

Factor analyzing ordinal items requires substantive knowledge of response marginals

Steffen Grønneberg and Njål Foldnes
BI Norwegian Business School

In the social sciences measurement scales often consist of ordinal items and are commonly analyzed using factor analysis. Either data are treated as continuous, or a discretization framework is imposed in order to take the ordinal scale properly into account. Correlational analysis is central in both approaches, and we review recent theory on correlations obtained from ordinal data. To ensure appropriate estimation, the item distributions prior to discretization should be (approximately) known, or the thresholds should be known to be equally spaced. We refer to such knowledge as substantive, because it may not be extracted from the data, but must be rooted in expert knowledge about the data-generating process. An illustrative case is presented where absence of substantive knowledge of the item distributions inevitably leads the analyst to conclude that a truly two-dimensional case is perfectly one-dimensional. Additional studies probe the extent to which violation of the standard assumption of underlying normality leads to bias in correlations and factor models. As a remedy, we propose an adjusted polychoric estimator for ordinal factor analysis that takes substantive knowledge into account. Also, we demonstrate how to use the adjusted estimator in sensitivity analysis when the continuous item distributions are known only approximately.

Keywords: Ordinal data, Factor retention, Factor Analysis, Polychoric correlation

Psychologists often work with theories that relate unobserved abilities or characteristics (e.g., self-esteem, anxiety, conscientiousness) to observed human behavior. Such constructs are themselves unobservable, but are each thought to reveal themselves indirectly through their effects on a scale, that is, a set of observable variables that all tap into the construct in question. In many empirical studies scale data are gathered through the use of questionnaires, and are ordinal in nature.

Factor analysis is a commonly applied technique in development and analysis of such scales. For instance, factor modeling plays a pivotal role in the important decision of how many factors to retain for a given scale. Factor analysis with continuous data can be performed without strong assumptions on the distribution of the data, e.g., normality is not required (Browne, 1984; Satorra, 1989). When the items that constitute a scale are ordinal in nature, e.g., of Likert-type, distributional assumptions become more important, and analysts should carefully consider how to proceed. One option, widely encountered in practice, is to simply proceed using continuous methodology, i.e., by fitting the factor model to regular Pearson correlations among the scale items. The most common estimator for this is normal-

theory based maximum-likelihood (ML), and we refer to this approach as continuous ML (cont-ML). Under very specific conditions reviewed later in this paper cont-ML may be a viable option, but it is generally inconsistent. Another popular option is to properly accommodate the ordinal nature of the data by adding a non-linear discretization model to the measurement model of the latent constructs. In the discretization model we posit that each ordinal variable is the result of dividing a continuous *latent response variable* into a finite number of values. The division is dictated by a set of cut-off values along the scale of the response variable. These cut-off values are referred to as *thresholds*. The unobserved latent response variable is continuous and is linearly related to latent factors in the same manner as a manifest variable in regular continuous factor analysis. At the expense of model complexity, the covariance structure model implied by the factor model is thereby accommodated. We refer to the correlations among the latent response variables as underlying correlations or response correlations. If we assume that two response variables are bivariate normally distributed the underlying correlation may be estimated (Olsson, 1979a), and we refer to this estimate as the *polychoric* correlation. Note that when the response variables are not bivariate normal, the polychoric correlation is likely biased with respect to the true underlying correlation. A common procedure is to fit the proposed factor or covariance model to the polychoric correlation matrix using least squares (LS) estimation. We refer

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

to this approach as categorical LS (cat-LS), and it is reported to outperform cont-ML in terms of estimation bias and fit inference (e.g., Flora & Curran, 2004; Foldnes & Grønneberg, 2021; Li, 2016). cat-LS 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.

The present study follows Foldnes and Grønneberg (2021) in adopting a copula perspective when investigating violations of the normality assumption and its consequences for the polychoric correlation estimator. As will be explained in a later section, any multivariate distribution may be separated into its marginals and its copula. Violations of normality can consequently occur in two ways: either some of the marginal distributions are non-normal, or the copula is non-normal. The Venn diagram in Figure 1 separates the class of multivariate distributions into four distinct regions, according to normality criteria. Distributions whose copula is not normal and whose marginals are not all normal are contained in region A. The class of multivariate normal distributions is characterised by normal marginal distributions coupled with normal copulas, represented by region C. Region B contains distributions whose copula is normal, but whose marginals are not all normal. Finally, region D consists of distributions whose marginals are all normal, but whose copula is not normal. In the present study the main focus is on region B, that is, on the consequences of marginal non-normality on cat-LS, while retaining a normal copula.

The effect of non-normality of the type represented by region D (normal marginals coupled by a non-normal copula) on ordinal factor analysis may be substantial, as shown by Foldnes and Grønneberg (2021). Fortunately such conditions are detectable using statistical tests (Foldnes & Grønneberg, 2019b; Jöreskog, 2005; Maydeu-Olivares, 2006). In the present work we provide evidence obtained from such tests that underlying normality in real-world empirical datasets can not be taken for granted. However, the non-normality detected in such tests pertain to the copula. As shown in the appendix, marginal non-normality in the underlying vector has no testable implications. We are therefore unable to test whether marginal normality holds in practical settings. Hence, we deem the topic of non-normal marginals of special interest, and make it the focus of the present article. Underlying normality is often justified by invoking forms of the central limit theorem. However, tests for underlying normality are not always passed, and it is therefore important to investigate how ordinal factor analysis is affected by marginal non-normality.

Several influential simulation studies have concluded that cat-LS is moderately robust to violation of the underlying normality assumption (Flora & Curran, 2004; Li, 2016; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Foldnes and

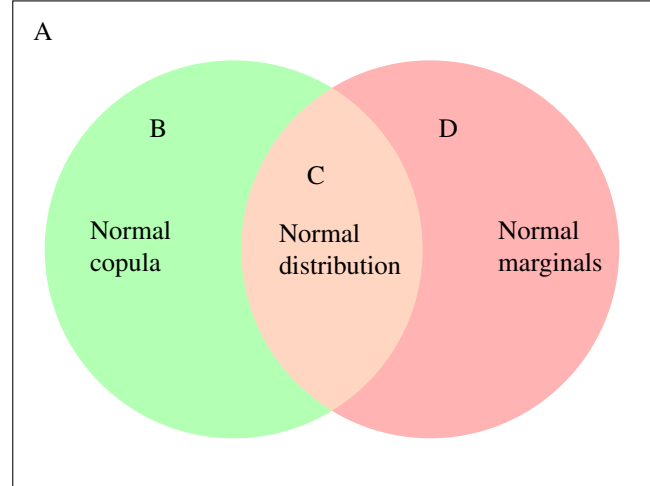


Figure 1. Venn diagram of multivariate distributions. Region A consists of all distributions whose copula is non-normal, and whose marginals are not all normal. Region B consists of all distributions whose copula is normal, but whose marginals are not all normal. Region C consists of all multivariate normal distributions. Region D consists of distributions whose copula is non-normal, and whose marginals are all normal.

Grønneberg (2021) identified problematic aspects of this conclusion, and presented conditions where cat-LS was sensitive to underlying non-normality. In the present paper we continue to study the extent to which cat-LS is sensitive to underlying normality, focusing on the consequences of marginal non-normality. There are data conditions where cat-LS inference is biased, conditions that unfortunately are not detectable by statistical tests. A theoretical analysis of the polychoric correlation given later in the paper shows that when the copula of the response vector is normal but its marginals are non-normal, the polychoric correlation will systematically overestimate size of the true correlation. This systematic overestimation may in turn lead to overoptimistic assessments of the strength of the factor structure in the data.

The misspecification involved in assuming normal response distributions can only be corrected by invoking expert substantive knowledge of the true distributional forms of the underlying response variables. In the present article we propose an approach to implement such a correction, which we refer to as cat-LS-adj. This adjustment requires the exact specification of the marginal distributions of the continuous response variables. In some situations, exact specification of the marginals may not be available. Then, cat-LS-adj is a tool for sensitivity analysis, when applied with many combinations of marginals that are compatible with approximate substantive knowledge.

This article makes several contributions to the literature. First, we explain how the copula perspective helps under-

stand different ways that multivariate normality may be violated, and how this may affect inference with ordinal data. Second, we demonstrate that copula normality may not be taken for granted in psychometric datasets. Third, using a newly developed class of non-normal distributions, we construct an informative worst-case scenario in the context of the number of factors problem, where an exact continuous two-factor solution is turned into an exact one-factor solution when analyzed with cat-LS. To solve this problem, a fourth contribution is a new adjusted cat-LS methodology based on a simple adjustment of the polychoric correlation estimator and its standard error. We demonstrate in a series of both Monte Carlo and population-level studies that cat-LS may be severely biased under marginal non-normality, while cat-LS-adj performs well. Fifth, we present a simple illustration of sensitivity analysis based on cat-LS-adj when only approximate knowledge of the marginals is available.

The present article is organized as follows. Next, we review the theoretical framework for cat-LS and present a evidence that underlying normality is not to be taken for granted in empirical datasets. We then present copula theory and illustrate its usefulness when analyzing the polychoric correlation under non-normality. We next present our illustrative worst-case factor retention scenario, where a two-factor solution is entirely lost, and replaced by a distinctive one-factor solution in the ordinal data. This is followed by a review of recent research in ordinal factor analysis. Subsequently, we propose an adjustment to the polychoric estimator in order to remedy the problem of marginal non-normality. The adjustment is thereafter evaluated in a series of five studies.¹ The implied need for substantive knowledge is then discussed, and some practical recommendations for ordinal factor analysis is provided.

While this paper focuses on ordinal factor analysis estimated via cat-LS, the paper has consequences for related models and related estimation methods as well. For example, our discussion is directly relevant also for the more general class of ordinal structural equation models.

Factor analysis for ordinal data

Factor analysis was historically first developed for variables that each consisted of a test or scale score (Bartholomew, 2007; Flora & Flake, 2017). Items were therefore considered approximately continuous and model estimation was based on continuous methodology, culminating in normal-theory based maximum likelihood confirmatory factor analysis (Jöreskog, 1969), with later extensions to non-normal continuous data (Browne, 1984; Satorra & Bentler, 1988).

However, ordinal data are widely encountered in social science, for instance in the context of sample surveys. In such situations it is often of interest to directly factor analyze the items, instead of forming subscores or parcels based

on more or less arbitrary collections of items. In practice, nothing stops the analyst from applying cont-ML to factor analyze ordinal items. However, it has long been known that Pearson correlations among ordinal variables tend to underestimate the underlying true response correlation (e.g., Coenders, Satorra, & Saris, 1997). The problems (biased factor loadings, standard errors and chi-square statistics) associated with factor-analyzing ordinal data with cont-ML are reported in many simulation studies (e.g., Beauducel & Herzberg, 2006; Foldnes & Grønneberg, 2021; Li, 2016; Rhemtulla et al., 2012). As an improvement of cont-ML, Foldnes and Grønneberg (2021) proposed a simple transformation of the ordinal sample data before calculating regular correlations. This new method, referred to as cont-ML-adj, was found consistently, but marginally, to outperform cont-ML.

cat-LS methodology for factor analyzing ordinal variables of dichotomous (Christoffersson, 1975) and polytomous (Muthén, 1984) nature is based on the concept of an underlying latent response variable, which, when discretized according to threshold values, produces the observed ordinal variable. Let us refer to the continuous response variables as X_i^* , and to the observed ordinal variables as X_i , for $i = 1, 2, \dots, p$. For each variable X_i^* there are thresholds denoted by $-\infty = \tau_{i,0} < \tau_{i,1} < \dots < \tau_{i,K_i-1} < \tau_{i,K_i} = \infty$. The discretization rule is as follows:

$$X_i = x_{i,j} \leftrightarrow \tau_{i,j-1} \leq X_i^* < \tau_{i,j}, \quad (1)$$

where $x_{i,1} < x_{i,2} < \dots < x_{i,K_i}$ constitute the discrete set of values realized by the variable X_i . Hence X_i is a K_i -category ordinal variable. A widespread practice² is to assign consecutive integers $x_{i,j}^* = j$ for $j = 1, 2, \dots, K_i$. cat-LS proceeds by fitting the factor model to the correlation matrix of the response vector $X^* = (X_1^*, X_2^*, \dots, X_p^*)'$, using unweighted or diagonally weighted least squares estimation. In the online appendix (p. 1) the necessity of the discretization model is further discussed.

It is crucial to observe that the underlying correlation matrix is not identified without making strong distributional assumptions concerning X^* (Grønneberg & Moss, 2021; Grønneberg, Moss, & Foldnes, 2020). Traditionally, it has been assumed that the response variables are bivariate normally distributed, which yields the *polychoric* correlation, estimated by ML estimation (Olsson, 1979a). Many studies have confirmed that when X^* is multivariate normal, cat-LS performs well and better than continuous methodology. When X^* departs from normality, the polychoric correlation and its standard error estimate may be biased, rendering factor analysis inference invalid (Foldnes & Grønneberg, 2019b, 2021). These findings challenge the view that cat-LS is moderately

¹The online supplementary material contains R computer code for the illustrative cases, numerical results and simulations.

²A risk with such integer encoding is that it appears to produce a scale on the interval level of measurement.

robust to violation of the underlying normality assumption (Flora & Curran, 2004; Li, 2016; Rhemtulla et al., 2012).

