get partial identification sets that are small. The required knowledge is considerable, and even when considering marginal distributions of X^* as known, the set $\mathcal{S}$ is a large set that is too large to be useful.

The calculation or even approximation of $\mathcal{S}$ is non-trivial. Further, in terms of psychometric models, the covariance matrix of X^* is really just a stepping-stone to the estimation of parameters of underlying factor or structural equation models that contain the parameters of fundamental interest. A non-trivial task is therefore to develop a partial identification machinery for the actual covariance model for X^* based on available information in the distribution of X. Simulation studies therefore appear to be necessary, even though they have the unfortunate arbitrariness of comparing statistical methodology based on X to distributional aspects of X^* which is just one out of many representations of underlying random variables that may generate X. The quantification of whether the resulting comparison is typical or not in the class of distributions which may generate X cannot be answered unless a partial identification analysis is performed, either exactly or approximately.

Proof and a technical discussion of Proposition 1

We first provide some insights into the intuition leading up to Proposition 1. Our core observation is that the covariance matrices $\text{Cov}(X^{(K)})$ and $\text{Cov}(X^*)$ are integrals where the integrand is the distribution of $X^{(K)}$ and X^* respectively. For these to be close, we may construct the distribution of $X^{(K)}$ so that the resulting integral defining $\text{Cov}(X)$ is a Riemann sum approximation to $\text{Cov}(X^*)$. To see how this may be achieved, consider the simpler case of the expectation of $X_k^{(K)}$, where $1 \leq k \leq p$. Let $F_k(x) = P(X_k^* \leq x)$ be the k-th marginal cumulative distribution function (CDF) of X^*. Assuming F_k to be continuous with density f_k, and letting $I\{A\}$ denote the indicator function of A, which is one if A is true, and zero

otherwise, we have

$$
\begin{aligned}
\mathrm{E}\, X_k^{(K)} &= \mathrm{E} \sum_{j=1}^K x_{k,j}^{(K)} I\{\tau_{k,j-1}^{(K)} < X_k^* \leq \tau_{k,j}^{(K)}\} \\
&= \sum_{j=1}^K x_{k,j}^{(K)} \mathrm{E}\, I\{\tau_{k,j-1}^{(K)} < X_k^* \leq \tau_{k,j}^{(K)}\} \\
&= \sum_{j=1}^K x_{k,j}^{(K)} P(\tau_{k,j-1}^{(K)} < X_k^* \leq \tau_{k,j}^{(K)}) \\
&= \sum_{j=1}^K x_{k,j}^{(K)} \left[F_k(\tau_{k,j}^{(K)}) - F_k(\tau_{k,j-1}^{(K)}) \right].
\end{aligned}
$$

When choosing the values $(x_{k,j}^{(K)})$ so that $\tau_{k,j-1}^{(K)} \leq x_{k,j}^{(K)} \leq \tau_{k,j}^{(K)}$ for $1 \leq k \leq p$ and $1 \leq j \leq K$, we recognize the above sum as what is known as a Riemann-Stieltjes sum, which approximates a Riemann-Stieltjes integral (Rudin, 1976, Chapter 6). Under further assumptions, we get $\lim_{K \to \infty} \mathrm{E}\, X_k^{(K)} = \int_{-\infty}^{\infty} x \, dF_k(x) = \int_{-\infty}^{\infty} x f_k(x) dx = \mathrm{E}\, X_k^*$, as desired. We now extend this analysis to the full covariance matrix of $X^{(K)}$, and identify conditions on the values of $(x_{k,j}^{(K)})$, the thresholds and the marginal distributions of X^* so that $\text{Cov}(X^{(K)})$ approximates $\text{Cov}(X^*)$.

We next discuss Assumption 1. Assumption 1 (1) is a moment condition required for our proof method. The assumption is close to minimal, in the sense that $\text{Cov}(X^*)$ exists only when $\mathrm{E}\,|X_k^*|^2 < \infty$ for $k = 1, \ldots, p$. Recall that if the k-th marginal is normal, which is the most common assumption, we have $\mathrm{E}\,|X_k^*|^q < \infty$ for any $q > 0$. Assumption 1 (2) specifies that the values $X^{(K)}$ are are getting increasingly closer to the values attained by X^*. Assumption 1 (3) requires that resolution of the threshold configuration increases indefinitely, and that the range of the thresholds will eventually cover the full support of X^*. Recall that for normal marginals, and any distribution with unbounded range, we have $F_k^{-1}(0) = -\infty$ and $F_k^{-1}(1) = \infty$. Assumption 1 (4) places a restriction on the rate of growth of the values $(x_{k,j}^{(K)})$ compared to the threshold values $(\tau_{k,j}^{(K)})$. Note that from Assumption 1 (2), the values of $(x_{k,j}^{(K)})$ are between the thresholds, whose range is expanding to the support of X_k^* and whose resolution increases indefinitely. These properties are therefore also transferred to the numbers $(x_{k,j}^{(K)})$.

Suppose X^* has standard normal marginals, and let us consider Assumption 1 (4) in detail. Since the standard normal distribution Φ is symmetric around zero, we restrict attention to the upper tail requirement $\lim_{K \to \infty} x_{k,K}^{(K)}[1 - \Phi(\tau_{k,K-1}^{(K)})] = 0$. We recall (Feller, 1968, Lemma 2, Chapter VII) that for $x > 0$ we have $1 - \Phi(x) \leq x^{-1}\phi(x)$ where $\phi(x) = \Phi'(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. From Assumption 1 (3), we have $\lim_{K \to \infty} \tau_{k,K-1}^{(K)} = \Phi^{-1}(1) = \infty$. The value $x_{k,K}^{(K)}$ is then to fulfill $\lim_{K \to \infty} x_{k,K}^{(K)}(\tau_{k,K-1}^{(K)})^{-1}\exp(-(\tau_{k,K-1}^{(K)})^2/2) = 0$, which

is implied by the weaker $\lim_{K\to\infty} x_{k,K}^{(K)} \exp(-(\tau_{k,K-1}^{(K)})^2/2) = 0$. For an illustration of this assumption, suppose $\tau_{k,K-1}^{(K)}$ grows at the very slow sub-logarithmic rate $\tau_{k,K-1}^{(K)} = 2(\log K)^{1/2}$. Then, $\exp(-(\tau_{k,K-1}^{(K)})^2/2) = 1/K$, so that for this slow growth we require that $x_{k,K}^{(K)}$ has a sub-linear growth to infinity as $K \to \infty$.

Proof of Proposition 1. The result follows from standard convergence results in probability. We give all technical details for completeness.

We show convergence of the relevant expectations using the following argument, proved as a corollary to Theorem 25.12 in (Billingsley, 1995, p. 348). For completeness, we reproduce the result here (using obvious modifications to the notations in that book). In its statement, $Y_K \implies Y$ as $K \to \infty$ denotes that Y_K convergences in distribution to Y as $K \to \infty$. From Theorem 25.2 in (Billingsley, 1995, p. 340), the required convergence in distribution is implied if Y_K converges in probability to Y. We will write this as $Y_K - Y = o_P(1)$ or equivalently $Y_K = Y + o_P(1)$, where $o_P(1)$ is used to denote a quantity converging to zero in probability. We recall that $Z_K = o_P(1)$ means that for any $\epsilon > 0$ we have that $\lim_{K\to\infty} P(|Z_K| > \epsilon) = 0$.

