Foldnes, N., & Grønneberg, S. (2019). On Identification and Non-normal Simulation in Ordinal Covariance and Item Response Models. Psychometrika, 84(4), 1000–1017. https://doi.org/10.1007/s11336-019-09688-z

ON IDENTIFICATION AND NON-NORMAL SIMULATION IN ORDINAL COVARIANCE AND ITEM RESPONSE MODELS

NJÅL FOLDNES AND STEFFEN GRØNNEBERG

Abstract. A standard approach for handling ordinal data in covariance analysis such as structural equation modeling is to assume that the data was produced by discretizing a multivariate normal vector. Recently concern has been raised that this approach may be less robust to violation of the normality assumption than previously reported. We propose a new perspective for studying the robustness towards distributional misspecification in ordinal models using a class of non-normal ordinal covariance models. We show how to simulate data from such models, and our simulation results indicate that standard methodology is sensitive to violation of normality. This emphasizes the importance of testing distributional assumptions in empirical studies. We include simulation results on the performance of such tests.

1. Introduction and summary

Empirical investigations in the social and behavioral sciences are often based on categorical data, which has been collected using ordinal scales (e.g., Likert-type scales). A popular method for modeling such data is to assume that a continuous latent variable underlies each categorical variable, so that the observations on each ordinal variable is the result of discretizing the corresponding continuous variable. In the context of covariance modeling such as structural equation modeling (SEM) and confirmatory factor analysis (CFA) this approach was initiated by **?** for the dichotomous case, with later expansions to the polytomous case (e.g., **??**).

In this article we investigate identifiability issues that arise from the assumption of an underlying random vector whose discretization produces the observed variables. Based on our identifiability findings we propose a new method of simulating ordinal data that allows true violation of the normality assumption. In the process, we propose a class of non-normal ordinal covariance models, whose estimation theory is not dealt with in this paper.

We are given n independent observations of an ordinal d-dimensional random vector $X = (X_1, X_2, \ldots, X_d)'$. In practice, each X_i may correspond to an item in a test or questionnaire, that is scored on an ordinal scale. We postulate an underlying continuous variable ξ_i that produces the observed variable X_i through discretization. Many aspects of the the underlying discretized vector $\xi = (\xi_1, \xi_2, \ldots, \xi_d)'$ are not identified. This means that there is a large class of distributions for ξ that will result in the exact same distribution for X, and we call members of this class of distributions discretize equivalent to ξ. One consequence is that a crucial assumption in ordinal SEM and CFA, namely that ξ is a multivariate normal vector, can not be consistently tested. Another consequence is that many simulation studies designed to address the robustness of model inference to violation of the normality assumption have generated data that only appear to violate the normality assumption, but that in fact is indistinguishable from discretizing a normal vector, as recently shown by **?**. This observation is the starting point for our paper: How should a proper simulation study outside the normality assumption be conducted?

We note that although the discussion in the present article deals almost exclusively with ordinal SEM and CFA, our findings extend also to the case of multidimensional IRT (**?**). The close relationship between IRT and CFA in terms of statistical procedures is well-established (**?**), and we provide extensions to this body of work in Appendix A.

This article is organized as follows. In Section 2 we review the ordinal covariance model discretization framework and establish identifiability results, and in Section 3 we summarize earlier results on testing for discretized normality. The results on identifiability lead us to Section 4, where we embed the normal theory ordinal covariance models into a larger model class which supports more general distributional assumptions. In Section 5 we then discuss how to simulate data for this model class, which enables us to investigate the robustness of conventional ordinal covariance models to violation of the underlying normality assumption in a controlled manner. Numerical illustrations are given in Section 6. We here also include a discussion and evaluation of a test of underlying normality that has been largely neglected in the literature. The present study points to the high importance of testing for underlying normality in empirical work. Concluding remarks are given in the last section.

2. On identifiability and normality in ordinal covariance models

Suppose each coordinate of X takes on $K > 1$ possible distinct values $x_1, x_2, \dots, x_K$. We assume further that X is the result of the discretization of a d-dimensional random vector ξ. In the following we refer to ξ as the *discretized variable.* Initially, we do not impose any further restrictions on ξ, and define X using the relation that for $i = 1, 2, \dots, d$ we have

$$X_i = \begin{cases} x_1, & \text{if } \tau_{i,0} < \xi_i \leq \tau_{i,1} \\ x_2, & \text{if } \tau_{i,1} < \xi_i \leq \tau_{i,2} \\ \vdots & \\ x_K, & \text{if } \tau_{i,K-1} < \xi_i \leq \tau_{i,K} \end{cases}$$

where $\tau_{i,0} = -\infty < \tau_{i,1} \leq \tau_{i,2} \leq \cdots \leq \tau_{i,K-1} \leq \tau_{i,K} = \infty$, and where $x_1 < x_2 \cdots < x_K$. Following **?**, a compact representation of each coordinate X_i of X for $1 \leq i \leq d$ is given by

$$X_i = \sum_{j=1}^{K} x_j I\{\tau_{i,j-1} < \xi_i \leq \tau_{i,j}\} \tag{1}$$

where we use that $|\xi_i| < \infty$ since ξ is assumed to be a random vector, and where $I\{A\}$ is the indicator function of A, i.e., it is one if A is true and zero otherwise. We next combine this discretization framework with covariance modeling to obtain the traditional normality-based ordinal SEM model.

Definition 1. *A normal ordinal covariance model has the data generating mechanism of eq.* (1)*, where ξ is assumed to be a multivariate normal vector with standard normal marginals, and a correlation matrix $\Sigma^\circ = \Sigma(\theta^\circ)$ where $\theta \mapsto \Sigma(\theta)$ is a covariance model.*

In the above definition the vector θ° contains the population values of the covariance model parameters, and Σ° is the population correlation matrix of ξ implied by θ°. Ordinal SEM/CFA as proposed by, e.g., **?**, is based on fitting the proposed structural model $\Sigma(\theta)$ to an estimate of the correlation matrix Σ° of ξ, the so-called *polychoric correlation matrix*. The first step in ordinal SEM estimation is therefore to estimate Σ°, which is only possible under additional assumptions concerning ξ. The model above makes the traditional assumption of multivariate normality of ξ, which allows Σ° to be estimated using normal theory maximum-likelihood (ML) estimation (**?**). This is the approach implemented by default in SEM software.

Researchers have been concerned with potential bias in the estimation of Σ°, should the normal ordinal covariance model not hold due to distributional misspecification. That is, when ξ is not multivariate normally distributed, the estimation of polychoric correlations may become biased, and the bias may propagate to parameter estimates and invalid inference for the structural model. Starting with **?**, the robustness of the normal ordinal covariance model to distributional misspecification have often been studied by discretizing a non-normal vector obtained through the approach of **?** (e.g., **??????????**). The consensus reached by these studies, is that the normality-based polychoric correlation estimator seems to be quite robust to violation of the underlying normality assumption. However, recently **?** showed that ordinal data stemming from discretizing a Vale-Maurelli (VM) vector is in most cases numerically equivalent to data stemming from discretizing a multivariate normal vector. Hence, these studies do not provide information about the robustness of the normal ordinal covariance model to distributional misspecification. This surprising finding is a consequence of the lack of identifiability of ordinal covariance models, combined with the fact that VM vectors in most cases have a normal copula (**?**).