Underlying normality is not always empirically tenable

Using a selection of real-world datasets, we next illustrate that the testable implications of response normality are not always empirically tenable. Seven datasets with psychometric items were collected from the psychometrics textbook Mair (2018). The datasets were selected among all datasets provided with the book using the following criteria: The dataset must contain ordinal item-level data from published psychometric studies with at least three categories, and have a sample-size of at least 200. The sample-size requirement comes from the evaluations of the bootstrap test given in Foldnes and Grønneberg (2019b), which showed that the test still performed adequately at this sample-size. The requirement of having at least three categories is that in the binary case, it is impossible to test underlying normality with bivariate data (Muthén & Hofacker, 1988), and in our illustration we will use a test statistic that is based on bivariate distributions and cannot therefore be used with binary data. All background variables, such as gender, were discarded, and we consider the unconditional distribution of all remaining items. This means that it could be that some subgroups of the datasets behave differently when it comes to tests of underlying normality than when considering the full datasets.

All datasets were subjected to the bootstrap test of Foldnes and Grønneberg (2019b) as implemented in the R package *discnorm* (Foldnes & Grønneberg, 2020) to assess joint and bivariate response normality. This test is a refinement of the testing procedures of Maydeu-Olivares (2006). The conclusion the tests are presented in Table 1. Two types of tests of response normality are reported: A global test for underlying normality for all item variables in the dataset, and individual tests for underlying normality among all pairs – i.e., a sequence of bivariate tests. The bivariate tests are performed since polychoric correlations among a pair of variables will be consistent if the underlying continuous variables are normal. The pairwise tests use the same test statistic as the joint test but applied to bivariate data.

In all of the seven empirical datasets we found strong support for the alternative hypothesis, namely that the underlying distribution has a non-normal copula. All joint tests of underlying normality have p-values numerically equal to zero. The sample-size reported in Table 1 is before any removal of missing values. For all tests, listwise deletion of missing variables for the variables used in the test was performed. For the global test, this means full listwise deletion of all cases with any missing values. For the pairwise tests, listwise deletion was applied only to the bivariate pair considered by the test. The sample size of all datasets are still above 200 also after the removal of cases with missing values. The item datasets we consider are analyzed using

different types of psychometric techniques, including network modeling and IRT modeling, where the precise modeling assumptions may be different from the assumption of underlying normality. Nevertheless, the findings show that underlying non-normality in psychometric scales may not be taken for granted. This may have important implications for statistical estimation and inference with cat-LS, as shown in Foldnes and Grønneberg (2019b); Foldnes and Grønneberg (2021) and the studies in the present article.

To investigate the sensitivity of the polychoric correlation to departures from normality, it is fruitful to adopt the perspective of copula theory, to which we turn next.

Copula theory and the polychoric correlation

The present section gives a conceptual introduction to copula theory and its usefulness for studying the polychoric correlation. We confine ourselves to the bivariate case, although the theory extends naturally to higher dimensions. The interested reader is referred to the textbook Nelsen (2007) for a formal treatment. A major application of copulas is to construct multivariate distributions and models. In the psychometric literature some applications of copulas are found in, e.g., Braeken, Tuerlinckx, and Boeck (2007); Foldnes and Grønneberg (2021); Krupskii and Joe (2013); Nikoloulopoulos and Joe (2015), and copulas have recently been used to handle the endogeneity problem in regression (Falkenström, Park, & McIntosh, 2021).

Copulas are used to describe the dependency structure between variables, when taking the marginal distributions out of the equation. Formally, a copula is a distribution with univariate margins that are uniform on the interval $[0, 1]$. Let us denote its cumulative distribution function (CDF) by $C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$. It may be used to construct statistical distributions as follows. Given univariate CDFs $G(x)$ and $H(y)$, a valid bivariate CDF $F(x, y) = P(X \leq x, Y \leq y)$ is obtained by combining these with the copula:

$$F(x, y) = C(G(x), H(y)). \quad (2)$$

It can be shown that the marginal distributions of $F(x, y)$ are given by $G(x)$ and $H(y)$.

Also, any distribution may be uncoupled into its copula and its marginal CDFs. Sklar's theorem (Sklar, 1959) states that for any CDF $F(x, y)$ with continuous marginal CDFs F_X and F_Y , there exists a unique copula C_F such that $F(x, y) = C_F(F_X(x), F_Y(y))$.

Consider the well-known class of bivariate normal distributions, where we employ the following notation. For the standard normal distribution, denoted by $N(0, 1)$, the CDF is denoted by $\Phi(z)$. For bivariate normal distributions with standard normal marginals the CDF is denoted by $\Phi_{\rho}^{12}(z_1, z_2)$, where ρ_0 is the correlation. For instance, with correlation $\rho = .5$ the density contours of the resulting distribution $\Phi_{\rho=.5}^{12}$ are depicted in Figure 2a, together with

	1	2	3	4	5	6	7
Dataset	ASTI	BSSS	CEAQ	condom	Rogers	SDOwave	Wenchuan
Global p-value	0	0	0	0	0	0	0
Sample size	1129	1626	208	500	408	612	362
Number of variables	25	8	16	6	26	20	17
Number of categories	3 or 4	5	3	4	4 or 5	7	5 or 6
Proportion of p-values = 0	0.16	0.68	0	0.73	0.06	0.46	0.24
Proportion of p-values < 5%	0.58	0.96	0.17	0.93	0.25	0.88	0.72
Mean of p-values	0.1	0.01	0.33	0.01	0.28	0.03	0.06

Table 1

Summaries of tests for underlying copula normality for all datasets from the R package Mair (2020) on psychometric items with more than two categories. The sample size is before any removal of missing values.

representations of the marginal $N(0, 1)$ -densities. This distribution is of Region C in Figure 1. The associated copula is a member of the class of normal copulas, and is given by

$$C_{\rho=.5}^N(u_1, u_2) = \Phi_{\rho=.5}^{12}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)).$$

There are many classes of copulas other than the normal class, and within each class there is typically a parameter that controls the strength of dependence. The dependency parameter in the specific normal copula described above is $\rho = .5$. This normal copula may be used to construct bivariate distributions whose copula will be normal, but whose marginal distributions may be freely specified. That is, given any pair of marginal CDFs $G(x)$ and $H(y)$, we may apply eq. (2) to obtain a valid bivariate CDF

$$K(x, y) = C_{\rho=.5}^N(G(x), F(y)),$$

that defines a distribution whose copula $C_{\rho=.5}^N$ is normal, and whose marginal distributions will be $G(x)$ and $H(y)$. We illustrate this by moderately distorting the normal marginals in Figure 2a as follows. We let $F(x)$ be the CDF associated with a gamma distribution with shape parameter 2, and $G(y)$ be the CDF associated with a Weibull distribution with shape parameter 20. The gamma and Weibull distributions are two-parameter families, and in each distribution the second parameter was chosen to yield unit variance. Also, both $F(x)$ and $G(y)$ are mean centered, so that $X \sim F$ and $Y \sim G$ have zero expectation and unit variance. The densities associated with F, G and $N(0, 1)$ are depicted in Figure 3. These distributional forms are sometimes found in empirical datasets of composite scores in psychometric studies, and belong to Region B in Figure 1. Skewness and excess kurtosis for X are 1.4 and 3, respectively, and for Y , -0.9 and 1.3 , respectively. To characterize moderate non-normality, typical skewness and kurtosis values used in simulation studies (Curran, West, & Finch, 1996, p.28) are 2 and 7, respectively. Accordingly, F and G depart only mildly from normality. The density contours of $K(x, y)$ are given in Figure 2b, together with representations of the distorted marginals. The correlation is distorted by replacing $N(0, 1)$ marginals with X and Y , from $\rho = .5$ in Figure 2a to $\rho = .45$ in Figure 2b.

Another way to generate non-normality is to distort the normal copula, while retaining the normal marginals. As an example, let us replace $C_{\rho=.5}^N$ associated with the distributions in Figures 2a and 2b with a copula belonging to the Clayton class of copulas. This class, in contrast to the normal copulas, allows lower tail dependency. In Figure 2c we have coupled the standard normal marginals in Figure 2a with a Clayton copula $C_{\theta=1.1}^{Cl}$ with dependence parameter $\theta = 1.1$. The dependency in the lower tails is visible. The correlation is distorted by replacing the normal copula by a Clayton copula, from $\rho = .5$ in Figure 2a, to $\rho = .525$ in Figure 2c.

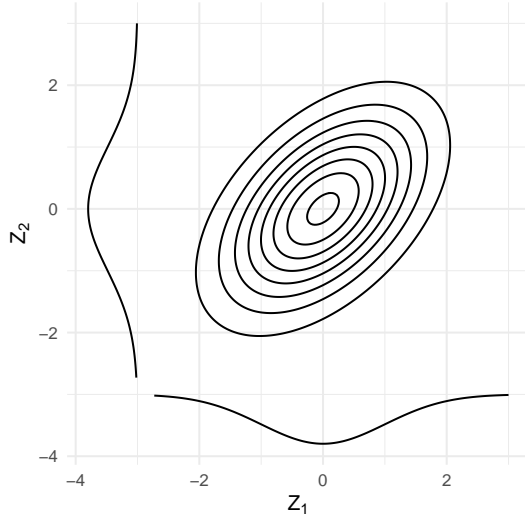
While the non-normality in the distribution in 2b was produced by distorting the normal marginals, and retaining the normal dependence structure, the non-normality in the distribution in Figure 2c is produced by distorting the normal copula, and retaining the normal marginals, and is therefore in Region D in Figure 1.

Finally, we distort both the marginals and the copula of the normal distribution in Figure 2a, and reach a distribution in Region A in Figure 1. The contours of the resulting distribution are depicted in Figure 2d. This distribution has non-normal marginals coupled together by the Clayton copula $C_{\theta=1.1}^{Cl}$. The correlation is severely distorted, from $\rho = .5$ in Figure 2a, to $\rho = .440$ in Figure 2d.

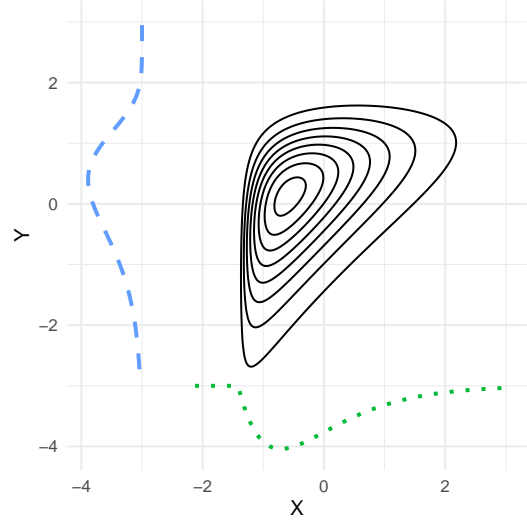
The four panels in Figure 2 illustrate how the choice of copula and marginals jointly affect the bivariate distribution. When investigating the sensitivity of a statistical procedure to distributional assumptions, the copula perspective allows us to disentangle violations pertaining to the marginal distributions from violations pertaining to the dependence structure. The uncoupling of marginal distributions from the dependence structure is well suited to study the sensitivity of the polychoric correlation to violations of the bivariate normality assumption, as next illustrated.

Investigating polychorics from a copula perspective

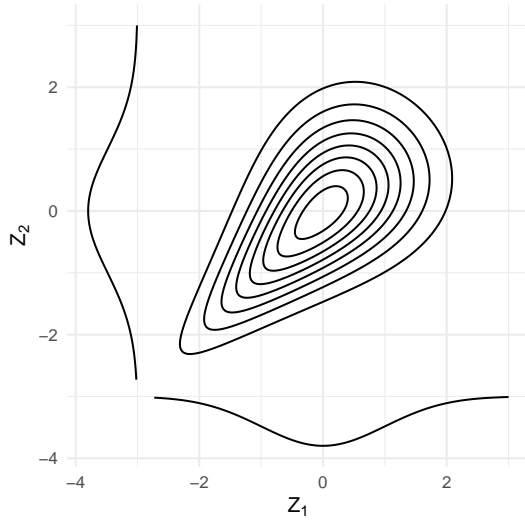
Consider the case illustrated in Figures 2a and 2b, i.e., the case where the underlying variables are coupled together by a normal copula $C_{\rho=\rho_0}^N$, and where the marginals are either normal or non-normal. In this condition it is straightforward to analyze the polychoric correlation: At the popula-



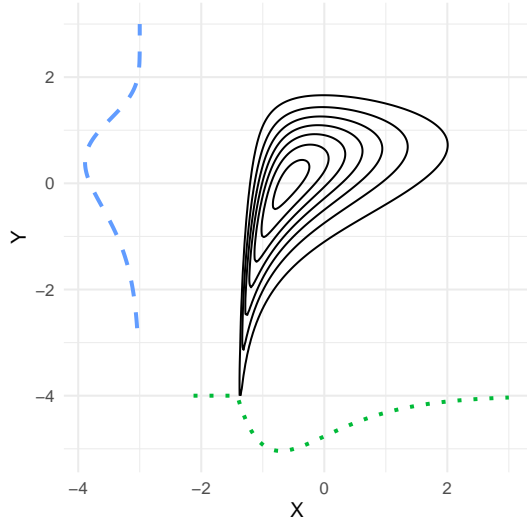
(a) Normal copula, normal marginals.



(b) Normal copula, non-normal marginals.



(c) Non-normal copula; normal marginals.



(d) Non-normal copula; non-normal marginals.