Corollary 2. *Let r be a positive integer. If $Y_K \implies Y$ and $\sup_{K\geq 1} E[|Y_K|^{r+\epsilon}] < \infty$ where $\epsilon > 0$, then $E[|Y|^r] < \infty$ and $\lim_{K\to\infty} E[Y_K^r] = E[Y^r]$.*

Step 1: Proof that $\lim_{K\to\infty} E X_k^{(K)} = E X_k^*$ for each $1 \leq k \leq p$. Lemma 3 shows that $X^{(K)} = X^* + o_P(1)$, which implies that for each $1 \leq k \leq p$ we have that $X_k^{(K)} = X_k^* + o_P(1)$. The conclusion $\lim_{K\to\infty} E X_k^{(K)} = E X_k$ therefore follows from Corollary 2 if we show $\sup_{K\geq 1} E|X_k^{(K)}|^{1+\epsilon} < \infty$. By assumption, $E|X_k^*|^{2+\epsilon} < \infty$, which gives $E|X_k^*|^{1+\tilde\epsilon} < \infty$ where $\tilde\epsilon = 1 + \epsilon > 0$. The required result $\sup_{K\geq 1} E|X_k^{(K)}|^{1+\epsilon} < \infty$ for some $\epsilon > 0$ therefore follows from Lemma 4.

Step 2: Proof that $\lim_{K\to\infty} E(X_k^{(K)} X_l^{(K)}) = E X_k^* X_l^*$ for all $1 \leq k, l \leq p$. Let $1 \leq k, l \leq p$. Since $\text{Cov}(X_k^{(K)}, X_l^{(K)}) = E(X_k^{(K)} X_l^{(K)}) - (E X_k^{(K)})(E X_l^{(K)})$ and we have already shown that $E X_k^{(K)} \to E X_k^*$ for each $1 \leq k \leq p$, we only need to show that $E(X_k^{(K)} X_l^{(K)}) \to E X_k^* X_l^*$.

Recall that a sequence of variables (Y_K) is said to be bounded in probability, denoted by $O_P(1)$, if for any $\epsilon > 0$ there exists an $M > 0$ such that $\sup_{K\geq 1} P(Y_K > M) < \epsilon$. Also recall that $O_P(1)o_P(1) = o_P(1)$, that $o_P(1) + o_P(1) = o_P(1)$, and that for any random variable Y, we have $Y + o_P(1) = O_P(1)$ (Van der Vaart, 2000, see e.g., Chapter 2.2). For any $1 \leq k \leq p$, we have $X_k = X_k^* + o_P(1) = O_P(1)$ from Lemma

3. Therefore,

$$
\begin{aligned}
X_k^{(K)} X_l^{(K)} &= (X_k^* + o_P(1))(X_l^* + o_P(1)) \\
&= X_k^* X_l^* + X_k^* o_P(1) + X_l^* o_P(1) + o_P(1) \\
&= X_k^* X_l^* + o_P(1).
\end{aligned}
$$

For Corollary 2 to give us the desired conclusion, we are therefore left with showing that for an $\epsilon > 0$ we have

$$
\sup_{K\geq 1} E |X_k^{(K)} X_l^{(K)}|^{1+\epsilon/2} < \infty.
$$

From the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
E |X_k^{(K)} X_l^{(K)}|^{1+\epsilon/2} &= E |X_k^{(K)}|^{1+\epsilon/2} |X_l^{(K)}|^{1+\epsilon/2} \\
&\leq \sqrt{E |X_k^{(K)}|^{2+\epsilon}} \sqrt{E |X_l^{(K)}|^{2+\epsilon}}. \quad (13)
\end{aligned}
$$

From Lemma 4, we have that $\sup_{K\geq 1} E |X_k^{(K)}|^{2+\epsilon} < \infty$ and that $\sup_{K\geq 1} E |X_l^{(K)}|^{2+\epsilon} < \infty$. Therefore, considering Equation (13), we also have that $\sup_{K\geq 1} E |X_k^{(K)} X_l^{(K)}|^{1+\epsilon/2} < \infty$. We have therefore verified the conditions for using Corollary 2, and we conclude that $E X_k^{(K)} X_l^{(K)} \to E X_k^* X_l^*$. □

We now provide the lemmas called upon in the first proof. We start by a preliminary lemma, which re-writes Assumption 1 into the forms which will be directly applied in the upcoming proofs.

Lemma 2. *1. Under Assumption 1 (4), we have that $\lim_{K\to\infty} x_{k,1}^{(K)} P(X_k^{(K)} = x_{k,1}^{(K)}) = 0$ and $\lim_{K\to\infty} x_{k,K}^{(K)} P(X_k^{(K)} = x_{k,K}^{(K)}) = 0$.*

2. Under Assumption 1 (3) we have $X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\}) = o_P(1)$

Proof. 1. We have $P(X_k^{(K)} = x_{k,1}^{(K)}) = P(X_k^* \leq \tau_{k,1}^{(K)}) = F_k(\tau_{k,1}^{(K)})$ and $P(X_k^{(K)} = x_{k,K}^{(K)}) = P(X_k^* > \tau_{k,K-1}^{(K)}) = 1 - F_k(\tau_{k,K-1}^{(K)})$.

2. We have $0 \leq I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\} \leq I\{X_k^* \leq \tau_{k,1}^{(K)}\} + I\{X_k^* > \tau_{k,K-1}^{(K)}\}$. Since $\tau_{k,1}^{(K)} \xrightarrow[K\to\infty]{} F_k^{-1}(0)$ we have that for any $E |I\{X_k^* \leq \tau_{k,1}^{(K)}\}| = E I\{X_k^* \leq \tau_{k,1}^{(K)}\} = P(X_k^* \leq \tau_{k,1}^{(K)}) \xrightarrow[K\to\infty]{} P(X_k^* \leq F_k^{-1}(0)) = F_k(F_k^{-1}(0)) = 0$. Since convergence in absolute value, i.e., L_1 convergence, implies convergence in probability, we have $I\{X_k^* \leq \tau_{k,1}^{(K)}\} = o_P(1)$. Similarly, $I\{X_k^* > \tau_{k,K-1}^{(K)}\} = o_P(1)$. Therefore, $I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\} = o_P(1)$. Now, $X_k^* = O_P(1)$ since X_k^* is a random variable. Therefore, $X_k^* I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\} = O_P(1)o_P(1) = o_P(1)$. □

Lemma 3. *Suppose given Assumption 1. We then have that $X^{(K)} = X^* + o_P(1)$.*