This points to the importance of taking identifiability in ordinal covariance models more fully into account. We start with the observation that there are many vectors ξ which, when discretized according to eq. (1), lead to the same distribution for X. That is, the distribution of ξ is not fully identified based on the distribution of X.

Definition 2. *If a d-dimensional random vector $\tilde{\xi}$ were to be discretized with appropriate thresholds and the resulting ordinal vector, say, $\tilde{X}$, has the same distribution as X, then we say that $\tilde{\xi}$ is* discretize equivalent *to ξ.*

That $\tilde{\xi}$ and ξ are discretize equivalent means that it is impossible to distinguish $\tilde{X}$ and X statistically, since their distributions are equal. That is, the thresholds and the distribution of ξ are not identified, as they cannot be uniquely determined from the distribution of X. As we now show, the class of discretize equivalent distributions always contain many members: there are many combinations of thresholds and distributions of ξ that lead through eq. (1) to the same ordinal distribution.

Using the above definition, we may briefly summarize the investigation of **?** as follows: If the polynomials of the VM-transformation are monotonous, the VM-distributed random vector $\tilde{\xi}$ is discretize equivalent to a multivariate normal random vector ξ.

We next provide two lemmas and a proposition on discretize equivalent random vectors. These results do not make any assumptions on the distribution of X, other than X has a finite number of outcomes. The first lemma indicates that we should analyze the class of ordinal covariance models with caution, firstly as we can generate any discrete random vector with a finite number of outcomes by eq. (1), and secondly as ξ is always discretize equivalent to a purely discrete random vector. Having ξ as a discrete random vector is far removed from the multivariate normal case. The lemma is proved simply by self-discretizing X. The assumption of a finite number of outcomes is made for simplicity and can be avoided.

Lemma 1. (1) *Let X be a discrete random vector with a finite number of possible outcomes. Then there exists a random vector ξ such that eq.* (1) *is fulfilled.*

(2) *Suppose eq.* (1) *is fulfilled. There exist a purely discrete $\tilde{\xi}$ with the same number of possible outcomes as X that is discretize equivalent to ξ.*

Proof. See Appendix C. □

The premise and motivation for the class of ordinal covariance models since the time of **?** has been that ξ is a continuous random vector. The following proposition shows that the marginals of ξ can be taken to be standard normal, or, by a trivial extension, any other continuous univariate distribution.

Proposition 1. *There exists a continuous random vector $\tilde{\xi}$ with standard normal marginals that is discretize equivalent to ξ.*

Proof. See Appendix C. □

Remark 1. *Note that by Lemma 1 (1) and Proposition 1, any discrete random vector with a finite number of outcomes can be thought of as being discretized from a random vector ξ with normal marginals using eq.* (1).

Note that in the argument underlying Proposition 1, the thresholds in the representation of eq. (1) are changed, as allowed by the definition of discretize equivalent. This has the consequence that the proposition only applies to models where the thresholds are free parameters. To our knowledge, this applies to all known statistical models for eq. (1).

Proposition 1 implies that the marginal distributions of ξ are not identified, i.e., we cannot deduce the marginal distributions of ξ when only observing X – unless further restrictions on the distribution of ξ are imposed. This has been noticed before, e.g., by **?**, who argued that the marginals therefore can be taken as uniform on $[0, 1]$, though their argument assumes that ξ has continuous marginals with strictly increasing cumulative distribution functions, and our argument is general.

When ξ is assumed to have uniform marginals, its joint distribution C is known as a copula. Also the copula of ξ is not fully identified, see **?**. Indeed, for any copula $\tilde{C}$ with the same probabilities over rectangles in $[0,1]^d$ defined by the thresholds $(\tau_{i,j})$, we have that $\tilde{\xi} \sim \tilde{C}$ is discretize equivalent to ξ.

3. Testing for underlying normality

As we will see in our numerical illustrations in Section 6, statistical methodology assuming a normal ordinal covariance model may be less robust to deviations of underlying normality than reported in previous studies. Testing the normality of ξ is therefore of practical importance in empirical studies. Due to the above lack of identifiability, testing whether ξ is multivariate normal based on observations from X, means testing whether ξ is discretize equivalent to a normal random vector (this interpretation of tests of normality was also noticed by **?**).

To the best of our knowledge, only one test has been proposed in the literature for detecting underlying multivariate non-normality in an ordinal dataset. **?**, section 4.2 proposed a test statistic T which is still understudied, and whose only empirical evaluation is a small simulation study under two approximations reported by **?**. The test statistic is based on the discrepancy between the observed bivariate proportions in the sample and the probabilities implied by assuming that ξ is multivariate normally distributed. Let $k \neq l$ and denote by $p_{kl,ij}$ the number of observations in the sample with $X_k = x_i$ and $X_l = x_j$, divided by the sample size. Likewise, we can estimate the thresholds and the polychoric correlation between ξ_k and ξ_l (**?**), and calculate $\pi_{kl,ij} = P(\hat{\tau}_{k,i-1} < \xi_k \leq \hat{\tau}_{k,i}, \hat{\tau}_{l,j-1} < \xi_l \leq \hat{\tau}_{l,j})$, assuming that ξ_k and ξ_l are bivariate normal with standard normal marginals and a correlation equal to the polychoric correlation. Note that in the probability defining $\pi_{kl,ij}$, the parameters estimated from data and are treated as fixed, and their distributions are not included in the probability calculation. Let $r_{kl,ij} = p_{kl,ij} - \pi_{kl,ij}$ be the residual between the observed proportion and the proportion implied by normality. There are $K^2d(d-1)/2$ such residuals. **?** derived the following asymptotic distribution, provided ξ is multivariate normal:

$$T := n\sum r_{kl,ij}^2 \xrightarrow[n\to\infty]{d} \sum_{i=1}^{m} \lambda_i\chi_1^2, \tag{2}$$

where $m = (K^2 - 2K)d(d-1)/2$. The coefficients $\alpha_1, \dots, \alpha_m$ are the eigenvalues of the matrix

$$M = (I - \Delta G)\hat{\Gamma}(I - \Delta G)', \tag{3}$$

where I is the identity matrix, and Δ is a Jacobian matrix defined as $\partial\pi/\partial\kappa$, where π contains the model-implied bivariate proportions, and κ contains the thresholds and the polychoric correlations. The matrix G is such that $\sqrt{n}(\hat{\kappa} - \kappa_0) \overset{a}{=} G\sqrt{n}(p - \pi_0)$, where π_0 contains the true bivariate proportions (**?**, eq. 14).