Figure 2. Four bivariate distributions. The normal copula in panels (a) and (b) has parameter $\theta = 0.5$. The non-normal copula in panels (c) and (d) is a Clayton copula with parameter $\theta = 1.1$. The marginals in panels (a) and (c) are standard normal. The marginals in panels (b) and (d) are standardized Gamma (X) and Weibull (Y) distributions.

tion level it will estimate ρ_0 , the parameter of the normal copula. This observation is general and holds regardless of the number and placement of thresholds. To show this, consider first the discretization of a univariate variable according to Equation (1). The resulting ordinal variable could equally well have been produced by discretizing any other continuous variable by transforming the thresholds using the quantile function of the continuous variable. For instance, discretization of the gamma distribution in Figure 3 with thresh-

olds $\tau_1 = -1, \tau_2 = 0, \tau_3 = 1$ is equivalent to discretization of the standard normally distributed Z_1 with thresholds $\tilde{\tau}_1 = -1.19, \tilde{\tau}_2 = 0.24$, and $\tilde{\tau}_3 = 1.06$, see Figure 4. This illustrates that the thresholds and the underlying continuous distribution are not jointly identifiable from data: Either must the thresholds be known for the distribution to be identified, or vice versa.

This discretization equivalence holds also when we move to the bivariate case, provided the two continuous distributions

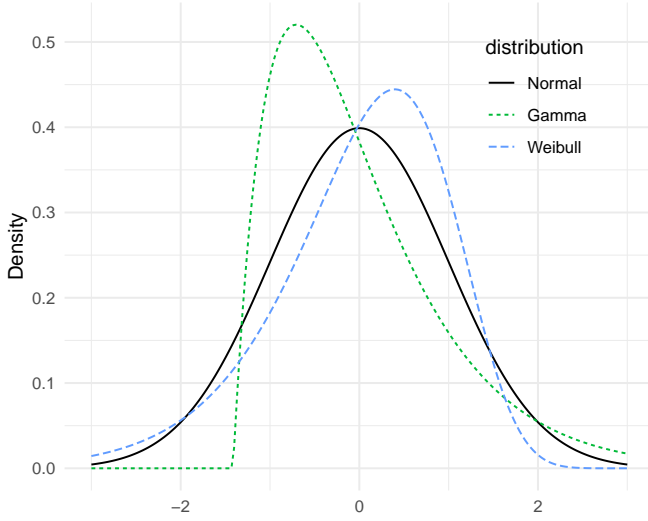


Figure 3. The normal, a Gamma, and a Weibull distribution. All three have zero mean and unit variance.

have the same copula. In our illustration the common copula in Figures 2a and 2b is $C_{\rho=.5}^N$. Now let us calculate the probability that both ordinal variables take on the lowest possible values. This probability is

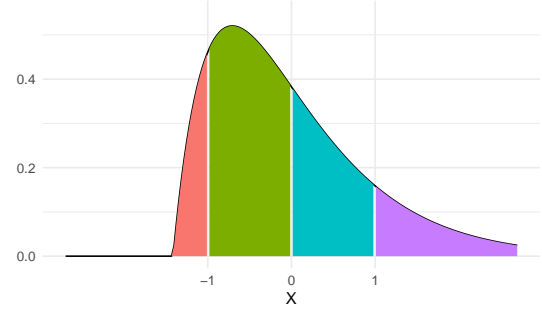
$$P(X \leq \tau_{1,1}, Y \leq \tau_{2,1}) = C_{\rho=.5}^N(G(\tau_{1,1}), H(\tau_{2,1})).$$

Transforming the thresholds into $\tilde{\tau}_{1,1} = \Phi^{-1}(G(\tau_{1,1}))$ and $\tilde{\tau}_{2,1} = \Phi^{-1}(H(\tau_{2,1}))$, yields

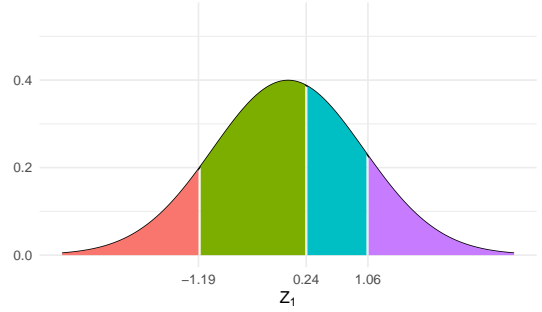
$$\begin{aligned} P(Z_1 \leq \tilde{\tau}_{1,1}, Z_2 \leq \tilde{\tau}_{2,1}) \\ &= C_{\rho=.5}^N(\Phi(\Phi^{-1}(G(\tau_{1,1}))), \Phi(\Phi^{-1}(H(\tau_{2,1})))) \\ &= C_{\rho=.5}^N(G(\tau_{1,1}), H(\tau_{2,1})). \end{aligned}$$

The calculation implies that for any set of thresholds, the discretization of the distribution in Figure 2b leads to the exact same ordinal distribution that is produced by the discretization of the distribution in Figure 2a, after transforming the thresholds. Now, since the polychoric correlation *assumes underlying normality*, it will interpret the ordinal distribution as stemming from the discretization of the bivariate normal distribution in Figure 2a. This means that the polychoric correlation will reach $\rho = .5$ at the population level, regardless of the number and values of the thresholds. Since the correlation in Figure 2b was equal to $\rho = .45$, the polychoric correlation is biased, with a relative bias of $100(.5 - .45)/.45 = 11\%$.

Therefore, with an underlying normal copula $C_{\rho=\rho_0}^N$, the situation is simple: the population value of the polychoric correlation is ρ_0 , regardless of threshold placement. When the underlying copula is non-normal, as in Figures 2c and 2d, the situation is more complex, and the threshold placements come into play. To illustrate, we consider the ordinal distributions obtained from discretizing all four distributions in Figure 2, with three different threshold placements



(a) Discretizing a gamma distribution with thresholds $-1, 0$, and 1 .



(b) Discretizing $N(0, 1)$ with thresholds $-1.19, 0.24$, and 1.06 .

Figure 4. Discretizations of X and of $Z_1 \sim N(0, 1)$ that produce the same ordinal distribution.

for $K = 7$. The three configurations yield observed ordinal distributions with ceiling effects, with floor effects and with symmetry. For each of the distributions in Figure 2 a very large sample is simulated and discretized according to the three threshold configurations³. Then, the population values of the polychoric correlation coefficient $\hat{\rho}_{NT}$ is calculated, with the following results

Figure	True ρ	$\hat{\rho}_{NT}$		
		Ceiling	Floor	Symmetrical
2a	.5	.5	.5	.5
2b	.45	.5	.5	.5
2c	.525	.611	.485	.526
2d	.440	.536	.499	.520

We see that with a normal copula the polychoric correlation is not dependent upon the thresholds, and will always reach the underlying copula parameter $\rho = .5$. In the non-normal copula case, we see that the placement of thresholds affects the polychoric correlation, and that its value may be markedly different from the true underlying correlation. The role of threshold placement on ordinal SEM was thoroughly investigated by Foldnes and Grønneberg (2021).

³See the R code in the supplementary material for the threshold values.

An illustrative case where non-normal response marginals impairs factor retention

In this section a case is presented where severe non-normality in the response marginals leads to a faulty conclusion in the number of factors problem. The purpose of the case is to demonstrate that substantive knowledge is important when factor analyzing ordinal items. Also, this example is revisited in a later Monte Carlo study which evaluates the adjusted cat-LS method.

The illustrative case is a two-dimensional scale with continuous items that appears exactly uni-dimensional after being discretized into ordinal items and estimated using established methodology. The fallacy is caused by lacking substantive knowledge about the univariate response distributions. That is, applying cat-LS and its default assumption of underlying bivariate normality, the two-dimensional scale will appear to fit perfectly to the uni-dimensional parallel factor model. Our construction holds at the population level for any discretization according to Equation (1), i.e., regardless of the number of categories and distributional shape of each ordinal variable. Also, the idea behind the construction of this case is general, and allows expansion to any factor number. For simplicity we limit ourselves to six indicators, but our construction can be extended to any number of indicators per factor.

A two-factor scale with continuous items

Consider a researcher that has gathered information on six variables on an ordinal-categorical scale in order to analyze the data using ordinal factor analysis. We eliminate sampling variability from our treatment by assuming that the researcher may draw arbitrarily large random samples from the population. In other words, the population distribution of the six-dimensional ordinal vector X is fully known. This vector is a discretized version of an underlying continuous response vector X^* , as in Equation (1). The distribution of X^* is perfectly in accordance with the two-factor model depicted in Figure 5. The figure also contains population model parameter values. The indicator variables are all standardized, that is, with zero mean and unit variance. Also, for identifiability, the two latent variables are standardized. The parameter values in Figure 5 imply the correlation matrix given in Table 2. It is clear that this matrix has poor fit to a uni-dimensional model. For instance, consider the population value of the root mean square error of approximation (RMSEA), defined as $\sqrt{F_0/d}$, where F_0 is the minimum of the normal-theory ML fit function, and d denotes the model degrees of freedom. In our case, the population RMSEA value when fitting a one-factor model to the correlation matrix in Table 2 is 0.23, suggesting very poor fit according to conventional cut-off values (e.g. Browne & Cudeck, 1992). That is, the six continuous variables are far from forming a uni-dimensional scale, but

fit perfectly to the model with two subscales depicted in Figure 5.

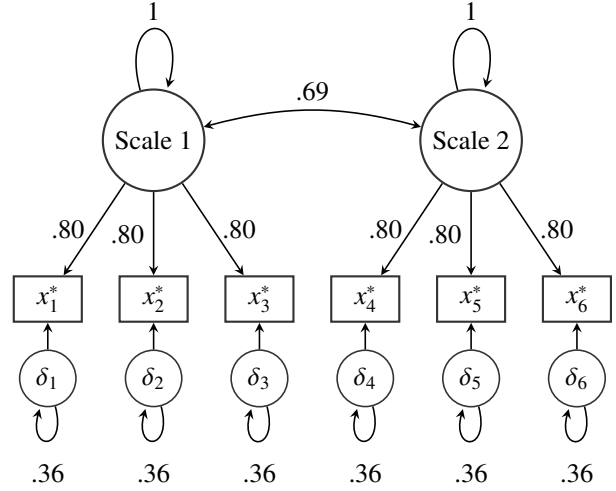


Figure 5. Two correlated constructs, with unit variance indicators.

Table 2

Covariance matrix for six continuous variables stemming from a two-factor model.

	X_1^*	X_2^*	X_3^*	X_4^*	X_5^*	X_6^*
X_1^*	1					
X_2^*	.64	1				
X_3^*	.64	.64	1			
X_4^*	.45	.45	.45	1		
X_5^*	.45	.45	.45	.64	1	
X_6^*	.45	.45	.45	.64	.64	1

Having specified the correlation matrix of the underlying continuous vector $X^* = (X_1^*, X_2^*, \dots, X_6^*)'$, our next aim is to give a stochastic representation of X^* , thereby fully specifying its distribution. Care must be taken, so that X , the observed ordinal vector after discretization of X^* , has a poly-choric correlation matrix that is perfectly uni-dimensional. One way to accomplish this is based on a novel approach for constructing multivariate random vectors given in Foldnes and Grønneberg (2021), where piecewise linear transformations of standard normal variables are used. This new approach may be used to construct non-normal multivariate distributions with pre-specified univariate skewness and kurtosis, as well as a pre-specified correlation matrix, similar to earlier methodologies (Foldnes & Olsson, 2016; Vale & Maurelli, 1983).

For our illustration, we will use the two piecewise linear

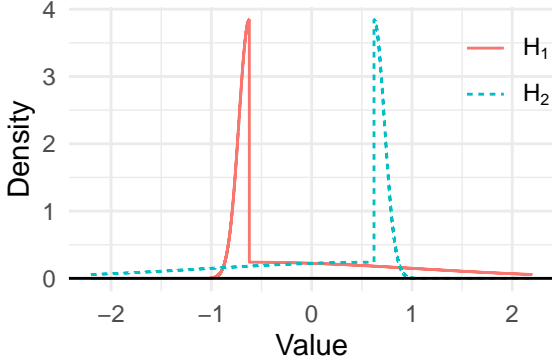


Figure 6. Densities of $H_1(Z)$ and $H_2(Z)$, where $Z \sim N(0, 1)$.

functions

$$H_1(Z) = \begin{cases} 0.1039 \cdot Z - 0.6216 & \text{if } Z < 0 \\ 1.6619 \cdot Z - 0.6216 & \text{if } Z \geq 0 \end{cases}, \quad (3)$$

and $H_2(Z) = -H_1(-Z)$. The values of the intercepts and slopes were chosen so that $H_1(Z)$ and $H_2(Z)$ both have zero mean and unit variance (Foldnes & Grønneberg, 2021), given that Z is a standard normal variable. However, $H_1(Z)$ and $H_2(Z)$ are not normal. For instance $H_1(Z)$ and $H_2(Z)$ have a skewness of 1.55 and -1.55 , respectively. The excess kurtosis of both $H_1(Z)$ and $H_2(Z)$ is 2.1. These skewness and kurtosis values are not considered extreme, and are deemed indicative of low to moderate non-normality (see, e.g., Kline, 2015, p. 76). Figure 6 depicts the densities of $H_1(Z)$ and $H_2(Z)$. Even though traditionally used skewness and kurtosis criteria indicate that these densities are only moderately non-normal, it is seen that the densities depart markedly from a normal density. We may speculate that such densities may arise in practice by mixing two subpopulations, one in which values are almost uniformly distributed, and one in which values are highly concentrated.