Proof. Let $\delta_i(k) := I\{\tau_{k,i-1}^{(K)} < X_k^* \leq \tau_{k,i}^{(K)}\}$. We have

$$
\begin{aligned}
X_k^{(K)} &= \sum_{i=1}^K (x_{k,i}^{(K)} - X_k^* + X_k^*)\delta_i(k) \\
&= \sum_{i=1}^K (x_{k,i}^{(K)} - X_k^*)\delta_i(k) + X_k^* \underbrace{\sum_{i=1}^K \delta_i(k)}_{=1} \\
&= \sum_{i=1}^K (x_{k,i}^{(K)} - X_k^*)\delta_i(k) + X_k^* \\
&= X_k^* + \sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k) + x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} \\
&\quad + x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\} \\
&\quad - X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)}\} + I\{X_k^* > \tau_{k,K-1}^{(K)}\}) \\
&= X_k^* + x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} + x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\} \\
&\quad - X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\}) \\
&\quad + \sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k). \quad\quad (14)
\end{aligned}
$$

We have $x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} = o_P(1)$ since by Markov's inequality, we have, for $\epsilon > 0$, that

$$
\begin{aligned}
P(x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} > \epsilon) &\leq \epsilon^{-1} \operatorname{E} x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} \\
&= \epsilon^{-1} x_{k,1}^{(K)} P(X_k^* \leq \tau_{k,1}^{(K)}) \xrightarrow[K\to\infty]{} 0
\end{aligned}
$$

by Lemma 2 (1). Also by Lemma 2 (1), $x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\} = o_P(1)$. We also have $X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\}) = o_P(1)$ by Lemma 2 (2). Finally, letting $M_K = \sup_{1\leq k\leq p, 2\leq j\leq K-1}\left[\tau_{k,j}^{(K)} - \tau_{k,j-1}^{(K)}\right]$, and recalling that by Assumption 1 we have $\lim_{K\to\infty} M_K = 0$, we see that

$$
\begin{aligned}
|\sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k)| &\leq \sum_{i=2}^{K-1} |x_{k,i}^{(K)} - X_k^*|\delta_i(k) \\
&\overset{(a)}{\leq} M_K \sum_{i=2}^{K-1} \delta_i(k) \overset{(b)}{\leq} M_K \quad (15)
\end{aligned}
$$

(a) When $\delta_i(k) = 1$, we have that both $x_{k,i}^{(K)}$ and X_k^* are contained within the interval $(\tau_{k,j-1}, \tau_{k,j}]$. The absolute difference between two numbers contained in an interval is bounded by the length of the interval. Therefore, $|x_{k,i}^{(K)} - X_k^*|\delta_i(k) \leq (\tau_{k,j} - \tau_{k,j-1})\delta_i(k) \leq M_K\delta_i(k)$. When $\delta_i(k) = 0$, this inequality is also true, giving the conclusion. *(b)* We have $\sum_{i=1}^K \delta_i(k) = 1$.

In conclusion, we have

$$
X_k^{(K)} = X_k^* + o_P(1).
$$

for each k. □

Lemma 4. *Suppose given Assumptions 1 (2), (3), and (4). If for some $\epsilon > 0$ we have $\operatorname{E}|X^*|^{m+\epsilon} < \infty$, we also have $\sup_{K\geq 1} \operatorname{E}|X_k^{(K)}|^{m+\epsilon} < \infty$.*

Proof. Recalling Equation (14), and writing the L_p norm of a random variable Y as

$$
\|Y\|_p = (\operatorname{E}|Y|^p)^{1/p}
$$

we use Minkowski's inequality (the triangle inequality for L_p norms) with $p = m + \epsilon$ to see that

$$
\begin{aligned}
(\operatorname{E}|X_k^{(K)}|^{m+\epsilon})^{1/(m+\epsilon)} &= \|X_k^{(K)}\|_{m+\epsilon} \\
&= \|X_k^* + x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\} + x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\} \\
&\quad - X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\}) \\
&\quad + \sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k)\|_{m+\epsilon} \\
&= \|X_k^*\|_{m+\epsilon} + \|x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\}\|_{m+\epsilon} + \|x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\}\|_{m+\epsilon} \\
&\quad + \|X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\})\|_{m+\epsilon} \\
&\quad + \|\sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k)\|_{m+\epsilon}.
\end{aligned}
$$

Now $\|X_k^*\|_{m+\epsilon} < \infty$ since X^* is assumed to have $n + \epsilon$ moments.

Recall that a sequence of numbers (x_K) is $o(1)$ if $\lim_{K\to\infty} x_K = 0$. Further,

$$
\begin{aligned}
\|x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\}\|_{m+\epsilon} &= (\operatorname{E}|x_{k,1}^{(K)} I\{X_k^* \leq \tau_{k,1}^{(K)}\}|^{m+\epsilon})^{1/(m+\epsilon)} \\
&= |x_{k,1}^{(K)}| P(X_k^* \leq \tau_{k,1}^{(K)}) = o(1)
\end{aligned}
$$

by Lemma 2 (1), recalling that if for some numbers c_K we have $\lim_{K\to\infty} c_K = 0$, then also $\lim_{K\to\infty} |c_K| = 0$. Similarly, $\|x_{k,K}^{(K)} I\{X_k^* > \tau_{k,K-1}^{(K)}\}\|_{m+\epsilon} = o(1)$.

Since the indicator function is bounded by 1, we have $\|X_k^*(I\{X_k^* \leq \tau_{k,1}^{(K)} \text{ or } X_k^* > \tau_{k,K-1}^{(K)}\})\|_{m+\epsilon} \leq \|X_k^*\|_{m+\epsilon} < \infty$ by the assumption of the lemma. Finally, from Equation (15), we have $\|\sum_{i=2}^{K-1} (x_{k,i}^{(K)} - X_k^*)\delta_i(k)\|_{m+\epsilon} \leq \|M_K\|_{m+\epsilon} = M_K = o(1)$. □

Proofs of Corollary 1, Lemma 1, and Proposition 2

Proof of Corollary 1. For each $1 \leq k \leq p$, let $\gamma_{k,j}^{(K)} = \tau_{k,j}$ for $j = 1, 2, \ldots, K - 1$ and let $\gamma_{k,K} = z_M^{(K)}$. By Assumption 2, we have that Assumption 1 (4) holds for $x_{k,j}^{(K)} = \gamma_{k,j}^{(K)}$. Let $\tilde{X}^{(K)} = (\tilde{X}_1^{(K)}, \ldots, \tilde{X}_p^{(K)})$ where for each $1 \leq k \leq p$ we let $\tilde{X}_k^{(K)} = \sum_{j=1}^K \gamma_{k,j}^{(K)}\delta_j(k)$ where $\delta_j(k) := I\{\tau_{k,j-1}^{(K)} < X_k^* \leq \tau_{k,j}^{(K)}\}$. Proposition 1 therefore implies that $\operatorname{Cov}(\tilde{X}^{(K)}) \xrightarrow[K\to\infty]{} \operatorname{Cov}(X^*)$. Therefore, $\lim_{K\to\infty} \operatorname{Cor}(\tilde{X}_k^{(K)}, \tilde{X}_l^{(K)}) = \operatorname{Cor}(X_k^*, X_l^*)$. We show that in fact $\operatorname{Cor}(\tilde{X}_k^{(K)}, \tilde{X}_l^{(K)}) = \operatorname{Cor}(X_k^{(K)}, X_l^{(K)})$, which finishes