It is important to note that there are various ways of approximating the distribution of T to obtain a p-value, see **?** for a thorough discussion. The small simulation study in **?** only included a mean-scaled and a mean-and-variance scaled approximation, but we deem it important to consider several approximations in order to best profit from the result in eq. (2). We therefore include not only the classical approximations, but also new developments proposed by **?**, which have yet been little evaluated in the literature. Hence, in Section 6, we evaluate four approximations to the limiting distribution. Two of these approximations are well-known, based on scaling (**?**) and scaling-and-shifting (**?**). In addition two approximations based on the recently proposed technique of eigenvalue block averaging (EBA) were evaluated (**?**). In full EBA we estimate the eigenvalues λ_j and obtain the p-value as

$$p_{\text{EBAF}} = P\left(\sum_{j=1}^{d} \hat{\lambda}_j Z_j^2 > T\right),$$

while in the split-half approach we sort the eigenvalues and split them at the median. In each of the two halves, we replace the eigenvalues with their group-based average to obtain the p-value as

$$p_{\text{EBAH}} = P\left(\sum_{j=1}^{\lceil d/2\rceil} \tilde{\lambda}^1 Z_j^2 + \sum_{j=\lceil d/2\rceil+1}^{d} \tilde{\lambda}^2 Z_j^2 > T\right),$$

where $\tilde{\lambda}^1 = \frac{1}{\lceil d/2\rceil}\sum_{j=1}^{\lceil d/2\rceil}\hat{\lambda}_j$ and $\tilde{\lambda}^2 = \frac{1}{d-\lceil d/2\rceil}\sum_{j=\lceil d/2\rceil+1}^{d}\hat{\lambda}_j$.

4. A non-normal ordinal covariance model

We now turn to the problem of simulating from a non-normal ξ that is then discretized into X. A central aim in conducting such simulations is to assess the performance of normal-theory methods for estimating the model in Definition 1, when ξ is in fact non-normal. For instance, one might study the bias in polychoric correlation estimates based on the popular two-step method of **?** in conditions where the discretized vector is truly non-normal.

The first step in identifying the types of distributions to simulate from is to extend the normal ordinal covariance model in Definition 1 to a model which supports non-normality. That is, we wish to define a model we may call a non-normal ordinal covariance model. The difficulty in identifying a proper extension to embed the normal ordinal covariance model into is that the marginals are not identified, see Proposition 1.

At a minimum, the model class should allow for non-normal ξ whose covariance matrix equals $\Sigma(\theta^\circ)$. Since the covariance between two random variables depends on the marginals as well as the copula of the variables, the choice of marginals will influence the meaning of the covariance matrix. In order that the normal ordinal covariance model is to be a special case of the non-normal ordinal covariance model, we fix the marginals to standard normal. Another less technical motivation for assuming normal marginals may be given on a priori grounds, see Appendix B.

Definition 3. *A non-normal ordinal covariance model (with normal marginals) fulfils eq.* (1)*, where ξ is assumed to have standard normal marginals and a correlation matrix $\Sigma^\circ = \Sigma(\theta^\circ)$ following a covariance model $\theta \mapsto \Sigma(\theta)$.*

By the Cauchy-Schwarz inequality, the covariance matrix of ξ always exists, since ξ has standard normal marginals and hence finite univariate moments of all orders.

Since the above model class is considerably larger than the normal ordinal covariance model, the problem of identifiability is also more complex. Indeed, while the marginals and covariance matrix are given, the copula of ξ is free to vary, meaning that the distribution of ξ – and therefore also X is only partly specified. There may be a large class of copulas which when joined with normal marginals yield the desired covariance matrix. We here use "model" in a rather loose sense: A non-normal ordinal covariance model does not completely specify the probability distribution of X, but instead specifies a space of probability distributions.

We note that also Σ° and θ° are not in general identified, i.e., cannot be deduced from the distribution of X – unless further restrictions on the distribution of ξ are imposed. If an estimation theory is to be developed for this model class, one either has to impose further restrictions on the distribution of ξ and thereby gaining identifiability, or one could analyze this model class in terms of partial identification (see e.g., **??**). The approach of partial identification would then not estimate Σ° or θ°, but instead identify sets which contain these parameters. We consider this issue outside the scope of this paper, in which we focus on simulation.

While Proposition 1 shows that the marginal distributions of ξ are not identified, this argument does not take into account potential knowledge of the covariance matrix of ξ belonging to the space of covariance matrices given by $\theta \mapsto \Sigma(\theta)$. We leave this issue open to further research.

To assess the distributional robustness of normal-theory methods in normal ordinal covariance models, we may simulate from a non-normal ordinal covariance model and assess how these methods perform.

5. Simulating from the non-normal ordinal covariance model using the VITA method

In order to simulate from a non-normal ordinal covariance model, we must discretize a random vector with normal marginals and a fixed covariance matrix Σ°. To the best of our knowledge, currently only the VITA simulation method of **?** is capable of constructing such random vectors.

Briefly stated, the VITA method identifies a so-called vine copula distribution whose covariance matrix under chosen marginals equals a target covariance matrix. Vine copula distributions are made up of a sequence of bivariate copulas, known as pair copulas, and are combined through a sequence of tree structures in a manner that always yield a valid high dimensional copula distribution. See **?** and **?** for more details on vines. While having marginals and the covariance matrix fixed, the VITA method allows the specification of the mentioned tree-structure and its pair copula classes. A large class of distributions fulfilling the required restrictions on the marginals and covariance matrix can be obtained in this manner.

In general, after having identified a simulation method for a non-normal random vector whose marginals are standard normal and whose covariance matrix equals Σ°, one can use this to generate a whole class of non-normal random vectors with varying degrees of non-normality and with standard normal marginals and covariance matrix Σ° (**?**, Section 3.1). Indeed, let us denote a multivariate normal vector whose covariance matrix equals Σ° by Z. And let us denote by V a VITA vector, generated independently from Z, whose covariance matrix is Σ°, and whose marginal distributions are standard normal. Then, for any $0 \leq \alpha \leq 1$, the vector

$$\xi = \sqrt{1-\alpha} \cdot Z + \sqrt{\alpha} \cdot V \tag{4}$$

has covariance matrix Σ° and standard normal marginals. By letting α run from 0 to 1, the generated vector ξ violates the underlying normality assumption to a higher and higher degree. When $\alpha = 1$ we arrive at a pure VITA vector V.

In simulation studies, after having identified the non-normal ξ that is to be discretized, either by using VITA or some other method yet to be proposed, it is important to test whether ξ is discretize equivalent to the multivariate normal. Since the marginals of ξ are fixed to normal in Definition 3, this will not happen in trivial ways, as happened in the VM simulation method examples discussed in **?** where the copula of ξ is exactly normal. Still, we recommend simulating a large sample from ξ, and test its discretized vector X for underlying normality using the test of **?** discussed in Section 3 and illustrated below.