Next, piecewise linear transformations are applied to each coordinate of a multivariate normal vector $Z = (Z_1, Z_2, \dots, Z_6)'$. We apply H_1 to the first three coordinates, and H_2 to the last three coordinates:

$$X^* = (H_1(Z_1), H_1(Z_2), H_1(Z_3), H_2(Z_4), H_2(Z_5), H_2(Z_6))' \quad (4)$$

The covariance matrix of X^* is a function of the correlations in Z . Importantly, we choose Z to conform to a parallel measurement model by setting all the correlations among Z coordinates to $\rho = 0.7$. Also, if Z_1 and Z_2 are bivariate normal with correlation .7, then $H_1(Z_1)$ and $H_1(Z_2)$ will have correlation .64. Moreover, $H_1(Z_1)$ and $H_2(Z_2)$ will have correlation .45. This means that the covariance matrix of X^* equals the correlation matrix in Table 2.

So far, we have constructed a continuous non-normal vector X^* for which a two-factor model fits perfectly, by trans-

forming coordinate-wise a normal vector Z for which a one-factor model fits perfectly. We next describe how the discretization of X^* removes the true two-factor structure and replaces it with the one-factor structure of Z , when seen through the lens of polychoric correlations. Mathematically, we have

$$\begin{aligned} \tau_1 \leq X_1^* < \tau_2 &\Leftrightarrow H_1^{-1}(\tau_1) \leq H_1^{-1}(X_1^*) < H_1^{-1}(\tau_2) \\ &\Leftrightarrow \tilde{\tau}_1 \leq H_1^{-1}(H_1(Z_1)) < \tilde{\tau}_2 \\ &\Leftrightarrow \tilde{\tau}_1 \leq Z_1 < \tilde{\tau}_2. \end{aligned} \quad (5)$$

Hence, discretizing X_1^* with thresholds $\tau_1 < \tau_2 < \dots < \tau_{K-1}$ is statistically indistinguishable from discretizing Z_1 with transformed thresholds $H_1^{-1}(\tau_1) < H_1^{-1}(\tau_2) < \dots < H_1^{-1}(\tau_{K-1})$. See Figure 4 for the same calculus when $H_1(Z)$ is replaced by a gamma distribution.

This leads to the following remarkable conclusion, when applying the argument to all six variables in Equation (4): Discretizing the original X^* using any set of thresholds produces the same ordinal vector as discretizing a multivariate normal vector Z whose pairwise correlations are all equal to .7, after transforming the thresholds by H_1^{-1} or H_2^{-1} . Now consider what happens when we use the default polychoric correlation estimator based on the population distribution X . As explained above, X could have been produced by discretizing a multivariate normal vector with pairwise correlations equal to .7. The polychoric correlation estimator by default assumes underlying normality. With underlying normality, polychoric correlations are consistent. Since the exactly normal Z can generate X , the assumption of underlying normality is fulfilled, and the polychoric correlation estimator will therefore reach the correlation .7, instead of the true values .64 or .45⁴. That is, although the latent correlation between, say, X_1^* and X_4^* is .45, because discretizing the bivariate distribution of (X_1^*, X_4^*) is equivalent to discretizing the bivariate normal vector (Z_1, Z_4) , the polychoric correlation will reach the correlation .7 between Z_1 and Z_4 . So for (X_1^*, X_4^*) , the polychoric correlation is biased at the population level, with a relative bias close to 30%.

In conclusion, when we discretize a continuous non-normal X^* that stems from a two-factor model, the polychoric correlation estimator used in cat-LS will equal

$$\Sigma_1 = \begin{pmatrix} 1 & & & & & \\ .7 & 1 & & & & \\ .7 & .7 & 1 & & & \\ .7 & .7 & .7 & 1 & & \\ .7 & .7 & .7 & .7 & 1 & \\ .7 & .7 & .7 & .7 & .7 & 1 \end{pmatrix}.$$

Fitting a one-factor model to this matrix will yield perfect fit. Furthermore, this factor model is parallel, meaning that all

⁴This situation is similar to the discussion of Figures 2a and 2b on p. 5, where the underlying copula is normal.

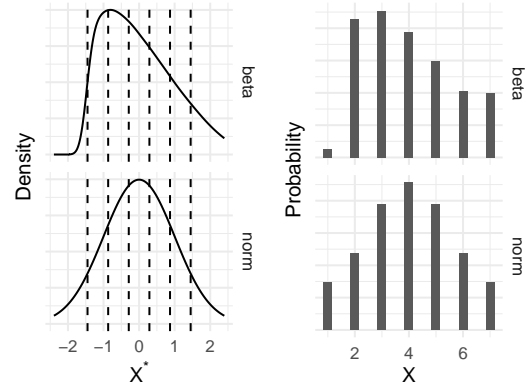
loadings are equal to $\sqrt{7}$. We may conclude that a researcher working with ordinal data obtained through an arbitrary discretization of the two-factor continuous vector X^* will at the population level find perfect support for a parallel one-factor measurement model.

This counter-example shows that the specification of correct marginals is required for ordinal factor analysis. As explained in more detail in the appendix, a troubling feature of the example is that any test for underlying non-normality will fail to detect that X^* is non-normal. The reason for this is that the non-normality is concentrated in the marginals, and the distribution of the ordinal observations do not place any restrictions on the marginals of the continuous response variable.

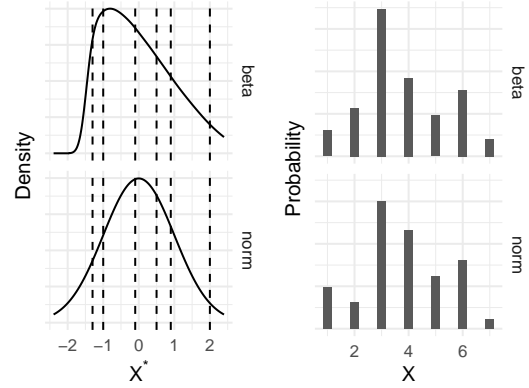
A review of recent results for ordinal factor models

Recent research (Foldnes & Grønneberg, 2019b; Monroe, 2018) demonstrated that the polychoric correlation and its standard error may be substantially biased under violations of bivariate normality. Foldnes and Grønneberg (2021) extended this line of research to structural equation models, by simulating 20-dimensional response vectors with normal marginal distributions. The multivariate dependence structure among the response variables, that is, the copula, was however not normal. Although cat-LS outperformed cont-ML, it was in some conditions substantially biased in terms of parameter and standard error estimation, and model fit assessment. Also, it was found in Foldnes and Grønneberg (2021) that the particular details of the discretization in Equation (1) influenced the bias, even within the same number K of categories, and the same type of observed ordinal distributions (ceiling, floor or symmetrical). While the performance of cont-ML and cat-LS was poor in many conditions defined by non-normal copulas, such conditions may reliably be detected by statistical means using, e.g., the bootstrap test⁵ proposed by Foldnes and Grønneberg (2019b). This test maintained Type I error control and exhibited acceptable power at sample size 300 for a medium-sized SEM (Foldnes & Grønneberg, 2021).

In the present study we investigate a less tractable non-normal condition than described above. Instead of isolating non-normality to the copula, we inject the latent response marginal distributions with non-normality, depicted as region B in Figure 1. That is, we consider the case where the univariate marginal distributions of the response variables X_i^* , $i = 1, 2, \dots, p$, are non-normal, and mainly coupled together by a normal copula. As explained by Foldnes and Grønneberg (2021) and further discussed in the appendix (p. 22), there is no way to formally detect whether ordinal data stem from discretizing a response vector with normal or non-normal marginal distributions, due to the confounding of threshold placement and response distribution illustrated in Figure 4.



(a) Discretizing a gamma distribution with thresholds $-1, 0$, and 1 .



(b) Discretizing $N(0, 1)$ with thresholds $-1.19, 0.24$, and 1.06 .

Figure 7. Two latent response distributions and their discretization with (a) Equally spaced thresholds (b) Unequally spaced thresholds.

Let us illustrate the interplay between the response marginal distribution and the threshold values τ in generating ordinal variables. We consider two underlying response marginal distributions, a skewed beta distribution and the standard normal. The beta distribution is scaled to have zero mean and unit variance. Also, we consider one set of thresholds that are equally spaced $(-1.45, -0.87, -0.29, 0.29, 0.87, 1.45)$, and one set of unequally spaced thresholds $(-1.3, -1, -0.1, 0.5, 0.9, 2)$. In Figure 7a, we observe that although discretization results in substantial loss of information with respect to the original response distribution, main aspects of the original distributional form are preserved under equally spaced thresholds. In Figure 7b we apply unequally spaced thresholds, and the original distributional form is almost completely lost after discretization.

⁵Available in the R package *discnorm* (Foldnes & Grønneberg, 2020).

The above example illustrates that, as K increases, and for equally spaced thresholds, the ordinal distribution increasingly approximates the underlying continuous response distribution. This is the intuition behind the recently proved theoretical result (Cor. 1, Foldnes & Grønneberg, 2021) that the correlation among integer-coded ordinal variables converges (as K increases) towards the true underlying correlation of response distributions, regardless of their distributional forms, provided the thresholds are equally spaced. Intuitively, integer coding works well for equally spaced thresholds, since it roughly preserves the intervals among values on the response X^* . For an illustration, let us consider four ($K = 5$ categories) and nine ($K = 10$ categories) equally spaced thresholds imposed on a bivariate normal response vector whose correlation is .7. Calculating the correlation directly on ordinal data gives, at the population level, .626 and .68 for $K = 5$ and $K = 10$, respectively.

Unfortunately, there is no way to test statistically whether thresholds are equally spaced. Claims of even threshold spacing can only be supported by resorting to substantive knowledge. Thresholds being equally spaced is roughly equivalent to claiming that the integer-coded observed ordinal scale is measured on an interval scale. If the analyst can substantively argue that the intervals between any two adjacent points on the ordinal scale correspond to the same interval length on the latent response scale, then this is equivalent to stating that the thresholds are equally spaced. How a substantive researcher might justify such a claim seems not a trivial matter, and we deem this important question outside the scope of the present article.

A second type of recent theoretical development concerns estimation where knowledge of the latent response marginal distributions is taken into account. Foldnes and Grønneberg (2021) introduced two new estimators which are consistent as K , the number of categories, increases indefinitely: cont-ML-adj, which encode ordinal data in such a way that justify using the standard Pearson correlation directly on the data, and an adjustment to cat-LS which estimate thresholds using the provided latent response marginal distributions. These methods are summarized in the online appendix. Both are consistent only as the number of categories increases indefinitely, as long as the latent response marginal distributions are correctly specified. This means that at least for large K , having substantive knowledge of the copula of the response variables is not needed, but knowing the response marginal distributions is still required.

The estimator methodology suggested in the present paper features a simple adjustment procedure for the polychoric correlation. When the response copula is normal, this estimator is consistent for all $K \geq 2$.

What happens if we just treat data from the illustrative two-factor model as continuous?

Let us return to the two factor model in Figure 5 (p. 8) with severely non-normal marginals. Although not the recommended approach (Foldnes & Grønneberg, 2021), cont-ML is still in wide use in empirical research (e.g., Marsh et al., 2013; Nilsen et al., 2020). This method estimates Pearson correlations directly on integer-encoded data. The practice⁶ is frequently encountered when variables have at least five categories and are approximately symmetrical. In the simulation design employed by Rhemtulla et al. (2012), cont-ML and cat-LS performed comparably well in such conditions when having at least 6-7 categories. We may therefore inquire whether the two-factor structure of X^* would be detectable by cont-ML. According to the recommendation of Rhemtulla et al. (2012), cont-ML may be used with symmetrically distributed items with five or more categories. Let us therefore assume that we have seven symmetrically distributed levels in all six categories. Unfortunately, cont-ML then mistakes the number of factors to be one instead of two. The reason is that any pair of observed variables, say X_1 and X_2 , may be thought of as discretizations of the identically distributed Z_1 and Z_2 , using the same set of thresholds. For $K = 7$ the Pearson correlation between X_1 and X_2 will equal .662. Therefore, the Pearson correlation matrix of X is structurally equal to Σ_1 , and with .7 replaced by .662. Such a correlation matrix fits perfectly to a uni-dimensional parallel factor model.

Paradoxically, only in a specific situation with asymmetrical ordinal observed distributions will cont-ML be approximately unbiased. That is, if there are many thresholds equally spaced along the scale of the response variables the Pearson correlation will approximate the correlations in X^* (Cor. 1, Foldnes & Grønneberg, 2021). This means that the observed ordinal distributional forms of X_1, X_2, X_3 and X_4, X_5, X_6 will resemble the densities of H_1 and H_2 , respectively, depicted in Figure 6 (see also the discussion of Figure 7). According to the recommendation of Rhemtulla et al. (2012), cont-ML should not be employed with such skewed distributions. Every simulation study is limited by its choice of distributions to simulate from, and the advice of Rhemtulla et al. (2012) that cont-ML may be used with symmetrically distributed items with five or more categories is strictly speaking valid only for data generated in a manner resembling the simulations in Rhemtulla et al. (2012). Clearly, our illustrative case represents a data condition not included in earlier simulation studies.

Our aim with the illustrative case has been to analytically construct a striking condition where the lack of substantive knowledge leads cat-LS and cont-ML to grossly miscalculate the number of factors. The bias incurred by the recom-

⁶The estimator is often referred to as MLR.

mended cat-LS approach with marginal distributions closer to normality will be dealt with in the upcoming numerical studies.