the proof, since this implies the required convergence. Using that $\sum_{j=1}^{K} \delta_j(k) = 1$ we get that

$$
\begin{aligned}
\tilde{X}_k^{(K)} &= \sum_{j=1}^{K} \left(z_m^{(K)} + (z_M^{(K)} - z_m^{(K)})(j-1)/(K-2) \right) \delta_j(k) \\
&= z_m^{(K)} + (z_M^{(K)} - z_m^{(K)}) \left(\sum_{j=1}^{K} j \delta_j(k) - 1 \right)/(K-2) \\
&= \left(z_m^{(K)} + (z_M^{(K)} - z_m^{(K)}) \left(X_k^{(K)} - 1 \right)/(K-2) \right) \\
&= a_k^{(K)} X_k^{(K)} + b_k^{(K)}
\end{aligned}
$$

for non-random numbers $a_k^{(K)}, b_k^{(K)}$. Since correlation is unchanged under affine transformations, the result follows. □

Proof of Lemma 1. Recall that $\phi_2(x_1, x_2; r) = \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)} \left[x_1^2 + x_2^2 - 2rx_1x_2 \right]\right)$. Let $H(r) = \mathrm{E}_{F_{k,l}} \log \phi_2(X_k^*, X_l^*; r)$. Using that X_k^*, X_l^* are standardized, and denoting $\mathrm{Cor}(X_k^*, X_l^*)$ by ρ, we have

$$
\begin{aligned}
H(r) &= -\log(2\pi) - \frac{1}{2}\log(1-r^2) \\
&\quad - \frac{\mathrm{E}(X_k^*)^2 + \mathrm{E}(X_l^*)^2 - 2r\,\mathrm{E}\,X_k^* X_l^*}{2(1-r^2)} \\
&= -\log(2\pi) - \frac{1}{2}\log(1-r^2) - \frac{2-2r\rho}{2(1-r^2)} \\
&= -\log(2\pi) - \frac{1}{2}\log(1-r^2) - \frac{1-r\rho}{(1-r^2)}.
\end{aligned}
$$

We have

$$
\begin{aligned}
H'(r) &= -\frac{1}{2}\frac{2r}{1-r^2} - \frac{-\rho(1-r^2) - (1-r\rho)(-2r)}{(1-r^2)^2} \\
&= \frac{r(1-r^2) + \rho(1-r^2) - (1-r\rho)2r}{(1-r^2)^2} \\
&= \frac{r - r^3 + \rho - \rho r^2 - 2r + 2\rho r^2}{(1-r^2)^2} \\
&= \frac{-r - r^3 + \rho + \rho r^2}{(1-r^2)^2} \\
&= \frac{(r^2+1)(\rho - r)}{(1-r^2)^2},
\end{aligned}
$$

so that $H'(r) = 0$ if and only if $r = \rho$. We have $H''(r) = -(1 + 6r^2 + r^4 - 6r\rho - 2r^3\rho)/(1-r^2)^3$ and $H''(\rho) = -(r^2 + 1)/[(r^2 - 1)^2(r+1)^2] < 0$, showing that $r = \rho$ is the global maximum. □

Proof of Proposition 2. Let $1 \leq k, l \leq p$ be given. Let $U_k^{(K)} = F_k^{(K)}(X_k^{(K)})$ and $U_k = F_k(X_k^*)$ for $1 \leq k \leq p$. Recall $\rho_{k,l}^{S,(K)} = \mathrm{Cor}(U_k^{(K)}, U_l^{(K)})$ and $\rho_{k,l}^S = \mathrm{Cor}(U_k, U_l)$. We have

$$
\begin{aligned}
\rho_{k,l}^{S,(K)} &= \mathrm{Cor}(U_k^{(K)}, U_l^{(K)}) \\
&= \frac{\mathrm{E}\,U_k^{(K)} U_l^{(K)} - (\mathrm{E}\,U_k^{(K)})(\mathrm{E}\,U_l^{(K)})}{\sqrt{\mathrm{E}(U_k^{(K)})^2 - (E U_k^{(K)})^2}\sqrt{\mathrm{E}(U_l^{(K)})^2 - (E U_l^{(K)})^2}}.
\end{aligned}
$$

It therefore suffices to prove $\mathrm{E}\,U_k^{(K)} = \mathrm{E}\,U_k + o(1)$ and $\mathrm{E}\,U_k^{(K)} U_l^{(K)} = \mathrm{E}\,U_k U_l + o(1)$ for every $1 \leq k, l \leq p$.

Let $F_k^{(K)}(x) = P(X_k^{(K)} \leq x)$. Since by Lemma 4 (p.4) we have $X_k^{(K)} = X_k^* + o_P(1)$ as $K \to \infty$, meaning $X_k^{(K)} \xrightarrow[K\to\infty]{P} X_k^*$, and since (as reviewed at the start of the proof of Proposition 1) convergence in probability implies convergence in distribution, we have that also $X_k^{(K)} \xrightarrow[K\to\infty]{d} X_k^*$. By the definition of convergence in distribution, this means that for all x where F_k is continuous, which by the assumption that F_k is continuous means for all x, we have $\lim_{K\to\infty} P(X_k^{(K)} \leq x) = \lim_{K\to\infty} F_k^{(K)}(x) = F_k(x) = P(X_k^* \leq x)$. Since F_k is continuous, this convergence is in fact uniform in x. Indeed, by e.g., Lemma 2.11 in Van der Vaart (2000), we have that $\lim_{K\to\infty} \sup_x |F_k^{(K)}(x) - F_k(x)| = 0$. Since

$$
U_k^{(K)} = F_k^{(K)}(X_k^{(K)}) = F_k(X_k^{(K)}) - [F_k^{(K)}(X_k^{(K)}) - F_k(X_k^{(K)})]
$$

and

$$
|F_k^{(K)}(X_k^{(K)}) - F_k(X_k^{(K)})| \leq \sup_x |F_k^{(K)}(x) - F_k(x)| = o(1),
$$

we get

$$
U_k^{(K)} = F_k(X_k^{(K)}) + o_P(1).
$$

Finally, since F_k is continuous, and $X_k^{(K)} = X_k + o_P(1)$, the continuous mapping theorem gives

$$
U_k^{(K)} = F_k(X_k + o_P(1)) + o_P(1) = F_k(X_k) + o_P(1) = U_k + o_P(1).
$$

Since $U_k^{(K)} \in [0, 1]$, its expectation is bounded by 1, as is its $1 + \epsilon$ moment. Similarly, $\mathrm{E}\,|U_k^{(K)} U_l^{(K)}|^{1+\epsilon} \leq 1$ for any $\epsilon > 0$. Therefore, by Corollary 2, the statement follows. □