6. Illustration of ordinal data simulation

To illustrate ordinal data simulation with proper violation of underlying normality, consider a two-factor model where the first factor has two indicators ξ_1 and ξ_2, while the second factor has three indicators ξ_3, ξ_4 and ξ_5. The structural parameters are five factor loadings $\lambda_1, \dots \lambda_5$ and the interfactor correlation ϕ. We fix these parameters to the following population values: $\theta^\circ = (0.95, 0.95, 0.95, 0.95, 0.95, 0.9)'$. That is, in the population the factor loadings are 0.95 and the interfactor correlation is 0.9. Each factor has unit variance. The implied covariance matrix of the discretized vector ξ is then

$$\Sigma^\circ = \begin{pmatrix} 1 & & & & \\ 0.902 & 1 & & & \\ 0.812 & 0.812 & 1 & & \\ 0.812 & 0.812 & 0.902 & 1 & \\ 0.812 & 0.812 & 0.902 & 0.902 & 1 \end{pmatrix}.$$

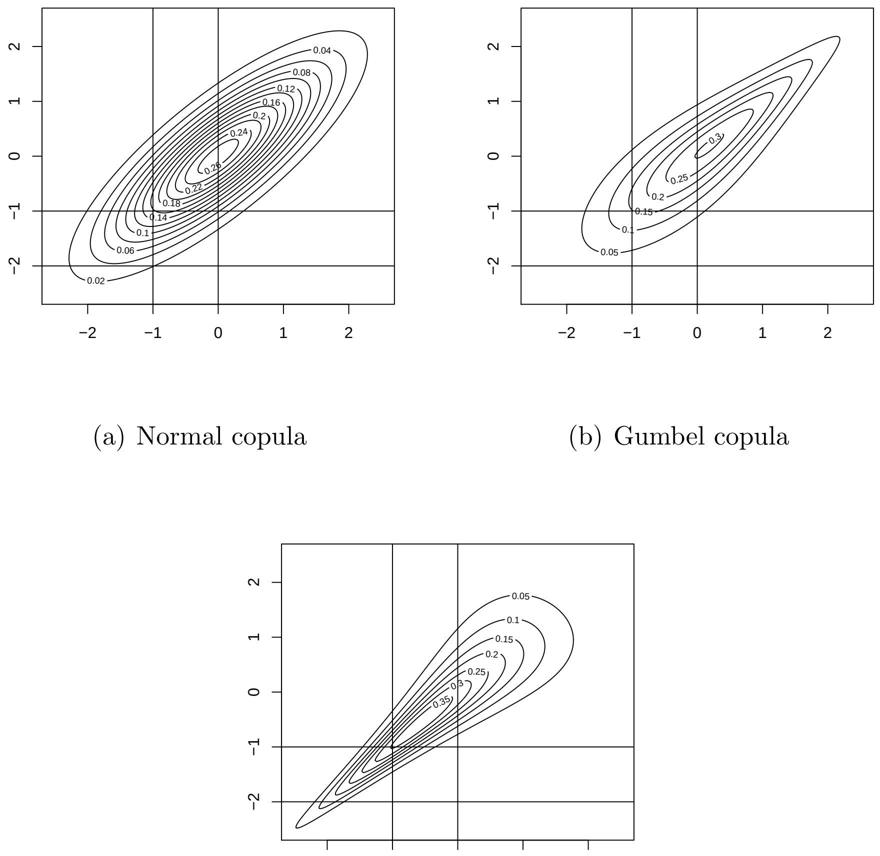
Our goal is to simulate ξ that matches this covariance matrix. Moreover, in accordance with the non-normal ordinal covariance model from Definition 3, each marginal is to be standard normally distributed: $\xi_i \sim N(0, 1)$ for $i = 1, \dots, 5$.

We are interested in how the polychoric estimates and subsequent model inference are affected when ξ violates the underlying normality assumption. In case non-normality has a deteriorating effect on these outcomes, we are also interested in investigating to what degree we can detect violation of the underlying normality assumption. Therefore, we will also evaluate the performance of the underlying normality test of **?**.

To investigate how sensitive the estimation of polychoric correlations and model parameters are to violation of normality, we will follow the interpolation method described in Section 5 to simulate under a sequence of conditions that interpolate between multivariate normality at one end, and a distinctively non-normal VITA condition at the other.

In the present study, we considered two such distinctively non-normal distributions as end conditions for our interpolation, each obtained using the VITA methodology of **?**. Note that the above covariance model was also studied in Section 3.1 in **?**, and we here use the same tree structure as in that paper. The choices of correlations and distributions were in that paper made to illustrate the effect of a high level of non-normality, and this is also the case here. The results we now report will therefore reflect a scenario of high non-normality in a highly correlated setting. A more complete and systematic simulation study with varying degrees of correlations and non-normality should be undertaken in future research.

Given the large class of non-normal copulas, we here included two VITA vectors that were based on different pair-copulas. Using exclusively **?** pair-copulas resulted in the regular vine here referred to as V_G. The second VITA vector V_C was based on using **?** pair-copulas to construct the regular vine. We emphasize that both V_G and V_C are random non-normal vectors of dimension 5 with standard normal marginal distributions and covariance matrix Σ°. Although both the Gumbel and Clayton copulas belong to the class of Archimedean copulas, their corresponding VITA vectors represent different kinds of non-normality. For instance, the Clayton copula captures lower tail dependence, while the Gumbel copula exhibits strong upper tail dependence, and we expect these characteristics to be partially reflected in their respective regular vines. We may illustrate these differences by restricting ourselves to the bivariate case. Figure 1 displays the contour plots for three bivariate distributions, all of which have a correlation of 0.812 and all having standard normal marginals, but with different copulas. We also included in the contour plots thresholds $\tau_{1,1} = -1, \tau_{1,2} = 0$



(a) Normal copula

(b) Gumbel copula

(c) Clayton copula

Figure 1. Three bivariate distributions with correlation 0.812 and standard normal marginals. The vertical and horizontal lines represent thresholds.

for ξ_1 and $\tau_{2,1} = -2, \tau_{2,2} = -1$ for ξ_2. Table 1 contains all five sets of thresholds used to obtain ordinal data X by discretizing ξ.

This yields $K = 3$ possible values for each of $X_1, \dots, X_5$, whose marginal distributions are given in Figure 2. To illustrate the difference in distributions when discretizing, we may again consider the bivariate case depicted in Figure 1. The thresholds of ξ_1 and ξ_2 illustrated in

	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
τ_1	-1.00	-2.00	-1.00	0.00	0.00
τ_2	0.00	-1.00	1.00	1.00	1.00

TABLE 1. Thresholds for discretizing $\xi_1, \dots, \xi_5$.

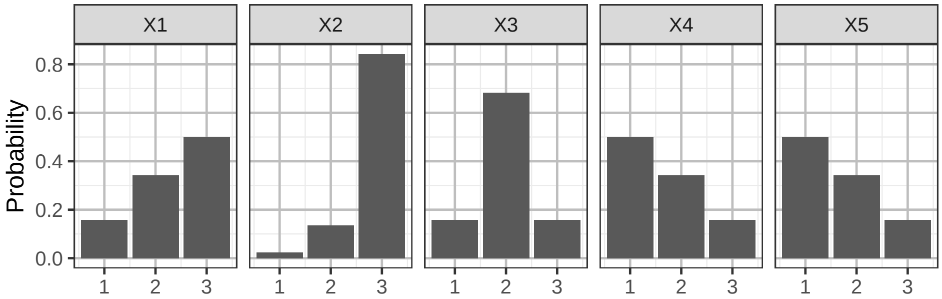


FIGURE 2. Marginal distributions of the ordinal variables $X_1, \dots, X_5$.