Adjusting ordinal factor analysis to non-normal response marginals

We have argued that substantive knowledge of the response marginal distributions of X^* and/or the thresholds is a requirement for using ordinal covariance models. We here assume that the response marginal distributions of X^* are known to have cumulative distribution functions $F_1^*, F_2^*, \dots, F_p^*$. Also, without loss of generality, we assume that the marginal distributions of X^* have unit variance. An adjusted version of cont-ML, cont-ML-adj, that takes substantive marginal information into account was proposed by Foldnes and Grønneberg (2021). We next propose a similar adjustment to cat-LS, denoted by cat-LS-adj, derived under the assumption that the copula of each pair of latent response variables is normal.

The idea behind cont-ML-adj is to use substantive knowledge of the marginals of X^* to encode the ordinal observations in such a way that continuous factor analysis methods work as intended for a large number of categories. Foldnes and Grønneberg (2021) also introduced a modification to the polychoric correlation which assumed that the number of categories were large. We here suggest an alternative adjustment to polychoric correlations which works well also when the number of categories is small. However, the price for good performance at small K is that we assume that X^* has a normal dependence structure, that is, that X^* has a normal copula. While the present article focuses on the case when X^* has a normal copula, a brief discussion of non-normal copulas is found in “What if the copula is non-normal?” on p. 2 in the online appendix.

Recall that cat-LS first estimates the correlation matrix of the continuous response vector X^* using polychoric correlations and then use least squares estimation methods to fit the proposed model to the correlation matrix (Muthén, 1984). We propose to adjust this procedure, by replacing standard polychorics with what we call adjusted polychorics. For valid inference based on adjusted polychorics, we also need to use an adjustment to the asymptotic covariance matrix of the polychoric correlations. This adjustment is described in “Standard errors for cat-LS-adj” on p. 2 in the online appendix.

Calculating the adjusted polychorics is simple, and does not require optimization. First, the standard polychoric estimator of Olsson (1979a) is computed, using standard software. Since the polychoric correlation assumes that X^* is fully normal, it is a normal theory estimator, and we denote it by $\hat{\rho}_{NT}$. Then a function Ψ is applied to $\hat{\rho}_{NT}$, giving

$$\hat{\rho}_{adj} = \Psi(\hat{\rho}_{NT}).$$

Here, Ψ is a function which depends on the marginal distributions F_1^*, F_2^* of X^* through the formula

$$\Psi(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_{\rho=r}^N(F_1^*(x_1), F_2^*(x_2)) - F_1^*(x_1)F_2^*(x_2) dx_1 dx_2$$

where $C_{\rho=r}^N$ is the normal copula with correlation r .

If X^* has a normal copula and marginals F_1^*, F_2^* , the adjusted polychoric correlation will be consistent and asymptotically normal. Technical details are given in the appendix (p. 21). Code to compute adjusted polychorics and their asymptotic covariance matrix is provided in the online supplementary material.

Polychoric correlations systematically overestimate the size of the true response correlations

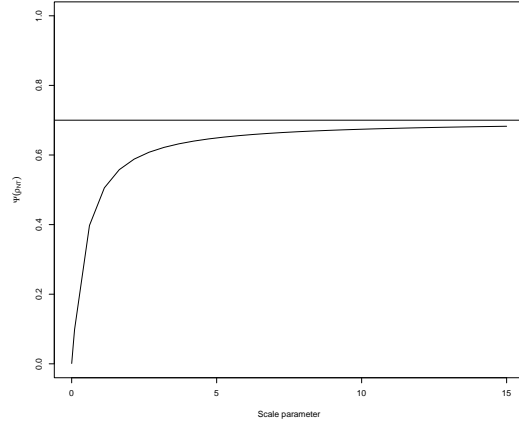


Figure 8. The function $\Psi(0.7)$ with one marginal following the standardized Gamma-distribution with varying scale parameter, and the other marginal following the same distribution but reflected around zero.

The polychoric correlation $\hat{\rho}_{NT}$ consistently estimates

$$\rho_{NT} = \rho(\Phi^{-1}(F_1^*(X_1^*)), \Phi^{-1}(F_2^*(X_2^*))) \quad (6)$$

when the copula of X_1^*, X_2^* is normal (Foldnes & Grønneberg, 2021, online appendix, “On polychoric correlations with misspecified marginals”). Here, $\rho(A, B)$ is the Pearson correlation between random variables A, B . Typically, ρ_{NT} does not equal the target correlation $\rho(X_1^*, X_2^*)$. By the continuity of Ψ , the adjusted polychoric correlation estimates $\Psi(\rho_{NT})$, which equals $\rho(X_1^*, X_2^*)$ under the conditions given in the appendix. The value of the function $\Psi(r)$ is therefore interpretable as the true response correlation when the polychoric correlation estimates the response correlation to be r .

An important and, to the best of our knowledge, new finding is that

$$|\rho(X_1^*, X_2^*)| \leq |\rho_{NT}|, \quad (7)$$

provided the underlying copula is normal. That is, unless the polychoric correlation is consistent, it will always overestimate the size of the true response correlation. This observation will shortly be proven. The overestimation will in turn bias ordinal factor models estimated via these correlations, and may lead to estimates of, e.g., unrealistically strong factor structures in applications.

Figure 8 illustrates the finding, and plots $\Psi(0.7)$ as a function of a single parameter that varies the marginal response distributions. The first marginal is the standardized Gamma-distribution with a scale parameter that varies, and the second marginal follows the same distribution as the first except that the distribution is reflected around zero. The response marginals are coupled together by a normal copula with correlation parameter .7. The straight line in the plot is $\rho_{NT} = 0.7$, the number which the polychoric correlation necessarily estimates. In contrast, the true response correlation estimated by the adjusted polychoric correlation varies from around zero up to around 0.7. For very low values of the scale parameter, which corresponds to extremely non-normal distributions, the polychoric correlation will severely overestimate the true correlation which is close to zero. For distributions closer to normal, corresponding to scale parameters some distance from 0, the degree of overestimation is vanishing. The gamma distributions in the numerical illustration both converge to normality as the scale parameter increases, and we see that for large scale parameters the observed overestimation is diminishing.

Due to eq. (7), it will never be the case that the true correlation is under-estimated by the polychoric correlation, and any similar plot with a different choice of marginals will have true correlations below the straight line at 0.7. The degree of overestimation depends on the marginals, a relation exactly quantified in the function Ψ .

Let us now explain why eq. (7) holds and how we can interpret this bound. Notice first that $\rho(X_1^*, X_2^*)$ and ρ_{NT} in eq. (6) are both of the form

$$\rho(a(X_1^*), b(X_2^*)) \quad (8)$$

for real functions a, b . The maximum correlation ρ_M of X_1^*, X_2^* is a correlation coefficient with a complex history (Klaassen & Wellner, 1997, Section 6). As far as we know, it has not been studied in the psychometric literature. It is defined as the supremum of eq. (8) over all functions a, b such that $\text{Var}(a(X_1^*))$ and $\text{Var}(b(X_2^*))$ are finite. When X_1^*, X_2^* have a normal copula, a classical result re-derived in Klaassen and Wellner (1997, Theorem 6.1) shows that this supremum is attained at $a(x) = \Phi^{-1}(F_1^*(x))$ and $b(x) = \Phi^{-1}(F_2^*(x))$. Therefore, eq. (6) gives

$$\rho_M = |\rho_{NT}|,$$

providing the claimed bound.

When the response copulas are normal but the response marginals are non-normal, the polychoric correlation there-

fore estimates not the Pearson correlation between the response marginals, but the maximum correlation between the response variables. We consider a full discussion of this new observation outside the scope of the present paper.

Adjusted polychoric correlations and sensitivity analysis

We have argued that to estimate response correlations, substantive knowledge of response marginal distributions is required. The requirement is similar to the specification of prior distributions in Bayesian statistical analysis (Berger, 1985). The prior is specified based on substantive knowledge, but in many cases, available substantive knowledge allows only a partial specification of the prior. A complete specification of the prior is required to calculate the posterior, and is constructed based on arbitrary choices (Leamer, 1985, p.309). The analysis of the consequences of these choices is known as Bayesian sensitivity analysis, see e.g. the general treatments of Berger et al. (1994); Insua and Ruggeri (2012), as well as Van Erp, Mulder, and Oberski (2018) in a psychometric setting. As shown in the upcoming Monte Carlo illustrations, estimates of response correlations and models based on such estimates may be sensitive to the marginal response distributions, motivating sensitivity analyses also when fitting ordinal factor models.

Establishing best practice methodology for such sensitivity analyses is a complex problem left to future research, and we here sketch a simple pragmatic procedure in this direction, assuming a normal response copula. Estimation and inference for the ordinal factor model should be performed with cat-LS-adj using a selection of marginals that are representative of distributions compatible with substantive knowledge. The resulting degree of robustness of statistical conclusions should be reported. For example, if the focus in a study is the statistical significance of a parameter, the p-value and the parameter estimated should be perturbed with a selection of marginals, and a plot of the resulting p-values and estimates should be studied. If the conclusion of, say, significance at a 5 % level holds under all marginal choices, the analysis indicates that this conclusion is robust. A simple illustration of a sensitivity analysis is given in Study 4 in the next section.

Studies of cat-LS and cat-LS-adj under marginal non-normality

This section comprises a sequence of five studies, some conducted at the population level, and some at the finite sample level using Monte Carlo simulation. The studies investigate how non-normal marginal distributions affect the polychoric correlation estimator and its adjusted version, and how any effect is propagated to factor analysis with cat-LS or cont-ML. Alongside, the performance of the adjusted polychoric estimator and cat-LS-adj is evaluated. Study 4 gives a simple empirical example of sensitivity analysis.

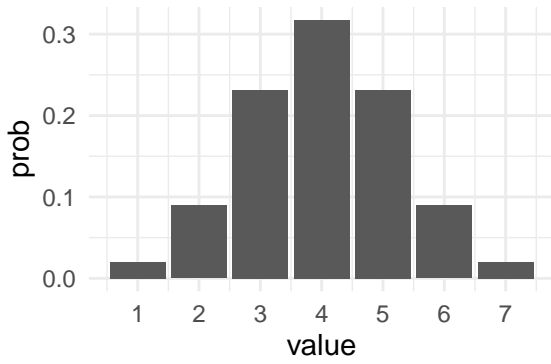


Figure 9. Probability distribution of X and Y , for $K = 7$.

Study 1: Correlation bias at the population level

In this study we examine the biasedness of Pearson and polychoric correlations in the absence of substantive knowledge. Also, we illustrate that the adjusted Pearson correlation, as the number K of categories increases, eliminates bias when substantive marginal knowledge is available. The adjusted polychoric correlation is demonstrated to be unbiased for all K . We remark that larger values of K ($K > 11$) are seldom encountered in practice, but these are included here to illustrate that the polychoric and Pearson estimator do not perform better with large K .

We consider ordinal bivariate vectors obtained by discretizing a non-normal vector $(X^*, Y^*)'$ whose copula is normal. As discussed previously, the same ordinal bivariate distribution can be obtained by discretizing (using other thresholds) a normal vector $(Z_1, Z_2)'$. The first case is taken as the true data-generating process, and we seek to calculate the correlation between X^* and Y^* . We let $X^* = H_1(Z_1)$ and $Y^* = H_2(Z_2)$ (see eq. (3)) where Z_1 and Z_2 are standard normal variables which together form a bivariate normal distribution with correlation $\rho = .7$. The correlation between X^* and Y^* is then approximately equal to 0.446 (Table 2).

We now let the number of categories increase from $K = 2$ to $K = 20$. For each K , the thresholds are taken to be equally spaced on the standard normal scale. For instance, for $K = 7$ we employ equally spaced thresholds at $-2.05, -1.23, -0.41, 0.41, 1.23$ and 2.05 . Applying H_1^{-1} and H_2^{-1} to these values yields unequally spaced values: $-0.83, -0.75, -0.66, 0.06, 1.42$ and 2.78 , and $-2.78, -1.42, -0.66, 0.66, 0.75$ and 0.83 , respectively. Discretizing $H_1(Z)$ and $H_2(Z)$ using these thresholds results in a common ordinal distribution that approximates a normal distribution, see Figure 9. For each K , in the same manner equally spaced thresholds are transformed to the X^* and Y^* scale. Approximate normality in ordinal data has been reported to be beneficial for inference when treating ordinal data as continuous (Olsson, 1979b; Rhemtulla et al., 2012). However, despite the approximately normal distribution of

the ordinal variables, we next illustrate that without substantive knowledge both Pearson and polychoric correlations are biased.

For each K we calculate four correlations: The regular Pearson and polychoric correlations, and their respective adjustments based on true substantive knowledge, i.e., that the underlying response variables X^* and Y^* follow the non-normal distributions depicted in Figure 6. Now, since the ordinal distribution of $(X, Y)'$ is consistent with discretizing a normal vector with correlation $\rho = 0.7$, the polychoric correlation will equal 0.7 for all K . The Pearson correlation will also reach this correlation as K increases, since the observed ordinal distribution is approximately normal. The adjusted correlations require that we correctly specify the marginals of X^* and Y^* . The adjusted Pearson correlation will approach zero bias with respect to the true underlying correlation 0.446 as K increases, while the adjusted polychoric correlation will be consistent for all K . This is confirmed in Figure 10, which is based on population-level calculations. The figure shows that the Pearson (for large K) and polychoric correlations are highly biased, with more than 50% relative bias. Also, incorporating correct substantive knowledge of the latent response distributions clearly reduces bias dramatically in the adjusted Pearson correlation adjustment, and totally removes bias for the adjusted polychoric correlation.