On polychoric correlations with misspecified marginals

For simplicity, we assume that both G_k and F_k are strictly increasing, for $k = 1, \ldots, p$. Recall that we assume that F_k for $k = 1, \ldots, p$ are standardized, so Lemma 1 will give consistency of the polychoric estimator if the marginals are chosen in such a manner that the thresholds are consistent. Recall that the true marginals of X^* are G_k for $k = 1, \ldots, d$. Since the marginals are not identified, they can be made to equal F_k with a change of threshold values. Let $U_k = G_k(X_k^*)$ so that U_k is uniform on $[0, 1]$, see Equation (16) (p.6.) Recall that if given a random variable U_k that is uniform on $[0, 1]$, then $\tilde{X}_k = F_k^{-1}(U_k)$ is F_k distributed. This follows as in Equation (19), i.e., we have

$$
P(\tilde{X}_k \leq x) = P(F_k^{-1}(U_K) \leq x) = P(U_k \leq F_k(x)) = F_k(x).
$$

Recall that since F_k is strictly increasing, also F_k^{-1} is strictly increasing. Since G_k is strictly increasing, also the composite map $F_k^{-1}G_k(x) = F_k^{-1}(G_k(x))$ is strictly increasing. We have

$$
X_k = x_{k,j}, \quad \text{when } \tau_{k,j-1} \leq X_k^* \leq \tau_{k,j}
$$

Emulating the argument from Equation (7) to Equation (8) on p. 8 shows that

$$X_k = x_{k,j}, \quad \text{when } \tilde{\tau}_{k,j-1} \le \tilde{X}_k^* \le \tilde{\tau}_{k,j}$$

where

$$\tilde{X}_k^* = F_k^{-1} G_k(X_k^*), \quad \tilde{\tau}_{k,j} = F_k^{-1} G_k(\tau_{k,j}).$$

Therefore, $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_p)'$ has marginals $F_1, \ldots, F_p$ and is discretize equivalent to X^*. Since the threshold estimators built on the assumption that X^* has marginals $F_1, \ldots, F_p$ are consistent estimators when this is correct, and we see that this in fact is correct for a discretize equivalent vector $\tilde{X}^*$, they are consistently estimating the thresholds $\tilde{\tau}_{k,j}$.

An introduction to copulas, vines and the VITA method

Vine distributions present a general method for combining a series of bivariate distributions into a valid full multivariate distribution, and was developed in Joe (1996) and Bedford and Cooke (2002), see Joe (2014). Vines are built around the concept of copulas, which we will introduce in the following. See Nelsen (2007) for a general introduction to copulas. Briefly, copulas allow the specification of a multivariate distribution with uniform marginals that are then connected with univariate marginal distributions, giving a full distribution with the attached marginals. The construction of multivariate distributions is therefore split into first constructing univariate marginal distributions, which is a well-studied problem, and then constructing a multivariate distribution with uniform marginals, i.e., copulas. The construction of copulas is in general difficult, although the vine technique allows a large class of multivariate copula distributions to be constructed in a simple manner.

We will not review all the technical details of the general vine construction. Our aim is instead to provide the necessary overview to gain an understanding of what the VITA simulation method of Grønneberg and Foldnes (2017) does. The VITA method is based on simulating from a particularly chosen vine distribution whose Pearson covariance matrix is fixed. We also explain enough technical details in some introductory cases for the reader to appreciate the main limitation of the vine construction which is given by a property called the simplifying assumption. This limitation is on the other hand the reason for the ease of use and flexibility of vines.

Since VITA is based on vines, which in turn are based on connecting bivariate copula distributions in a certain manner, we start by reviewing copulas, we then present the bivariate copulas we use in our applications, and then we briefly present the core idea of vines. Finally, a brief introduction to the core idea of the VITA simulation method is summarized.

Copulas are distribution functions with marginal distributions that are uniform on [0, 1]. Let us therefore review these concepts first. Consider a random vector $Y = (Y_1, \ldots, Y_p)'$. Its distribution is fully characterized by its CDF given by

$$F(y_1, \ldots, y_p) = P(Y_1 \le y_1, \ldots, Y_p \le y_p) = P(\cap_{j=1}^p \{Y_j \le y_j\}).$$

For any $1 \le j \le p$, let $[a_j, A_j]$ be the shortest interval (possibly infinite) where Y_j is always contained (i.e., the support of Y_j). That is, we always have $a_j \le Y_j \le A_j$. If Y_j is normal, we have $a_j = -\infty$ and $A_j = \infty$. Since $\{Y_j \le A_j\}$ is always the case, we have that for any $1 \le k \le p$ that

$$\begin{aligned}
F_k(y_k) &= F(A_1, \ldots, A_{k-1}, y_k, A_{k+1}, \ldots, A_p) \\
&= P(\{Y_k \le y_k\} \cap \cap_{1 \le j \le p, j \ne k} \{Y_j \le A_j\}) \\
&= P(Y_k \le y_k).
\end{aligned}$$

Therefore, F_k is the k-th marginal distribution of Y, i.e., the CDF of Y_k.

Recall that a distribution F_k is uniform on [0, 1] if $F_k(x) = xI\{0 \le x \le 1\} + I\{x > 1\}$ where $I\{A\}$ is 1 if A is true, and zero otherwise. For $0 \le x \le 1$ we have simply $F_k(x) = x$. The support of F_k is then $[a_k, A_k] = [0, 1]$. Therefore, a copula is a CDF C with the property that for each $1 \le k \le p$ and each $0 \le u_k \le 1$ we have

$$C(1, \ldots, 1, u_k, 1, \ldots, 1) = u_k.$$

The importance of copulas comes from Sklar's theorem (Sklar, 1959), see also Nelsen (2007), which informally states that we may always express any CDF in terms of its copula and its marginal distributions, and that if we start with a copula and some marginal distributions, we may combine them in a manner that always lead to a valid full distribution having the chosen marginal distributions.

For simplicity we assume that for each $1 \le k \le p$ the function F_k is continuous and strictly increasing, and hence has a continuous and strictly increasing inverse F_k^{-1}. Let Y be a random vector with CDF F, which we will write as $Y \sim F$. Then we define the random vector

$$U = (F_1(Y_1), \ldots, F_p(Y_p))'.$$

Denote the CDF of U by C. That is, $U \sim C$. The variable U has uniform marginals, which is seen by

$$\begin{aligned}
P(U_1 \le u_1) = P(F_1(Y_1) \le u_1) &= P(Y_1 \le F_1^{-1}(u_1)) \\
&= F_1\left(F_1^{-1}(u_1)\right) = u_1 \quad (16)
\end{aligned}$$

when $0 \le u_1 \le 1$. Hence C is a copula, and is in fact the copula of F. We calculate the fundamental relations

$$\begin{aligned}
C(u_1, \ldots, u_p) &= P(U_1 \le u_1, \ldots, U_d \le u_p) \\
&= P(F_1(Y_1) \le u_1, \ldots, F_d(Y_d) \le u_p) \\
&= P(Y_1 \le F_1^{-1}(u_1), \ldots, Y_d \le F_d^{-1}(u_p)) \\
&= F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_p)\right) \quad (17)
\end{aligned}$$

and

$$
\begin{aligned}
F(y_1, \ldots, y_p) &= P(Y_1 \le y_1, \ldots, Y_p \le y_p) \\
&= P(Y_1 \le y_1, \ldots, Y_p \le y_p) \\
&= P(F_1(Y_1) \le F_1(y_1), \ldots, F_p(Y_p) \le F_p(y_p)) \\
&= C(F_1(y_1), \ldots, F_p(y_p)). \qquad (18)
\end{aligned}
$$