	Normal copula			Gumbel copula			Clayton copula		
	X_2			X_2			X_2		
	1	2	3	1	2	3	1	2	3
	0.021	0.078	0.059	0.018	0.072	0.069	0.023	0.108	0.027
X_1	0.002	0.053	0.287	0.004	0.056	0.281	0.000	0.027	0.314
	0.000	0.005	0.495	0.000	0.008	0.491	0.000	0.001	0.499

TABLE 2. Probability tables for (X_1, X_2) obtained by discretizing the distributions in Figure 1.

Figure 1 were used to discretize the normal, the Gumbel and the Clayton bivariate distributions in Figure 1, with the resulting probability tables given in Table 2. Note that the row and column sums of the tables are equal (up to rounding error) across the three distributions. However, the different copulas imply different pairwise probabilities in the three contingency tables.

The simulation design was as follows. For $\alpha = 0, 0.1, 0.2, \dots, 0.9, 1$ we simulated ξ for both $V = V_G$ and $V = V_C$, and at three different sample sizes: $n = 100, 300$ and $n = 1000$. This results in $3 + 10 \cdot 2 \cdot 3 = 63$

conditions, in each of which 2000 samples were generated. For each such sample we estimated

- the polychoric correlations using the method of **?**.
- the model parameters ($\hat{\lambda}_i$ for $i = 1, \ldots, 5$ and the interfactor correlation $\hat{\phi}$) using diagonally weighted least squares (DWLS) estimation based on the polychoric correlation matrix.
- the p-value of the test of correct CFA model, using the scaled-and-shifted statistic. Note that the scaling and shifting in this case is applied applied to n times the DWLS fit function used to estimate the model parameters.
- the p-value under four approximations of the **?** test of underlying non-normality, namely the mean-scaled test (S), the scaled-and-shifted test (SS), and the full and two-block EBA tests (EBAF and EBAH). Note that these are approximations to the test statistic in eq. (2).

Data generation was conducted in the R computing environment (**?**) with the help of the VineCopula package (**?**) and model estimation was conducted using the lavaan package (**?**).

6.1. **Results.** In the following we mostly exclude the results for the intermediate sample size $n = 300$ to simplify the presentation. Sample size was not found to affect the estimation of polychoric correlations or model parameters to a large degree. Figure 3 contains the mean values of the estimated polychoric correlations. As expected, under multivariate normality ($\alpha = 0$) the polychoric estimator is unbiased. Moving away from normality by letting α increase is associated with larger and larger bias in the polychoric estimator. Finally, when $\alpha = 1$ we reach the distributions V_G (upper panels) and V_C (lower panels). It is clearly seen that while some polychoric correlations are rather robust to the underlying non-normality (e.g., at $n = 1000$ we have $\overline{\hat{\rho}}_{14} = 0.809$ under V_G, close to the population value of $\rho_{14} = 0.812$), other polychoric correlations are severely biased at $\alpha = 1$ (e.g., at $n = 1000$ we have $\overline{\hat{\rho}}_{23} = 0.957$ under V_C, not close to the population value of $\rho_{23} = 0.812$). Also, it is noteworthy that the polychoric estimator is sensitive to the kind of underlying non-normality as represented by

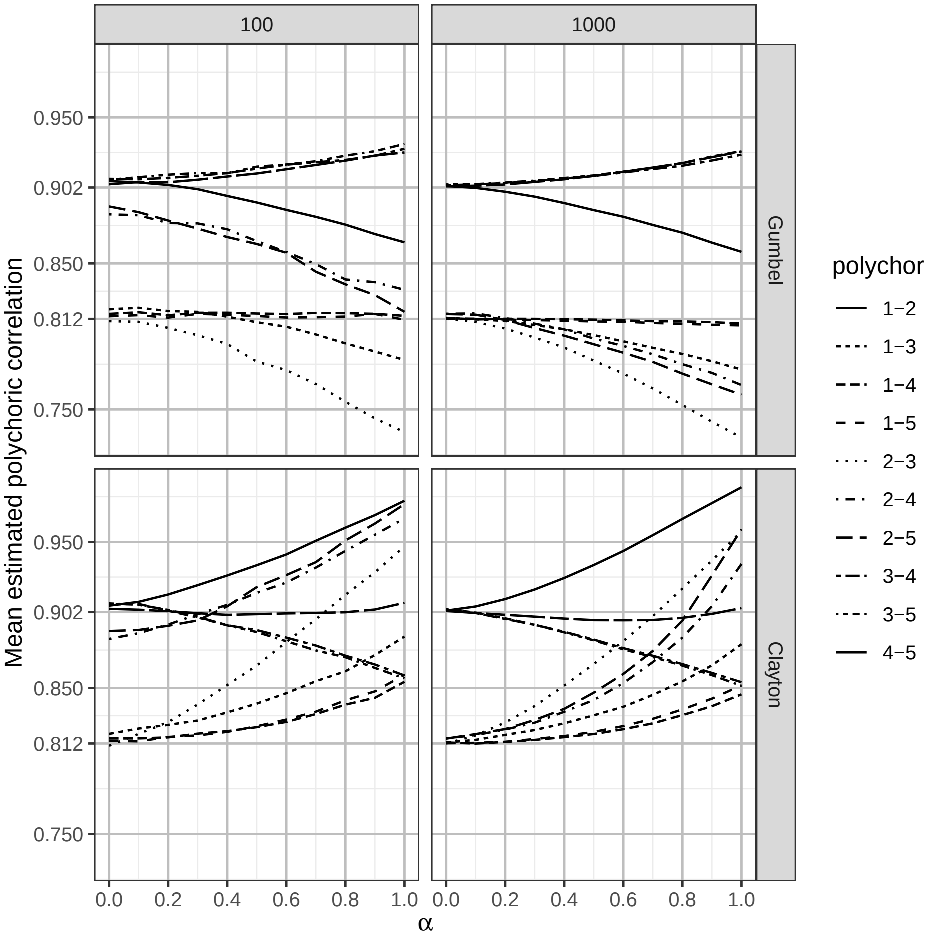


Figure 3. The mean of estimated polychoric correlations. i-j refers to the polychoric correlation ρ_{ij} between ξ_i and ξ_j.

V_G and V_C. For instance, for the polychoric correlation between ξ_2 and ξ_3, we have for $n = 1000$ that $\bar{\hat{\rho}}_{23} \approx 0.957$ under the Clayton VITA, compared to $\bar{\hat{\rho}}_{23} \approx 0.731$ under the Gumbel VITA. In other words, we find that the polychoric estimator is severely biased for some pairs of variables, under both V_G and V_C, and that the bias is in opposite directions.

In Figure 4 are depicted the mean of DWLS model estimates as we move from the multivariate normal case ($\alpha = 0$) towards the VITA distributions V_G and V_C. Given that DWLS estimation is based on polychoric correlations, which are increasingly biased as α increases, it is not surprising to see this reflected in the model estimates. However, of the six model parameters, four remain close to their population values, despite increasing non-normality in the underlying vector, under both V_C and V_G. In a sense, it seems that the distributional misspecification is absorbed as estimation bias for the two remaining parameters, λ_2 and ϕ. Again we see that the nature of the underlying non-normality strongly affects the bias. Under V_G both $\hat{\lambda}_2$ and $\hat{\phi}$ have a negative bias, with $\overline{\hat{\lambda}}_2 = 0.898$ at $n = 1000$ compared to the population value $\lambda_2 = 0.95$, and $\overline{\hat{\phi}} = 0.871$ compared to the population value $\lambda_2 = 0.9$. The bias under V_C is even more pronounced, although in the opposite direction: $\overline{\hat{\lambda}}_2 = 1.043$, and $\overline{\hat{\phi}} = 0.968$.