An alternative interpretation of Study 1. We here recall (Foldnes & Grønneberg, 2021, Section “Marginal Distributions and Thresholds are Confounded”) that when simulating a vector of ordinal observations through discretizing a continuous random vector, we are in a sense simulating discretizations from all possible continuous random vectors that is capable of being discretized to the ordinal observations. This means that the above simulation can just as well be interpreted as simulating from the continuous normal random vector that follows a one-factor model with all correlations equal to 0.7. Our illustration therefore symmetrically illustrates that cat-LS-adj and cont-ML-adj may reach grossly inaccurate conclusions when the marginal distributions are misspecified. This further underlines the importance of having substantive knowledge of the marginals. This type of symmetric interpretation holds for all upcoming studies.

Study 2: Monte Carlo study of the factor retention problem under substantial non-normality

In this study we examine by Monte Carlo simulations the finite-sample behaviour of nested model testing. Our context is to determine whether a one-factor solution is sufficient for the six dimensional scale in Figure 5. We know that the six-dimensional vector X^* defined in Equation (4) fits a two-factor solution depicted in the figure, but has a poor fit to a one-dimensional solution. In our simulations we repeatedly draw samples of sizes $n = 100$ and $n = 500$ from X^* , which we then discretize using the same thresholds as de-

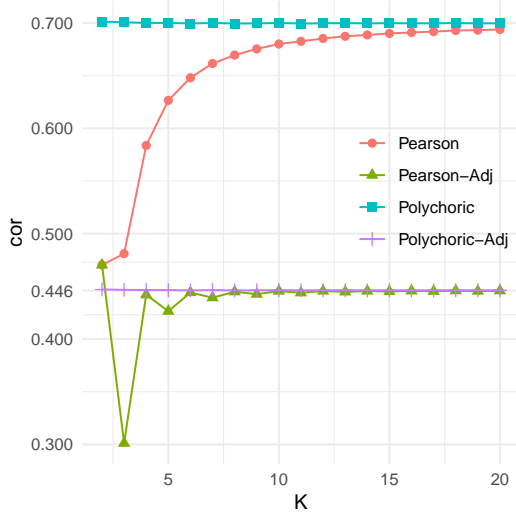


Figure 10. Population level correlation values for the Pearson and polychoric estimators, and their adjustments. K =number of categories.

scribed in the previous study, which produces symmetrical observed ordinal variables, for $K = 4$ and $K = 7$. For each of the resulting four conditions, 1000 samples were drawn and discretized. Four estimators of the 1-factor and 2-factor models were evaluated: cont-ML and cat-LS, together with their adjusted versions. Nested model testing based on the scaled-and-shifted statistic (Asparouhov & Muthén, 2010) was employed for all estimators. This test statistic, and standard error estimates, require an estimate of the asymptotic covariance matrix Γ of the correlation vector. For cont-ML-adj, Γ was estimated using bootstrapping with 1000 bootstrap samples. For cat-LS-adj, we implemented the formula deduced in the online supplementary material.

Results are provided in Table 3. Let us first consider the mean estimated correlation $\hat{\phi}$ between the factors in the 2-factor model. As expected, the unadjusted versions tend to estimate perfect correlation between the factors, indicating that only one factor is necessary to describe the scale. This is reflected in the unadjusted estimators having rejection rates close to the nominal Type I error rate associated with the nested model test. For cat-LS-adj, the mean $\hat{\phi}$ value is close to the true value 0.692, while the rejection rates of the nested model test are close to 100%, even at $n = 100$. It is seen that cont-ML-adj improves upon cont-ML, but it is still markedly biased at $K = 4$, with reduced bias at $K = 7$. At $K = 7$ cont-ML-adj almost always rejects the misspecified 1-factor model, even at the smallest sample size. In order to validate the standard errors based on new estimates of Γ , Table 3 also reports the % relative bias of standard errors, and it is seen that bias reduces generally with increasing n .

We conclude that the adjusted estimators performed well

in identifying a 1-factor model as untenable, in contrast to the unadjusted estimators. Also, the adjusted standard error estimation procedures were found to perform quite well at $n = 500$.

Table 3

Study 3: Monte Carlo evaluation of four estimators.

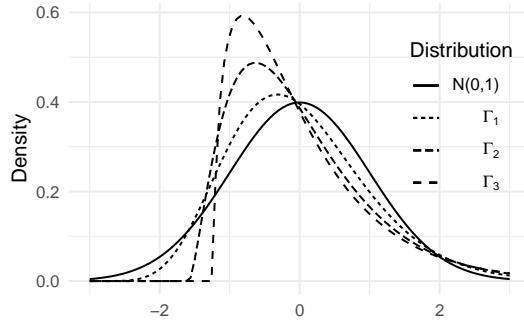
Estimator	K	n	Mean of $\hat{\phi}$	RB	SE RB	Rej
cat-LS	4	100	1.00	0%	-1%	0.05
		500	1.00	0%	1%	0.04
	7	100	1.00	0%	-4%	0.06
		500	1.00	0%	0%	0.04
cat-LS-adj	4	100	0.69	-1%	-5%	1.00
		500	0.69	0%	0%	1.00
	7	100	0.69	0%	-8%	1.00
		500	0.69	0%	4%	1.00
cont-ML	4	100	1.00	0%	-7%	0.06
		500	1.00	0%	-3%	0.05
	7	100	1.00	0%	-7%	0.07
		500	1.00	0%	-1%	0.04
cont-ML-adj	4	100	0.80	16%	-16%	0.62
		500	0.81	17%	-3%	1.00
	7	100	0.72	4%	-14%	0.97
		500	0.72	4%	0%	1.00

Note. K = number of categories. n = sample size. RB = Relative bias of $\hat{\phi}$. SE RB = relative bias of standard error of $\hat{\phi}$. The targeted estimand for $\hat{\phi}$ is .692. Rej = Rejection rate of chi-square difference test.

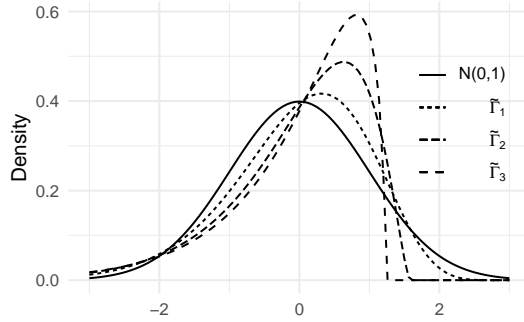
Study 3: Monte Carlo study of polychoric inference under moderate non-normality

The previous study involved a rather striking case where marginal non-normality was carefully chosen to illustrate a breakdown of conventional ordinal inference. In this study we consider the finite-sample performance of the polychoric correlation and its adjustment under a set of mild to moderate deviations from marginal normality, depicted in Figure 11a. The densities $\Gamma_1 - \Gamma_3$ are standardized gamma distributions which are increasingly non-normal, with skewness 0.6 and excess kurtosis 0.6 for Γ_1 , skewness 1.3 and excess kurtosis 2.4 for Γ_2 , and skewness 1.6 and excess kurtosis 4.0 for Γ_3 . These values of skewness and excess kurtosis represent mild to moderate non-normality (Kline, 2015, p. 76). To include negatively skewed distributions, reversed versions $\tilde{\Gamma}_1 - \tilde{\Gamma}_3$ are included in the design, see Figure 11b.

Marginal variables were selected from $N(0, 1), \Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2$, and $\tilde{\Gamma}_3$ to form a total of 28 pairs. The pairs represent a range of distributions with marginals spanning from normal to moderate non-normality, and with skewness in the marginals being positive or negative, and of the same, or of opposite, signs. Each pair was coupled together by a normal copula, calibrated with the vita function in package covsim (Grønneberg, Foldnes, & Marcoulides, 2021), so that the resulting continuous



(a) The standard normal and scaled gamma distributions.



(b) The standard normal and scaled and reversed gamma distributions.

Figure 11. The standard normal $N(0, 1)$ and standardized gamma distributions Γ_1, Γ_2 , and Γ_3 , with shape parameters 10, 2.5, and 1.5, respectively. The associated reversed distributions are denoted by $\tilde{\Gamma}_1, \tilde{\Gamma}_2$, and $\tilde{\Gamma}_3$.

bivariate distribution had (underlying) correlation .2, .4 and .7. Three sample sizes were specified: $n = 100, 300, 1000$. In each of the resulting $28 \cdot 3 \cdot 3 = 252$ conditions, 1000 samples were generated. Each generated continuous data sample was then discretized with thresholds specific to each of the seven marginal distributions, chosen so that in the population the resulting ordinal marginals distributions had seven symmetrically distributed levels, see Figure 12. That is, for every simulated dataset of sufficiently large sample size, a plot of a marginal variable would resemble the distribution in Figure 12.

The polychoric correlation and its standard error, as well as the adjusted polychoric correlation and its adjusted standard error, were calculated in each generated ordinal sample. Also, the 95% confidence interval (CI) for the underlying correlation was calculated based on the polychoric estimator and on its adjusted value. Relative bias for the two estimators are depicted in Figure 13, at the largest sample size $n = 1000$. The results are essentially the same for the smaller sample sizes, and are not reported here. We remark that bias in all conditions was positive, thereby verifying that the polychoric correlation will overestimate the size of the true underlying

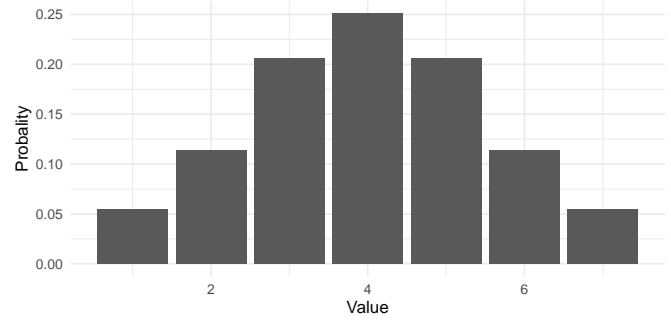


Figure 12. Study 3: Symmetric ordinal distribution with seven levels.

correlation.

Under bivariate normality (both marginals $N(0, 1)$), there is no bias. However, when at least one of the marginals departs from $N(0, 1)$, the polychoric estimator is biased. The overall relative bias associated with the polychoric estimator is 7.3%, across all 28 pairs, sample sizes and ρ values. The adjusted polychoric estimator, in contrast, is unbiased, with overall a mean relative bias of -0.1% . The polychoric estimator is severely biased when the marginals are skewed in opposite directions. For instance, with marginal distributions Γ_2 and $\tilde{\Gamma}_2$, the relative bias across all sample sizes of the polychoric estimator is 10.8%, 14.1%, and 17.4% for $\rho = .2, .4$, and $.7$, respectively. In the most extreme condition, with oppositely skewed marginals Γ_3 and $\tilde{\Gamma}_3$, and with underlying correlation $\rho = .7$, the polychoric bias was 33% across all sample sizes. This means that the mean polychoric estimate was .93, missing severely the true underlying value of .7. Although bias was not associated with sample size, there was an association between the underlying correlation ρ and polychoric estimation bias, as exemplified above with Γ_2 and $\tilde{\Gamma}_2$. In general, when the marginals were oppositely skewed, the higher underlying correlation lead to higher bias, while for marginals skewed in the same direction, an increase in underlying correlation led to decreased bias.

Coverage rates for CIs at the 95% level of confidence are reported in Table 4 for the largest sample size⁷. The adjusted polychoric correlation attains the nominal rate in all conditions. The polychoric correlation are close to the nominal rate only in mildly non-normal conditions. When the underlying correlation ρ increases, coverage rates tend to fall markedly below the nominal 95% level, especially when the marginals are skewed in opposite directions. For instance, when $\rho = .4$, under $\tilde{\Gamma}_3 - \Gamma_1$ the coverage is 62%. When $\rho = .7$ there are conditions where the polychoric confidence interval failed to cover the value .7 in all the 1000 simulated datasets. The polychoric estimator had acceptable coverage

⁷Tables for $n = 100$ and $n = 300$ are provided in the online supplementary material.