From Equation (18), we gain a useful method to simulate Y in two stages. We may first simulate $U \sim C$, and then perform the simple transformation

$$
\tilde{Y} = (F_1^{-1}(U_1), \ldots, F_p^{-1}(U_p))'.
$$

The resulting vector, here tentatively called $\tilde{Y}$, has the distribution of F, and is therefore a valid method to simulate from Y. This follows from Equation (18) since

$$
\begin{aligned}
&P(F_1^{-1}(U_1) \le y_1, \ldots, F_p^{-1}(U_p) \le y_p) \\
&= P(U_1 \le F_1(y_1), \ldots, U_p \le F_p(y_p)) \\
&= F(y_1, \ldots, y_p). \qquad (19)
\end{aligned}
$$

In our applications, U will be a vine copula, whose simulation methods are developed in Joe (1996), Bedford and Cooke (2001), and presented, e.g., in Dissmann, Brechmann, Czado, and Kurowicka (2013). The simulation methods are also implemented in several packages in the R system (R Core Team, 2020), such as the vinecopula (Schepsmeier et al., 2018) and rvinecopulib (Nagler & Vatter, 2019) packages.

By the above observation, it is therefore easy (once U is simulated) to set the marginals to what the application requires. In our applications the marginal distributions will be standard normal. When we now discuss vines, we only focus on vine copulas, which are vine distributions with uniform marginals. The full vine distribution will be a vine copula connected to marginal distributions as in the above.

As vines are constructed from a sequence of bivariate copulas in a manner that produces a full multivariate distribution, let us first review the three central bivariate copulas used in the present study. The Gaussian, or normal copula is defined as follows. Let $Y = (Y_1, Y_2) \sim N(\mu, \Sigma)$ where $\mu = (\mu_1, \mu_2)'$ and Σ is a covariance matrix. We may identify the copula of Y using Equation (17). The marginal distributions of Y are

$$
F_k(y_k) = P(Y_k \le y_k) = P\left(\frac{Y_k - \mu_k}{\sigma_k} \le \frac{y_k - \mu_k}{\sigma_k}\right) = \Phi\left(\frac{y_k - \mu_k}{\sigma_k}\right).
$$

Suppose $F_k(y_k) = x$, which means $\Phi\left(\frac{y_k - \mu_k}{\sigma_k}\right) = x$. Therefore, $\Phi^{-1}(x) = \frac{y_k - \mu_k}{\sigma_k}$ which means that $F_k^{-1}(x) = \sigma_k \Phi^{-1}(x) + \mu_k$. Let us write the density F_Y of Y via its correlation ρ and its

standard deviations σ_1, σ_2, so that

$$
\begin{aligned}
f_Y(y_1, y_2) = &\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big(-\frac{1}{2(1-\rho^2)}\Big[\frac{(y_1-\mu_1)^2}{\sigma_1^2} \\
&+ \frac{(y_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\Big]\Big)
\end{aligned}
$$

Note that $f_Y(y_1, y_2) = \phi_2([y_1 - \mu_1]/\sigma_1, [y_2 - \mu_2]/\sigma_2; \rho)$ where $\phi_2(\cdot, \cdot; \rho)$ is the density of a bivariate normal Z with standardized marginals and correlation ρ. The copula of Y is found using Equation (17) and the formula for the inverse of F_1, F_2, and is given by

$$
\begin{aligned}
&C_N(u_1, u_2; \rho) \\
&= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \\
&= \int_{-\infty}^{F_1^{-1}(u_1)} \int_{-\infty}^{F_2^{-1}(u_2)} f_Y(y_1, y_2)\, dy_1 dy_2 \\
&= \int_{-\infty}^{\sigma_1\Phi^{-1}(u_1)+\mu_1} \int_{-\infty}^{\sigma_2\Phi^{-1}(u_2)+\mu_2} \phi_2\Big(\frac{y_1-\mu_1}{\sigma_1}, \\
&\qquad \frac{y_2-\mu_2}{\sigma_2}; \rho\Big) dy_1 dy_2 \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \phi_2(x_1, x_2; \rho)\, dx_1 dx_2
\end{aligned}
$$

using the change of variables $x_1 = [y_1 - \mu_1]/\sigma_1$ and $x_2 = [y_2 - \mu_2]/\sigma_2$. We therefore see that the copula of Y is the same as the copula of Z, and the normal copula is therefore parameterized by the Pearson correlation ρ. There is no simple formula for C_N. In contrast, the Joe and Clayton copulas, denoted by C_J and C_C, are defined directly in terms of their CDFs:

$$
C_J(u_1, u_2; \theta) = 1 - \Big[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\Big]^{1/\theta},
$$

$$
C_C(u_1, u_2; \theta) = \Big[\max\big\{u^{-\theta} + v^{-\theta} - 1; 0\big\}\Big]^{-1/\theta},
$$

where the dependence parameter $\theta \in [1, \infty)$ for the Joe copula, and $\theta \in [-1, \infty) \setminus \{0\}$ for the Clayton copula. Both copulas express positive dependence, but can be rotated to express also negative dependence, see e.g. the documentation of the R package copula (Hofert et al., 2013).

These copulas produce different types of dependencies. The density of the distribution of $Y = (\Phi^{-1}(U_1), \Phi^{-1}(U_2))'$ when $(U_1, U_2)'$ come from a normal, a Joe or a Clayton copula is plotted in Figure 2 (p.5) when the dependency parameters are all chosen such that the Pearson correlation of the final Y is fixed to 0.56. For the normal copula, the dependence parameter is precisely the correlation of the copula when joined with normal marginals, which is exactly what we do here, so that the dependence parameter of the copula is simply $\rho = 0.56$. In contrast, for the Joe and the Clayton copulas, the dependence parameters θ are found via a

numerical search, computing for each candidate value θ the resulting Pearson correlation of Y until a match is found.

We next introduce vines. Our technical discussion is limited to the three-dimensional case, which illustrates the main features of the construction. The general case is later outlined, but without several complicated rules and definitions. A full presentation is given e.g. in Joe (2014), and is summarized in Dissmann et al. (2013), and Grønneberg and Foldnes (2017).

In psychometrics, we are used to summarizing dependencies using a covariance matrix. For a multivariate normal random vector $Z = (Z_1, Z_2, Z_3)'$, the covariance matrix does indeed summarize the multivariate dependence properties of Z perfectly. Since we will be interested in copulas, we assume that Z is standardized. Therefore, we assume $Z \sim N(0, \Sigma)$, where Σ is a correlation matrix.