To study to what degree the underlying non-normality affects the test of correct model specification, we depict in Figure 5 the rejection rate at the 5% significance level of the scaled-and-shifted statistic, which is the default in lavaan under DWLS estimation. Previous studies (**??**), have reported that this test tends to underreject a correctly specified model, and this is confirmed in our findings for $n = 100$ and $n = 300$, where $\alpha = 0$ corresponds to correct model and distributional specification. At sample sizes $n = 100$ and $n = 300$ the test of model fit is only moderately affected by underlying non-normality. At the largest sample size, $n = 1000$, at $\alpha = 1$, the correctly specified covariance model is rejected in 84% of the V_C samples, and in 26% of the V_G samples, when estimated using normal theory estimators.

Given the effect of underlying non-normality on the polychoric correlations and model inference depicted in Figures 3-5, we next proceed to investigate whether the underlying non-normality is detectable. As expected, the power to detect non-normality generally increases with increasing α. The rejection rates of the four approximations to the test statistic of **?** are shown in Figure 6. It is clear that only two of the approximations are able to properly control Type I error, namely EBAF and SS, with EBAF Type I error slightly superior to that of SS.

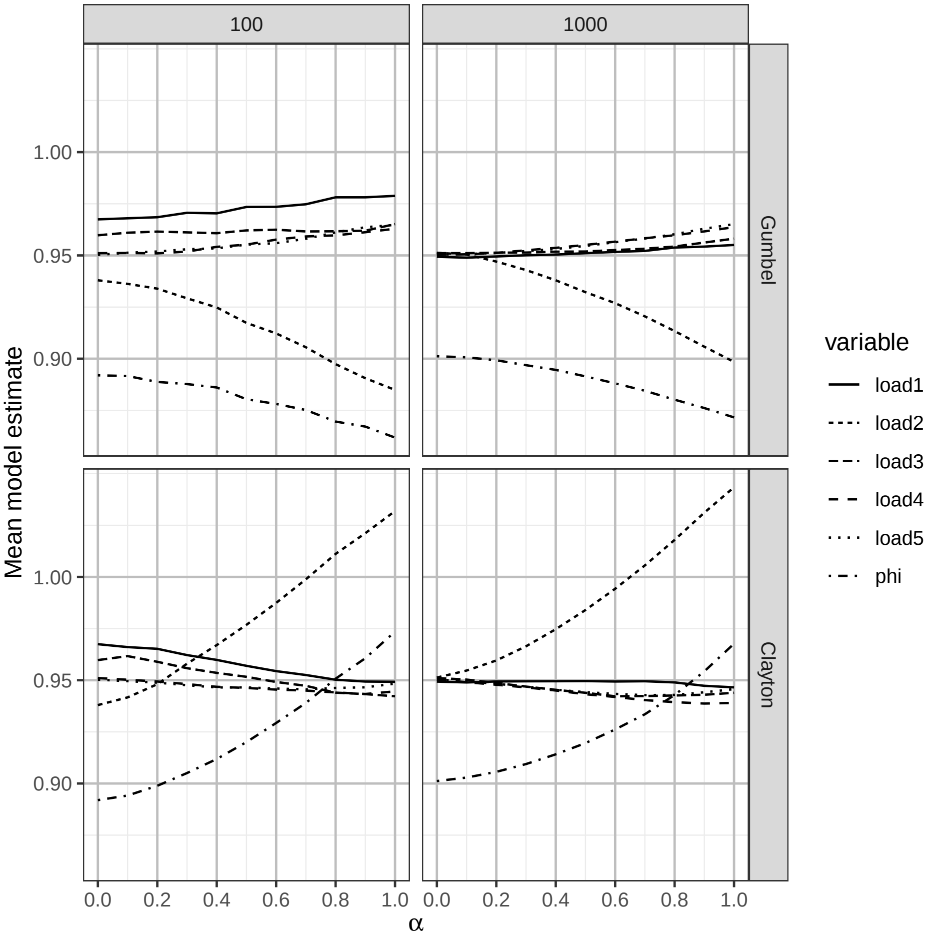


FIGURE 4. Mean of estimates for five factor loadings and the intrafactor correlation. load1-load5= factor loadings $\lambda_1, \ldots, \lambda_5$. phi= the intrafactor correlation ϕ.

Under interpolation toward the Gumbel distribution V_G the tests have very low power to detect the increasing underlying non-normality, unless the sample size is $n = 1000$. The statistics again differ between V_G and V_C, and although the tests show poor power to detect underlying non-normality of the Clayton VITA distribution at $n = 100$, the power significantly increases at $n = 300$, especially as α approaches 1.

6.2. **Discussion of results.** We have seen that polychoric correlations, and therefore also model estimates and goodness-of-fit tests, are

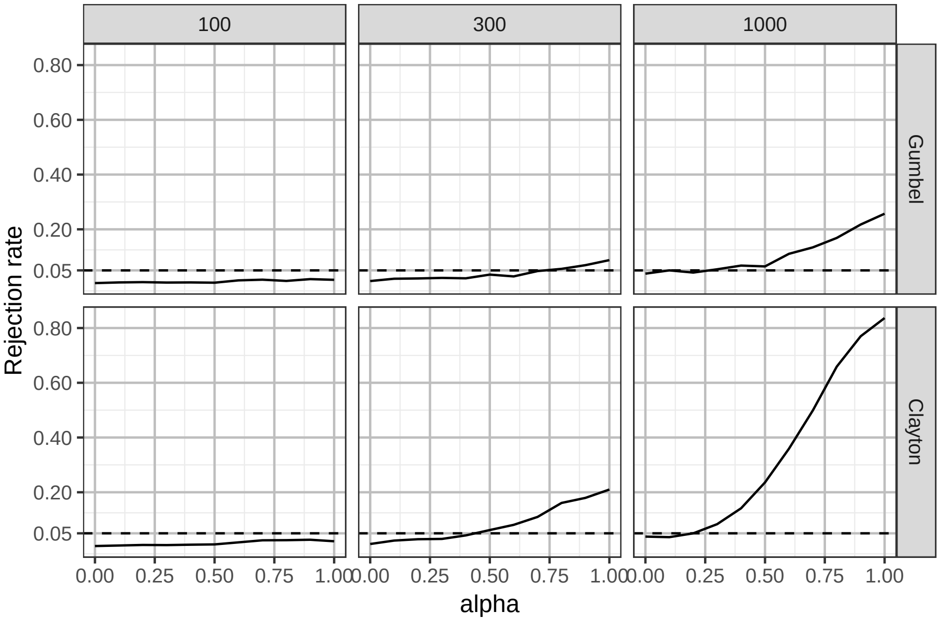


FIGURE 5. Rejection rate at the 5% significance level of the DWLS scaled-and-shifted test of correct model spesification.

affected by underlying non-normality. We have demonstrated that the type of non-normality (Gumbel VITA versus Clayton VITA) has a pronounced effect on the direction and magnitude of bias introduced by non-normality. For polychoric correlations and model estimates we saw this manifested particularly in statistics related to variable X_2 (e.g., ρ_{12}, ρ_{23} and λ_2). We believe that this is related to the fact that X_2 is the most asymmetrical of the ordinal variables, see Figure 2. Although we deem this topic outside the scope of the present illustration, we conjecture that asymmetrical ordinal distributions combined with tail dependence in the corresponding variables may be particularly detrimental to the performance of the polychoric estimator. Also, asymmetrical copulas were not used in our illustrations, but may have a pronounced effect on parameter estimates.