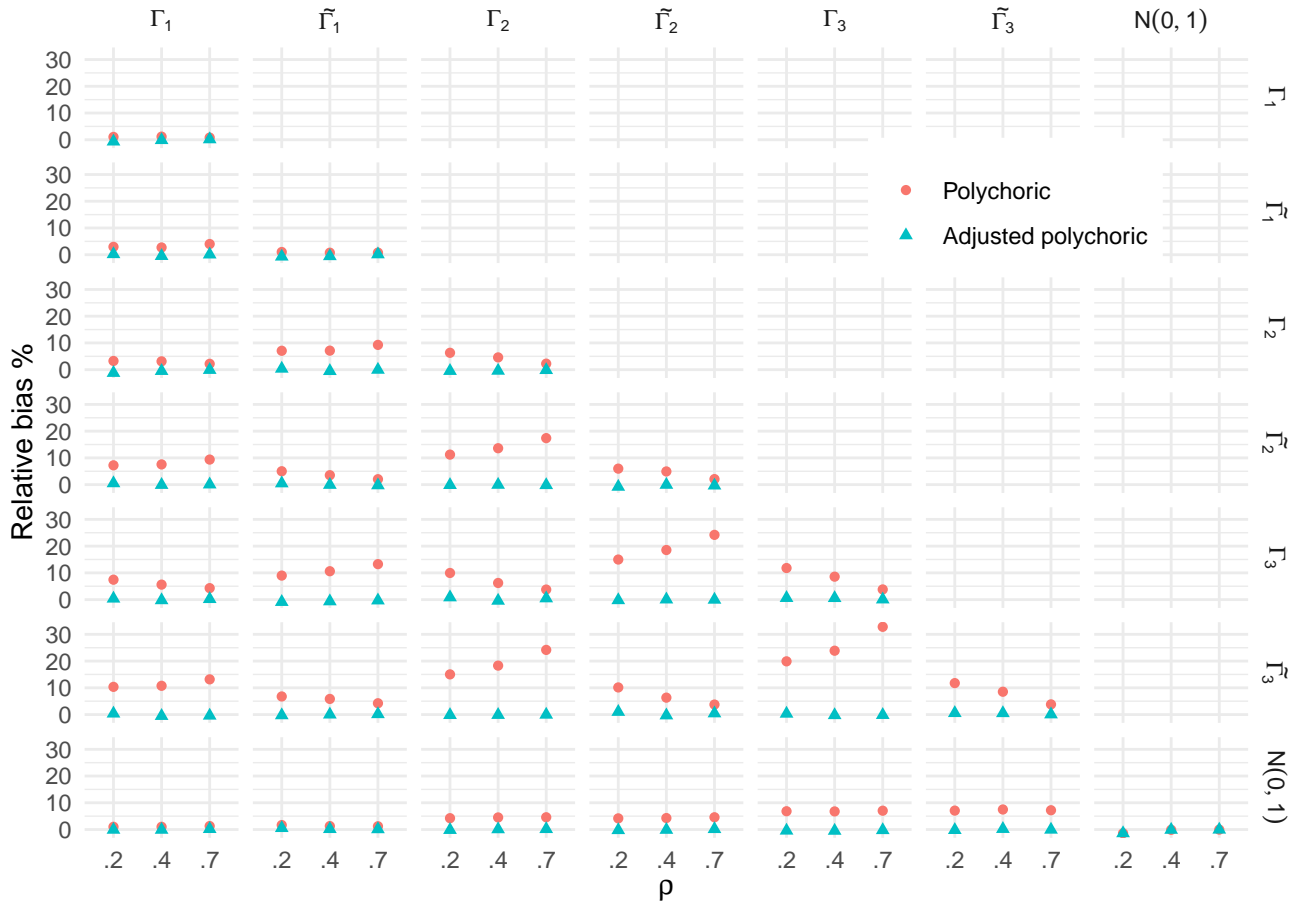


Figure 13. Study 3: Relative bias for the polychoric estimator and its adjusted version as a function of marginal distributions. Sample size $n = 1000$.

in only one condition, namely the bivariate normal condition. For the smaller sample sizes, coverage rates expectedly were higher for the polychoric estimator, given the larger standard error estimates under smaller sample sizes. For $n = 100$, the overall polychoric estimator coverage rates (across all 28 distributions) were 91%, 89%, and 70% for $\rho = .2, .4$, and $.7$, respectively, while the overall adjusted polychoric coverage rates were 92%, 92%, and 93%, for $\rho = .2, .4$, and $.7$, respectively. For $n = 300$, the overall polychoric estimator coverage rates (across all 28 distributions) were 93%, 86%, and 61% for $\rho = .2, .4$, and $.7$, respectively, while the overall adjusted polychoric coverage rates were 94%, 94%, and 94%, for $\rho = .2, .4$, and $.7$, respectively.

Study 4: Sensitivity to marginal non-normality in an empirical case

Included in the MPsychoR package (Mair, 2020) is provided a dataset taken from Mair et al. (2015) who examined motivation among R package developers. In this study we follow Mair (2018, Chapter 2.4.1) and consider a two-

factor model for intrinsic and extrinsic motivation, where intrinsic motivation was measured by five binary indicators, and extrinsic motivation by 12 binary indicators. The sample size is $n = 794$, and our focus in this sensitivity analysis is the correlation between intrinsic and extrinsic motivation. Cat-LS estimation based on the tetrachoric correlations⁸ yields an inter-factor correlation of .148 with a p-value of .002. To conduct a sensitivity analysis with respect to the assumption of normal marginals, we need to specify non-normal marginal distributions corresponding to each of the 17 marginals. This may be done in many ways. Our focus here is to illustrate, so we posit that the departure from marginal non-normality is not extreme, and that each marginal may be distributed according to one of $\Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3$ defined above in Study 3. We follow Mair (2018, Chapter 2.4.1) and assume that the underlying bivariate copulas are normal. Since the items are binary, there is to the best of our knowledge presently no available

⁸Polychoric correlations with binary data are known as tetrachoric correlations (Pearson, 1900).

Margins	Polychoric			Adjusted polychoric		
	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	94.4	95.5	91.8	94.3	96.0	95.0
$\tilde{\Gamma}_1\Gamma_1$	94.1	93.7	57.4	94.7	95.3	94.3
$\tilde{\Gamma}_1\tilde{\Gamma}_1$	94.0	94.1	92.6	94.1	95.3	95.3
$\Gamma_2\Gamma_1$	93.2	92.6	82.9	93.7	94.3	95.0
$\Gamma_2\tilde{\Gamma}_1$	92.4	79.9	1.1	93.8	93.9	95.3
$\Gamma_2\Gamma_2$	92.2	88.4	83.4	94.4	94.5	95.5
$\tilde{\Gamma}_2\Gamma_1$	90.7	78.4	1.4	93.3	95.9	95.2
$\tilde{\Gamma}_2\tilde{\Gamma}_1$	92.9	90.9	83.3	95.2	94.7	93.9
$\tilde{\Gamma}_2\Gamma_2$	87.4	46.3	0.0	96.0	93.8	93.6
$\tilde{\Gamma}_2\tilde{\Gamma}_2$	92.3	88.6	83.4	93.8	95.0	95.0
$\Gamma_3\Gamma_1$	93.3	86.6	51.1	94.4	95.8	95.2
$\Gamma_3\tilde{\Gamma}_1$	90.6	61.4	0.0	93.9	95.7	94.9
$\Gamma_3\Gamma_2$	89.6	83.9	60.0	93.3	94.6	94.1
$\Gamma_3\tilde{\Gamma}_2$	81.6	21.2	0.0	92.7	94.0	94.9
$\Gamma_3\Gamma_3$	87.4	73.9	60.5	93.8	95.1	94.7
$\tilde{\Gamma}_3\Gamma_1$	87.7	62.1	0.0	95.0	95.6	95.7
$\tilde{\Gamma}_3\tilde{\Gamma}_1$	92.6	85.3	54.1	95.1	94.2	94.8
$\tilde{\Gamma}_3\Gamma_2$	83.2	19.6	0.0	93.6	93.4	93.8
$\tilde{\Gamma}_3\tilde{\Gamma}_2$	90.7	84.0	63.2	95.8	94.4	93.9
$\tilde{\Gamma}_3\Gamma_3$	74.0	4.3	0.0	95.1	93.9	94.2
$\tilde{\Gamma}_3\tilde{\Gamma}_3$	87.5	72.9	60.4	94.5	94.6	95.1
$N\Gamma_1$	95.0	94.0	90.0	95.5	94.5	94.3
$N\tilde{\Gamma}_1$	94.3	94.8	91.0	94.4	95.0	95.4
$N\Gamma_2$	93.9	87.1	47.4	95.2	94.4	94.1
$N\tilde{\Gamma}_2$	94.1	89.4	48.5	95.4	94.9	95.1
$N\Gamma_3$	91.7	80.7	12.8	93.9	94.7	95.2
$N\tilde{\Gamma}_3$	92.7	79.8	12.0	95.9	96.0	93.2
NN	94.8	95.2	94.8	94.8	95.2	94.8
Mean	90.7	75.9	47.3	94.5	94.8	94.7

Table 4

Study 3: Coverage rates at the 95% level of confidence, sample size $n = 1000$.

implementation of tests for underlying normality, such as the test from Muthén and Hofacker (1988). After having specified the 17 marginals, we assume a normal copula, and employ the transform in Equation (6) to the tetrachoric correlation calculated in the original sample, to calculate a new correlation matrix. Then the two-factor model is estimated from this new correlation matrix, and its associated asymptotic covariance matrix. To each specification of the 17 marginals corresponds a new inter-factor correlation and an associated p-value. The p-value is calculated based on an adjusted asymptotic covariance matrix derived in the online appendix. To get a representative set of marginal configurations, we randomly generated 1000 marginal configurations and calculated the inter-factor correlation and its associated p-value in each configuration. The result is depicted in Figure 14, where it is seen that the inter-factor correlation is rather robust to moderate underlying marginal non-normality. Also, the p-value is not markedly affected by deviations from underlying non-normality. We may therefore conclude that our estimate of .148 and its statistical significance is not very sen-

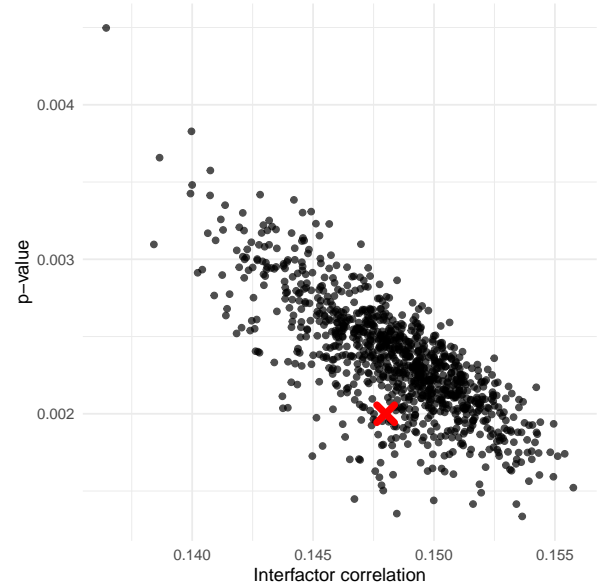


Figure 14. Study 4: Sensitivity analysis of the correlation between intrinsic and extrinsic motivation in an empirical dataset. The red cross represents the values obtained from the empirical dataset using cat-LS.

sitive to moderate marginal non-normality. We note that our illustration concerns only one parameter in the model.

Study 5: The consequences of mild marginal misspecification at the population level on cat-LS

We now consider a population study where we generate distributions that exactly follow a two-factor model. Each generated distribution has different non-normal response distribution. This is done across four conditions generated by varying some of the parameters in the two-factor model. We then study how cat-LS performs in terms of inter-factor correlation bias and RMSEA. The variation in this study is driven by the different marginal distributions within each of the four conditions, and not by sampling error. Therefore, the obtained inter-factor correlation and RMSEA values may be considered population values. We therefore omit cat-LS-adj in this study, since it will always reach the population ϕ value, and since its RMSEA will always equal zero.

This study employed a two-factor model as illustrated in Figure 5. Four conditions were specified by varying the residual error variances and the inter-factor correlation ϕ . The residual error variances θ_δ were either all equal to .4 or to .7, while ϕ was equal to either .4 or to .7. Factor variances and factor loadings were all equal to one in all four conditions. Each of the four specifications imply a correlation ma-

trix:

$$\begin{aligned}\Sigma_{\theta_\delta=0.4, \phi=.4} &= \begin{pmatrix} 1 & & & & & \\ .71 & 1 & & & & \\ .71 & .71 & 1 & & & \\ .29 & .29 & .29 & 1 & & \\ .29 & .29 & .29 & .71 & 1 & \\ .29 & .29 & .29 & .71 & .71 & 1 \end{pmatrix} \\ \Sigma_{\theta_\delta=0.7, \phi=.4} &= \begin{pmatrix} 1 & & & & & \\ .59 & 1 & & & & \\ .59 & .59 & 1 & & & \\ .24 & .24 & .24 & 1 & & \\ .24 & .24 & .24 & .59 & 1 & \\ .24 & .24 & .24 & .59 & .59 & 1 \end{pmatrix} \\ \Sigma_{\theta_\delta=0.4, \phi=.7} &= \begin{pmatrix} 1 & & & & & \\ .71 & 1 & & & & \\ .71 & .71 & 1 & & & \\ .50 & .50 & .50 & 1 & & \\ .50 & .50 & .50 & .71 & 1 & \\ .50 & .50 & .50 & .71 & .71 & 1 \end{pmatrix} \\ \Sigma_{\theta_\delta=0.7, \phi=.7} &= \begin{pmatrix} 1 & & & & & \\ .59 & 1 & & & & \\ .59 & .59 & 1 & & & \\ .41 & .41 & .41 & 1 & & \\ .41 & .41 & .41 & .59 & 1 & \\ .41 & .41 & .41 & .59 & .59 & 1 \end{pmatrix}\end{aligned}$$

We generated 1000 random configurations of marginal distributions, where each marginal was randomly and independently chosen from $\Gamma_1, \Gamma_2, \Gamma_3, \tilde{\Gamma}_1, \tilde{\Gamma}_2, \tilde{\Gamma}_3$. For each marginal configuration a large sample of size $n = 2 \cdot 10^5$ was drawn from a six-dimensional distribution whose correlation matrix was equal to $\Sigma_{\theta_\delta, \phi}$, with marginal distributions dictated by the marginal configuration. Hence, a very large sample of continuous data with non-normal marginals was drawn from a population whose correlation matrix was equal to $\Sigma_{\theta_\delta, \phi}$, so that the two-factor model fitted perfectly. Next, the sample was discretized into an ordinal dataset with $K = 10$ categories in each variable. The thresholds were chosen separately for each marginal so that the ordinal distribution was uniform, i.e., the thresholds are selected in such a way that each of the ten categories are equally likely. Then, we fitted the two-factor model with cat-LS to the ordinal dataset, and extracted the inter-factor correlation $\hat{\phi}$ and the RMSEA model fit index. The RMSEA is a measure of accumulated bias that is contained in the polychoric correlation matrix in relation to estimating the two-factor model. It is often used as a measure of closeness of fit, where values exceeding 0.05 are typically interpreted as lack of close fit. This cut-off is based on the normal-theory ML fit function, which differs from the DWLS fit function utilized in cat-LS. The cut-off 0.05 is used in the present study, since the DWLS based RMSEA tends to be lower than the normal-theory ML based RMSEA (Xia & Yang, 2019). Therefore, lack of close fit using the DWLS based RMSEA in general implies lack of poor fit under the normal-theory ML based RMSEA. The

very large sample size ensures that $\hat{\phi}$ and the RMSEA value are approximately equal to their population counterparts.