The fact that the correlation matrix suffices to describe the multivariate dependence structure of Z is somewhat surprising, since it only contains numbers that quantify bivariate features of the distribution. Indeed, Σ consists of the numbers $\rho_{1,2} = \mathrm{Cov}(Z_1, Z_2), \rho_{1,3} = \mathrm{Cov}(Z_1, Z_3)$ and $\rho_{2,3} = \mathrm{Cov}(Z_2, Z_3)$. From these bivariate numbers, we may deduce for example genuine multivariate features of the distribution, such as the conditional distribution of Z_1, Z_2 given Z_3. Since the formula for this conditional distribution will motivate the construction of vines in general, we review its derivation. Using general results for the normal distribution (e.g., Mardia, Kent, & Bibby, 1979), and writing

$$\Sigma = \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{1,2} & 1 & \rho_{2,3} \\ \rho_{1,3} & \rho_{2,3} & 1 \end{pmatrix} = \begin{pmatrix} \Sigma_{(1,2),(1,2)} & \Sigma_{(1,2),3} \\ \Sigma'_{(1,2),3} & \Sigma_{3,3} \end{pmatrix}$$

where

$$\Sigma_{(1,2),(1,2)} = \begin{pmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{pmatrix}, \quad \Sigma_{(1,2),3} = \begin{pmatrix} \rho_{1,3} \\ \rho_{2,3} \end{pmatrix}, \quad \Sigma_{3,3} = (1).$$

we have that

$$(Z_1, Z_2 | Z_3) \sim N(\Sigma_{(1,2),3} \Sigma_{3,3}^{-1} Z_3, \Sigma_{(1,2),(1,2)} - \Sigma_{(1,2),3} \Sigma_{3,3}^{-1} \Sigma'_{(1,2),3})$$

$$\sim N\left(\begin{pmatrix} \rho_{1,3} Z_3 \\ \rho_{2,3} Z_3 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{pmatrix} - \begin{pmatrix} \rho_{1,3} \\ \rho_{2,3} \end{pmatrix} (\rho_{1,3}, \rho_{2,3}) \right)$$

$$\sim N\left(\begin{pmatrix} \rho_{1,3} Z_3 \\ \rho_{2,3} Z_3 \end{pmatrix}, \begin{pmatrix} 1 - \rho_{1,3}^2 & \rho_{1,2} - \rho_{1,3}\rho_{2,3} \\ \rho_{1,2} - \rho_{1,3}\rho_{2,3} & 1 - \rho_{2,3}^2 \end{pmatrix} \right).$$

Conditional on Z_3, we therefore have that Z_1, Z_2 is again normal. Therefore, it has a normal copula, with correlation equal to

$$\rho_{1,2|3} = \frac{\rho_{1,2} - \rho_{1,3}\rho_{2,3}}{\sqrt{1 - \rho_{1,3}^2} \sqrt{1 - \rho_{2,3}^2}}.$$

The number $\rho_{1,2|3}$ is known as the partial correlation, and under normality will exactly specify the dependency properties

of Z_1, Z_2 given Z_3. We have also shown that

$$\mu_{1|3} = \mathrm{E}[Z_1 | Z_3] = \rho_{1,3} Z_3,$$
$$\mu_{2|3} = \mathrm{E}[Z_2 | Z_3] = \rho_{2,3} Z_3,$$
$$\sigma_{1|3} = \sqrt{\mathrm{Var}(Z_1 | Z_3)} = 1 - \rho_{1,3}^2,$$
$$\sigma_{2|3} = \sqrt{\mathrm{Var}(Z_2 | Z_3)} = 1 - \rho_{2,3}^2.$$

From our earlier derivations on the normal copula, we have that

$$P(Z_1 \le z_1, Z_2 \le z_2 | Z_3)$$
$$= C_N(\sigma_{1|3} \Phi^{-1}(z_1) - \mu_{1|3}, \sigma_{2|3} \Phi^{-1}(z_2) - \mu_{2|3}; \rho_{1,2|3})$$
$$= C_N(\sigma_{1|3} \Phi^{-1}(z_1) - \rho_{1,3} Z_3, \sigma_{2|3} \Phi^{-1}(z_2) - \rho_{2,3} Z_3; \rho_{1,2|3}),$$

where the dependence on Z_3 is only present through $\mu_{1|3}$ and $\mu_{2|3}$. Importantly, the copula C_N and its parameter $\rho_{1,2|3}$ do not depend on the value attained by Z_3. The fact that the copula C_N stays the same for every value of Z_3 need not be the case for general non-normal distributions, and we will call this assumption "the simplifying assumption" (Joe, 2014). Joe (1996) defined (a subset of) the distributions later called vine copulas as a generalization of this simplifying assumption. Instead of the normal copula, we could replace C_N by another copula, say $\tilde{C}$, and we would still have a valid probability distribution. From knowing the marginal distributions of Z_1, Z_2, Z_3 when seen separately, the bivariate distributions of (Z_1, Z_3) and (Z_2, Z_3), and finally the bivariate conditional distribution of (Z_1, Z_2) conditioned on Z_3, we have enough information to perfectly reconstruct the full distribution of (Z_1, Z_2, Z_3). To see this, we use the rule of iterated expectations together with the rule $\mathrm{E}[h(X)Y|X] = h(X)\mathrm{E}[Y|X]$ so that

$$P(Z_1 \le z_1, Z_2 \le z_2, Z_3 \le z_3)$$
$$= \mathrm{E}\, I\{Z_1 \le z_1, Z_2 \le z_2, Z_3 \le z_3\}$$
$$= \mathrm{E}\, I\{Z_3 \le z_3\} I\{Z_1 \le z_1, Z_2 \le z_2\}$$
$$= \mathrm{E}\,(\mathrm{E}(I\{Z_3 \le z_3\} I\{Z_1 \le z_1, Z_2 \le z_2\} | Z_3))$$
$$= \mathrm{E}\,(I\{Z_3 \le z_3\} \mathrm{E}(I\{Z_1 \le z_1, Z_2 \le z_2\} | Z_3))$$
$$= \mathrm{E}\,(I\{Z_3 \le z_3\} P(Z_1 \le z_1, Z_2 \le z_2 | Z_3))$$
$$= \mathrm{E}(I\{Z_3 \le z_3\} C_N(\sigma_{1|3} \Phi^{-1}(z_1) - \rho_{1,3} Z_3,$$
$$\quad \sigma_{2|3} \Phi^{-1}(z_2) - \rho_{2,3} Z_3; \rho_{1,2|3}))$$
$$= \int_{-\infty}^{\infty} I\{x_3 \le z_3\} C_N(\sigma_{1|3} \Phi^{-1}(z_1) - \rho_{1,3} x_3,$$
$$\quad \sigma_{2|3} \Phi^{-1}(z_2) - \rho_{2,3} x_3; \rho_{1,2|3}) \phi(x_3)\, dx_3.$$