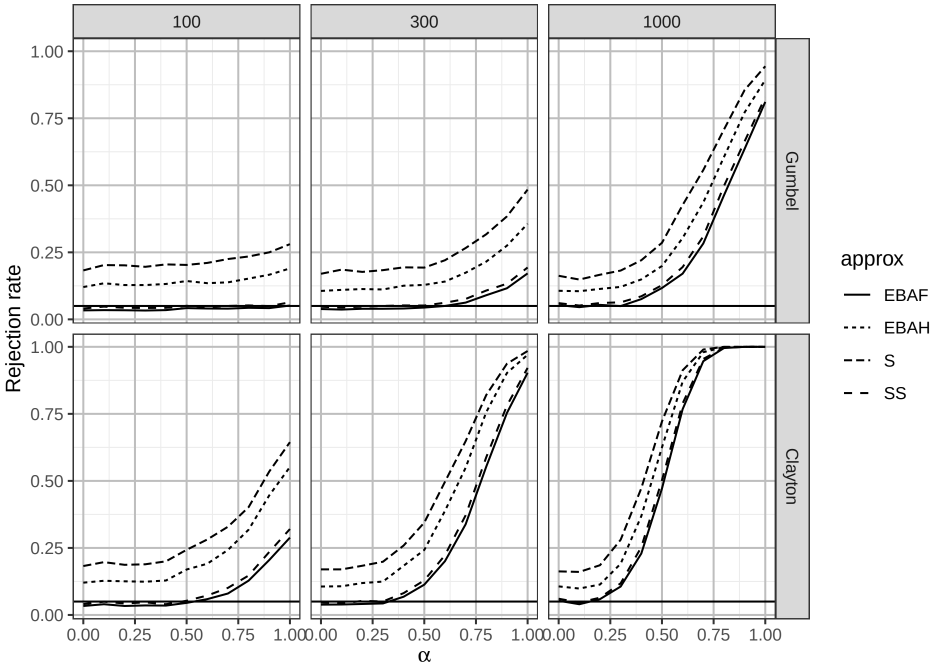


FIGURE 6. Rejection rates at the 5% significance level for four tests of underlying non-normality. EBAF=full eigenvalue block-averaging test. EBAH = eigenvalue block-averaging with two blocks. S= Mean-scaled test. SS= Scaled-and-shifted test.

7. CONCLUDING REMARKS

? recently reported that many influential studies on the robustness of ordinal SEM against underlying non-normality employed simulation methods that were equivalent to discretizing a multivariate normal random vector. Therefore, these investigations did not in fact study robustness against underlying non-normality. The degree to which non-normality influences polychoric estimates and related quantities is therefore an understudied problem that deserves further study, as the present paper only presents results from a limited simulation design.

That this surprising finding has not been detected before may be due to quite subtle identifiability issues that arise when assuming that the ordinal data at hand was produced by discretizing some underlying vector. The purpose of the present article was to shed light on this issue, and to formulate a non-normal ordinal covariance model that may serve as a basis for future empirical investigations into the robustness of ordinal SEM, CFA and IRT to non-normality of the discretized vector. It was demonstrated how one may simulate ordinal data based on this model that properly violated the normality assumption. The numerical results of this simulation study showed that non-normality embedded in the discretized vector affected polychoric correlation estimates, model parameter estimates and model fit statistics, introducing more substantial bias than previously reported. In addition, the specific type of non-normality embedded in the discretized vector was shown to affect both size and the direction of this bias.

Given these findings, it is important for users of ordinal SEM and CFA to try to detect whether the underlying normality assumption is plausible for their data. In the final simulation study we evaluated a test statistic for underlying non-normality, and found it to have rather poor power at small and moderate sample sizes. Further work is needed to develop tests with improved performance. Also, our simulation study was of limited scope, studying a low dimensional model with very high correlations. A systematic and more complete simulation study ought to be undertaken.

Acknowledgements: We thank the reviewers for several useful comments that improved the manuscript, and especially the reviewer who noticed Remark 1. The simulations were performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR).

Appendix A. On dichotomous multidimensional IRT models

For simplicity we limit the discussion to the dichotomous IRT case. We derive a stochastic representation of the IRT model under weak assumptions, which to our knowledge is a new result, and this representation immediately shows that IRT models are of the form of eq. (1). This representation is then applied to analyze how marginal assumptions (usually called link functions) in the IRT models transfer to the present discussion.

Assumption 1. *Consider a random vector $X = (X_1, X_2, \ldots, X_d)'$ where each coordinate takes on the value 0 or 1.*

1. *There is a p-dimensional random vector f which is such that for $i \neq j$ we have that X_i and X_j are independent conditional on f.*
2. *We assume for $i = 1, 2, \ldots, d$ that $\pi_i(f) := P(X_i = 1|f) = H(\zeta_i)$ where ζ_i is a function of f, and H is a CDF with density h with respect to Lebesgue measure.*

A standard assumption (**?**) is that $f \sim N(0, I)$ and that

$$\zeta_i = \alpha_{i,0} + \sum_{j=1}^{d} \alpha_{i,j} f_j. \tag{5}$$

This implies that $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_d)' \sim N(\mu, \Sigma)$ for some μ and Σ which are functions of the $(\alpha_{i,j})$ parameters. The link function H is typically assumed to either be the normal CDF, or the logistic CDF.

Let $Z = (Z_1, Z_2, \ldots, Z_d)$ consist of IID random variables with marginal distribution H, and Z is independent from $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_d)$, where ζ is defined in Assumption 1 (2). The proof of the following result is given in the online supplementary material.

Proposition 2. *A stochastic representation of X fulfilling Assumption 1 is $X = (I\{\xi_1 \leq 0\}, \ldots, I\{\xi_d \leq 0\})'$ where $\xi = Z - \zeta$.*

Since $Z_1, \ldots, Z_d$ are IID and independent to ζ, we have $\text{Cov}\,(\xi) = \text{Cov}\,(Z - \zeta) = \text{Cov}\,(Z) + \text{Cov}\,(-\zeta) = \sigma_Z^2 I + \text{Cov}\,(\zeta)$ where $\sigma_Z^2 = \text{Var}\,(Z_1)$ and I is the identity matrix. This simple correspondence means that the covariance structure of Z is that of ζ, except for changes in the variances. However, the choice of H influences the marginals of ξ, and the mathematical definition of the covariance of ξ depends on both the marginals and the copula of ξ. Hence, H plays a major role in the interpretation of the covariance of ξ, since it dictates at what "scale" the covariance model is to be interpreted. When ζ is multivariate normal, ξ will not be multivariate normal unless H is a normal CDF. Indeed, copulas are not preserved under marginal convolution, so that not even the copula of ξ is normal when H is not a normal CDF. This means that when ζ follows a normal covariance model but when H is not the normal CDF, the resulting IRT model does not follow even a non-normal covariance model (with respect to the covariance model of ζ) as defined in Definition 3, since we there insist that the marginals are standard normal.