This is therefore a type of expansion of the full distribution in terms of unconditional and conditional bivariate distributions, as well as the marginal distributions. Note that the bivariate distributions of (Z_1, Z_3) and (Z_2, Z_3) and the marginal distributions of Z_1, Z_2, Z_3, as well as the conditional distribution of (Z_1, Z_2) conditioned on Z_3 were used in forming the above expression.
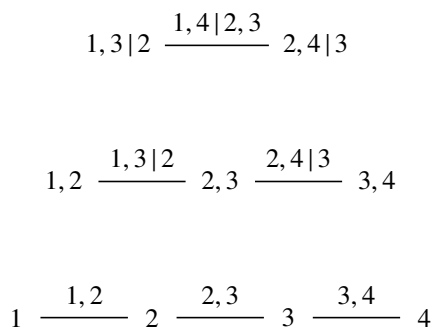
$$1,3\,|\,2 \quad \overline{\begin{array}{c} 1,4\,|\,2,3 \end{array}} \quad 2,4\,|\,3$$

$$1,2 \quad \overline{\begin{array}{c} 1,3\,|\,2 \end{array}} \quad 2,3 \quad \overline{\begin{array}{c} 2,4\,|\,3 \end{array}} \quad 3,4$$

$$1 \quad \overline{\begin{array}{c} 1,2 \end{array}} \quad 2 \quad \overline{\begin{array}{c} 2,3 \end{array}} \quad 3 \quad \overline{\begin{array}{c} 3,4 \end{array}} \quad 4$$

Figure 14. A four-dimensional regular vine.

If instead C_N was replaced by e.g., a Joe copula C_J, say with dependence parameter θ, the resulting function

$$\int_{-\infty}^{\infty} I\{x_3 \leq z_3\} C_J(\sigma_{1|3}\Phi^{-1}(z_1) - \rho_{1,3}x_3,$$
$$\sigma_{2|3}\Phi^{-1}(z_2) - \rho_{2,3}x_3; \theta)\phi(x_3)\,dx_3$$

would be a valid distribution function. The remaining components of the distribution would still be normal, but the copula of the conditional distribution of (Z_1, Z_2) conditioned on Z_3 would be a highly non-normal Joe copula. By the same types of steps, we could have changed the marginals, as well as the bivariate distributions of (Z_1, Z_3) and (Z_2, Z_3). The resulting distribution would be highly non-normal, even though it still would fulfill the algebraic similarity to the normal distribution by fulfilling the simplifying assumption, i.e., that the conditional distribution of Z_1, Z_2 given Z_3 has a copula that does not change with the value of Z_3.

In general, vine copulas can be formed by emulating the formulas for the expanded versions of the distribution of a multivariate normal distribution in terms of conditional and unconditional bivariate distributions, and then, while preserving the simplifying assumption, changing the normal copulas to copulas of our choosing. The only normal characteristic then remaining is the simplifying assumption. Without this assumption, the connecting of sequences of the conditional distributions would be much more complicated. In the tri-variate case, we could for example have expanded the distribution in terms of $(Z_1, Z_2), (Z_1, Z_3)$ and then the conditional distribution of (Z_1, Z_3) conditional on Z_2. In higher dimensions, there are many options for such expansions.

The expansion of a copula distribution into conditional and unconditional bivariate distributions can be represented by a sequence of trees, called a vine. A four-dimensional example of such an expansion is visualized in Figure 14. Let the resulting copula distribution be C, and let $(U_1, U_2, U_3, U_4)' \sim C$. Figure 14 shows a vine comprised by three trees.

The lowest tree specifies which unconditional bivariate distributions are to be specified by bivariate copulas. The figure shows that we are required to directly specify the bivariate distributions of (U_1, U_2), of (U_2, U_3) and of (U_3, U_4). The next tree knits together the copulas from the first tree, through saying that we are further to specify the copula of the conditional distribution of (U_1, U_3) given U_2, as well as the copula of the conditional distribution of (U_2, U_3) given U_3. These copulas are not to depend on the conditioning variables, thereby fulfilling the same type of simplifying assumption which we saw was fulfilled by the normal distribution. As in that case, the conditional marginal distributions will in general depend on the conditioning variables, but these conditional marginal distributions are consequences of the choices at the first level, and have therefore already been indirectly specified. The final level of the tree completes the linkage of a complete multivariate distribution, and requires the specification of a bivariate copula for the conditional distribution of (U_1, U_4) when we condition on the remaining variables, i.e., (U_2, U_3).

There are other four-dimensional vines, and with higher dimensions there is a great number of possible vine configurations. Each bivariate copula distribution represented in the figure is a copula distribution that we can choose freely, and the resulting distribution will be valid. In our applications for simplicity we considered only cases where every copula is of the same class, and fix the dependence parameters of the copulas separately in order to control the covariance matrix of the resulting full distribution once we attach marginal distributions to the copula. Since there are as many copulas as there are elements in a correlation matrix, this matching is often possible. This idea is the core idea of the VITA method of Foldnes and Grønneberg (2017), where the practical and computational details for achieving this match is discussed. Through certain simplifications originating from the simplifying assumption, simulation is easily implemented without the need for numerical integration, at least when using common copulas such as the Joe or Clayton copula, as shown in Joe (1996) and discussed in Dissmann et al. (2013); Schepsmeier et al. (2018). The R package covsim (Foldnes & Grønneberg, 2020) can be used for identifying a vine with dependency parameters so that for given marginal distributions and for a given sequence of bivariate copulas, their combination into a full vine distribution results in a distribution that has a pre-specified covariance matrix.

References

Bedford, T., & Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, *32*(1-4), 245–268.

Bedford, T., & Cooke, R. M. (2002). Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, *30*(4), 1031–1068.

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons.

Dissmann, J., Brechmann, E. C., Czado, C., & Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, *59*, 52–69.

Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed., Vol. 1). John Wiley & Sons.

Foldnes, N., & Grønneberg, S. (2020). covsim: Simulate from distributions with given covariance matrix and marginal information [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=covsim (R package version 0.1.0)

Foldnes, N., & Grønneberg, S. (2017). The asymptotic covariance matrix and its use in simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–16.

Foldnes, N., & Grønneberg, S. (2019). On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika*, *84*(4), 1000–1017.

Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *Psychometrika*, *82*(4), 1035–1051.

Grønneberg, S., Moss, J., & Foldnes, N. (2020). Partial identification of latent correlations with binary data. *Psychometrika*. (submitted)

Hofert, M., Kojadinovic, I., Maechler, M., & Yan, J. (2013). copula: Multivariate dependence with copulas [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=copula (R package version 0.999-7.)

Joe, H. (1996). Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, 120–141.

Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.

Nagler, T., & Vatter, T. (2019). rvinecopulib: High performance algorithms for vine copula modeling [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rvinecopulib (R package version 0.5.1.1.0)

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). New York: McGraw-Hill.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., & Erhardt, T. (2018). Vinecopula: Statistical inference of vine copulas [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=VineCopula (R package version 2.1.8)

Sklar, M. (1959). *Fonctions de repartition a n dimensions et leurs marges*. Université Paris 8.

Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.