Consider the popular choice of H given by the logistic CDF. Then ξ is not multivariate normal even when ζ is multivariate normal. Also, ξ will not have a normal copula. While ξ does have the same covariance matrix as ζ, the covariance matrix is given at a scale where the marginals F_{ξ_i} are convolutions between a logistic and a normal distribution. Since the marginal distributions are not identified when observing only copies of X, it seems difficult to interpret what the covariance matrix of ζ means. If the marginals are transformed to standard normal, one would instead of ξ study the discretize equivalent variable $\tilde{\xi} = (\Phi^{-1}F_{\xi_1}(\xi_1), \ldots, \Phi^{-1}F_{\xi_d}(\xi_d))'$, whose covariance is neither $\operatorname{Cov}(\xi)$ nor has a simple relation to $\operatorname{Cov}(\zeta)$. Finally, using arguments given in the upcoming Appendix B, a more natural a priori class of marginals for ξ is often normal, and not the convolution of a logistic and a normal.

Appendix B. An a priori justification for marginal normality of ξ that may be plausible in certain applications

We assume that the continuous discretized vector ξ have a covariance matrix obtained from a SEM model, that is, certain equations among latent variables are to hold, and these equations have error terms that fulfil certain restrictions in terms of correlation. The covariance model $\theta \mapsto \Sigma(\theta)$ for ξ is therefore motivated independently of the distributional class of ξ. Now, in many psychometric settings a central limit theorem argument can be used to make an a priori assumption of normality of ξ plausible. Indeed, let us suppose that ξ can be written as a sum of N random vectors $\xi_N^{(1)}, \ldots, \xi_N^{(N)}$. Under mild conditions, the simplest being that $\xi_N^{(i)} = \varepsilon_i/\sqrt{N}$ where $\varepsilon_1, \ldots, \varepsilon_N$ are IID random vectors, the multivariate distribution of ξ is close to that of a multivariate normal when $N \to \infty$ by a central limit theorem. If this approximation is very good, then the normal theory ordinal model in Definition 1 is appropriate. However, the quality of the approximation need not be very good for finite N, especially when the dimensionality d is high, which is the case in many applications of ordinal covariance models: Indeed, consider the ordinal confirmatory factor analysis model underlying many standard measurement instruments in empirical psychology, containing hundreds of items. In these cases, the marginal distributions of ξ may still be close to normal, since each marginal distribution is not affected by the relation between N and d, but the full distribution of ξ may be far from normal. If the marginals are close to normal but the full distribution is not, then the copula of ξ is not close to normal, and we have marginal normality but not joint normality. This may in certain cases make the marginal normality of ξ plausible, while the full copula of ξ is not normal.

Appendix C. Proofs for Section 2

Proof of Lemma 1. Self-discretize X, i.e., let $\tilde{\xi} = X$ and apply the transformation in eq. (1). The thresholds can be chosen in such a way that the discretization transformation is the identity transformation. The discretized version of X is then equal to X, which clearly has the same distribution as X, as required by discretize equivalence. □

We need the following preliminary lemma to prove Proposition 1.

Lemma 2. *There exists a continuous random vector $\tilde{\xi}$ which is discretize equivalent to ξ.*

Proof. We here only give a compressed version of the argument. The online supplementary material contains a detailed verification of technical details. By Lemma 1, we may without loss of generality assume that $\xi = X$. Define $x_0 = x_1 - 1$, and let $\mathcal{Q} = \{x = \otimes_{l=1}^{d}(x_{j_l}, x_{j_l+1}] : j_l \in \{0, 1, \ldots, K-1\}$ for $l = 1, 2, \ldots, d\}$ contain the hyper-rectangles contained between the points of the support S_X^d of X. Let $Q_1, Q_2, \ldots, Q_N$ be the sets in $\mathcal{Q}$, and note that they are disjoint. We now define a density $\tilde{f}$, which smears the probability that X is in Q_i uniformly over each Q_i. I.e., we let $\tilde{f}(x) = \sum_{i=1}^{N} \frac{P(X \in Q_i)}{V_i} I\{x \in Q_i\}$, where $V_i = \int_{\mathbb{R}^d} I\{x \in Q_i\}\,\mathrm{d}x \neq 0$ for $i = 1, 2, \ldots, N$, and $I\{A\}$ is the indicator function of A, which is one if A is true and zero otherwise. Let $\tilde{\xi}$ have $\tilde{f}$ as density. Then $\tilde{\xi}$ has the same probability as ξ (i.e., X) of being within the thresholds defined by the limits of the rectangles in Q_k for $k = 1, 2, \ldots, N$, completing the proof. □

Proof of Proposition 1. By Lemma 2, we may assume that ξ is a continuous random vector. This implies that any marginal cumulative distribution function F_i is continuous and increasing. Since it is illustrative, we here give a proof that assumes that F_i is also strictly increasing. A proof of this special case is also given in **?** (see their eq. (12)), and our argument follows closely Section 3 in **?**. The general case, which appears to be new, is proved in the online supplementary material. Since $F_i(\xi_i)$ is uniform on

$[0,1]$ we have that $\Phi^{-1}(F_i(\xi_i))$ is standard normal, where Φ^{-1} is the quantile function of the standard normal distribution. Since both F_i and Φ^{-1} are strictly increasing, so is $\Phi^{-1} \circ F_i$. For each coordinate X_i of X, we may therefore apply $\Phi^{-1} \circ F_i$ to each part of the inequalities defining X_i, and get $X_i = \sum_{j=1}^K x_j I\{\tau_{i,j-1} < \xi_i \leq \tau_{i,j}\} = \sum_{j=1}^K x_j I\{\Phi^{-1}(F_i(\tau_{i,j-1})) < \Phi^{-1}(F_i(\xi_i)) \leq \Phi^{-1}(F_i(\tau_{i,j}))\} = \sum_{j=1}^K x_j I\{\tilde{\tau}_{i,j-1} < \tilde{\xi}_i \leq \tilde{\tau}_{i,j}\}$ where $\tilde{\tau}_{i,j-1} = \Phi^{-1}(F_i(\tau_{i,j-1}))$, $\tilde{\xi}_i = \Phi^{-1}(F_i(\xi_i))$ and $\tilde{\tau}_{i,j} = \Phi^{-1}(F_i(\tau_{i,j}))$. □

Department of Economics, BI Norwegian Business School, Stavanger, Norway 4014

E-mail address: njal.foldnes@bi.no

Department of Economics, BI Norwegian Business School, Oslo, Norway 0484

E-mail address: steffeng@gmail.com