The results are depicted in Figure 15. It is seen that both bias and lack of close fit become more substantial in the condition with largest underlying correlations, namely $\theta_\delta = 0.4$ and $\phi = .7$. In this condition two-thirds of the marginal configurations result in lack of poor fit, and there are some marginal configurations where $\hat{\phi}$ exceeds .8 (see Table 5). The most extreme bias occurs when all response distributions in one factor are skewed in the same direction, while the response distributions in the second factor are all skewed in the opposite direction. For instance, when the first three response variables are Γ_3 distributed, while the last three response variables are $\tilde{\Gamma}_3$ distributed, the inter-factor correlation was estimated at .853. This exceeds the max value reported in Table 5, so we may infer this extreme configuration was not among the 1000 randomly chosen configurations.

Overall, the estimated $\hat{\phi}$ does not stray markedly from its target value, as indicated by short interquartile ranges and close-to-nominal median values in all four conditions. In contrast, model fit assessment is adversely affected by moderate marginal non-normality. Only in the condition with the lowest correlations, i.e., $\theta_\delta = .7$ and $\phi = .4$, did the model pass the RMSEA test with high certainty. Note that all distributions simulated from in this study fitted perfectly to the two-factor model. Therefore, it is problematic that a correctly specified model was indicated as lacking good fit in an overall of 31% of all 4000 distributions generated.

ϕ	θ_δ	$\hat{\phi}$				RMSEA > .05 %
		Min	Max	Median	IQR	
.4	0.4	.372	.458	.395	.019	23.3
.4	0.7	.376	.448	.396	.016	0.2
.7	0.4	.664	.830	.697	.031	67.3
.7	0.7	.660	.804	.693	.025	39.9

Table 5

Study 5: Summary statistics for the inter-factor correlation $\hat{\phi}$ across four conditions, and the percentage of RMSEA exceeding the close fit value of .05. IQR=interquartile range.

Discussion

Recently, progress has been made concerning the sensitivity of ordinal factor and structural equation modeling to underlying non-normality. Foldnes and Grønneberg (2021) focused on region D in Figure 1, with normal marginal distributions and non-normal response copula, and found that the polychoric estimator may be substantially biased, leading to bias in estimation and fit assessment in a medium sized structural equation model. These new insights into the sensitivity of cat-LS and cont-ML to underlying non-normality were made possible by employing newly developed simulation methodology. Fortunately, such conditions may be reli-

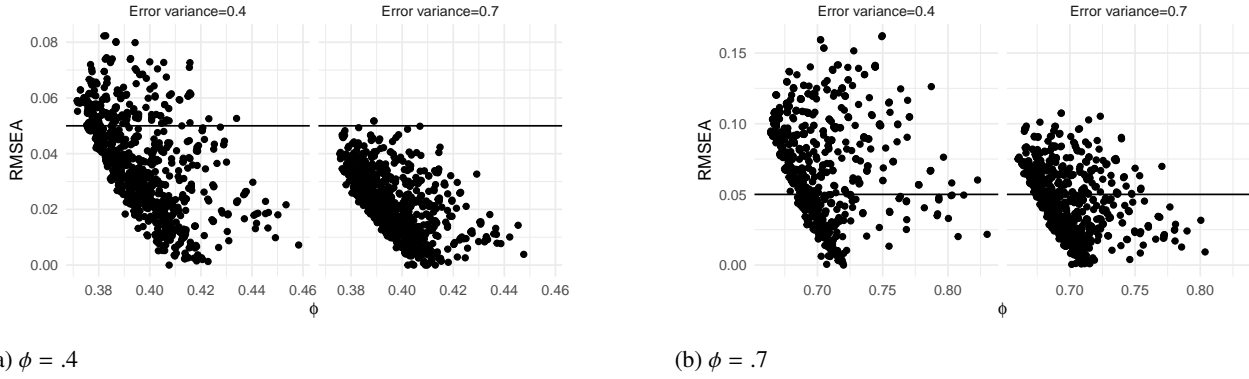


Figure 15. Study 5: Population-level analysis from 1000 randomly generated non-normal marginal configurations in each panel. The horizontal line represents the RMSEA cutoff 0.05 for close model fit. Upper and lower panels correspond to the inter-factor correlation ϕ being equal to .7 and .4, respectively.

ably detected in ordinal data by using, e.g., a bootstrap test (Foldnes & Grønneberg, 2019b).

In the present study we considered the reverse kind of underlying non-normality, as represented by Region B in Figure 1, where the copula is normal, and the response marginals are non-normal. Through the use of illustrative examples and probability arguments, we have argued that such conditions are impossible to detect by purely statistical means. It is therefore of importance to evaluate the extent to which cat-LS is sensitive to non-normality in the response marginal distributions. We have seen that if the response marginals specified in the estimation methodology is far away from the true marginals, the resulting factor model estimates may be severely biased. This highlights the need for substantive knowledge concerning the response marginal distributions. While exact knowledge of the marginals enables the analyst to reach the strongest type of conclusions, we have also illustrated that substantive knowledge of a more approximate form may be incorporated through the use of sensitivity analysis methods. For example, if marginals are known to be close to normal, the analyst may try out cat-LS-adj using different approximately normal marginal distributions and inspect how stable the statistical conclusions are. The cost of this is that the analyst ends up with one conclusion per marginal configuration, and standard methodology such as reporting a single p-value is no longer possible to follow.

A natural extension of the present research is to consider the case when neither the copula class nor the marginals are known exactly. Some notes on this case is given in the online appendix “What if the copula is non-normal?” (p. 2). We consider the full development of this problem outside the scope of the present paper.

Practitioners ought to start their ordinal data analysis with a test for the normality of the copula of the response vector using the bootstrap test of Foldnes and Grønneberg (2019b) as implemented in the R package *discnorm* (Foldnes &

Grønneberg, 2020). If the test for underlying copula normality is passed, and this is compatible with substantive knowledge, the material in the present article is highly relevant.

In contrast, if the test for underlying copula normality is clearly rejected and the copula is unknown, there are to the best of our knowledge just three available options in the literature, all derived in Foldnes and Grønneberg (2021), and all assuming that the number of categories K is large: First, cont-ML may be used if equally spaced thresholds are known to hold. And cont-ML-adj and cat-LS-thr (both reviewed in the online appendix of the present paper) can be used when the response marginals are known. Both cont-ML-adj and cat-LS-thr can be used in sensitivity analyses if the marginals are not known exactly.

If the number of categories K is not large, knowledge of both the response marginal distributions and of the copula is required. Otherwise the underlying correlation matrix is not identified from the ordinal data. For example, if the response marginals are known, but the response copula is unknown, we may calculate the set of possible correlation values an underlying vector may have and still be able to generate the ordinal dataset (Grønneberg & Moss, 2021; Grønneberg et al., 2020). These sets are unfortunately always large, and this is in contrast to the single value required to use cat-LS.

The present work shows that using methodology which wrongly assumes normal response marginals may reach statistically invalid conclusions. Wrongly assuming other marginals will also be problematic, and this can happen with cat-LS-adj. As mentioned at the end of Study 1, our numerical illustrations can be interpreted in a symmetric fashion by viewing the the true response distribution as having normal marginals. They therefore also illustrate how the cat-LS-adj is sensitive towards response marginal misspecification when the response marginals are normal but non-normal marginals are specified.

Our focus has been solely on traditional covariance based

ordinal factor models. Other types of factor models exist, for instance the factor copula model (Nikoloulopoulos & Joe, 2015). This model is based solely on underlying copulas, and therefore has identified parameters without assumptions on continuous marginals. There is in general no simple correspondence between a factor copula model and a covariance based copula model, except under normality. Also, the covariance calculations that usually motivate ordinal covariance models do not lead directly to a factor copula model.

Conclusion

We have provided an analysis of ordinal factor analysis with non-normal response marginal distributions, mainly under the assumption that the response copula is normal. Current software operates on the assumption that these distributions are normal, and given the impossibility of statistically testing for non-normal response marginal distributions, we deemed it important to examine whether cat-LS is sensitive to their occurrence. We provide empirical support for the relevance of non-normal response distributions, and show that their effect on cat-LS may be highly detrimental. Using an analytically constructed example, we show that the important factor retention problem in ordinal factor analysis cannot be solved through fully empirical means. The factor retention problem is therefore shown to rest on substantive knowledge not derivable through statistical means. We have shown through a series of illustrations and studies that ordinal factor analysis rests on substantive knowledge about the response marginal distributions provided by an expert. The more exact such knowledge is, the more exact conclusions can be reported. With only partial knowledge of the response marginal distributions, sensitivity analysis can help the analyst demarcate the robustness of the statistical conclusions. Methodological and theoretical developments for specifying response marginals distributions and for assessing the consequences of erroneous response distributional assumptions should be an important topic for further research, as the sensitivity of cat-LS and related techniques to response marginal distributional misspecification may have important consequences in foundational topics in psychological and behavioral research, such as scale development.

Appendix

A formal definition of the adjusted polychorics

Since polychoric correlations are based only on bivariate information, we consider the case when $p = 2$ without loss of generality. We note that our adjustment is not tied to the polychoric correlation, but will work with any consistent procedures for the covariance matrix of X^* when assuming normality, such as full information estimates.

Assumption 1. 1. We assume that F_1^*, F_2^* are the marginals of $X^* = (X_1^*, X_2^*)'$, and that they are standardized, continuous, and strictly increasing.

2. We assume that $(X_1^*, X_2^*)'$ has a normal copula.

We wish to estimate the underlying correlation, namely

$$\rho_{X^*} = \text{Cov}(X_1^*, X_2^*), \quad (9)$$

which equals the correlation of X_1^*, X_2^* since F_1^*, F_2^* are standardized.

Assume that we observe n independent and identically distributed random vectors, distributed as $X = (X_1, X_2)'$ fulfilling Equation (1). Based on these observations, we compute the full or the two-step normal theory polychoric correlation of Olsson (1979a) using standard software. Let the resulting estimator be denoted $\hat{\rho}_{\text{NT}}$. Unless F_1^*, F_2^* are standard normal distributions, $\hat{\rho}_{\text{NT}}$ will be an inconsistent estimator of ρ_{X^*} . We here derive a correction for $\hat{\rho}_{\text{NT}}$ to estimate ρ_{X^*} in a consistent manner.

Recall that X^* has a normal copula with correlation ρ_{Z^*} if

$$Z^* = (\Phi^{-1}(U_1^*), \Phi^{-1}(U_2^*))$$

is bivariate normal with standardized marginals and correlation ρ_{Z^*} , where

$$(U_1^*, U_2^*)' = (F_1^*(X_1^*), F_2^*(X_2^*))'.$$

From Sklar's Theorem, the copula and the marginal distributions uniquely describe the full probability distribution of X^* , and hence also X (Nelsen, 2007, Chapter 2.3). The parameter of interest is the Pearson correlation ρ_{X^*} of X^* . The normal copula is parameterized by the Pearson correlation ρ_{Z^*} of Z^* . Since F_1^*, F_2^*, Φ^{-1} are all strictly increasing functions from Assumption 1, and copulas are preserved under strictly increasing coordinate-wise transformations (Nelsen, 2007, Theorem 2.4.3), X^* and Z^* have the same copula, and X^* and Z^* can both be discretized to X as shown in Equation (5) (p. 9).

Therefore, the normal theory polychoric correlation $\hat{\rho}_{\text{NT}}$ will consistently estimate the correlation of Z^* , which is the parameter of the normal copula. It is a ML type estimator for ρ_{Z^*} , and the parameter we wish to estimate is ρ_{X^*} of Equation (9). The relation between these parameters is given by the following result, whose proof is in the appendix.

Proposition 1. Under Assumption 1, we have that

$$\rho_{X^*} = \Psi(\rho_{Z^*}),$$

where

$$\Psi(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_r(F_1^*(x_1), F_2^*(x_2)) - F_1^*(x_1)F_2^*(x_2) dx_1 dx_2,$$

and C_r is the normal copula with correlation r . We have that Ψ is a strictly increasing function, and that

$$\Psi'(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{2,r}(\Phi^{-1}(F_1^*(x_1)), \Phi^{-1}(F_2^*(x_2))) dx_1 dx_2$$

where $\phi_{2,r}$ is the density of a bivariate normal random vector with standardized marginals and correlation r .

Since Ψ in the proposition is strictly increasing, estimating $\rho_{X^*} = \Psi(\rho_{Z^*})$ is a reparametrization of the previous ML estimator $\hat{\rho}_{NT}$ for ρ_{Z^*} . Since ML estimators are invariant under reparametrizations, the ML estimator for ρ_{Z^*} equals

$$\hat{\rho}_{X^*} = \Psi(\hat{\rho}_{NT}). \quad (10)$$

This is the adjusted polychoric correlation. A simple adjustment for calculating standard errors for adjusted polychorics, and therefore for cat-LS-adj, is given in the online appendix “Standard errors for cat-LS-Adj” on p. 2.

The impossibility of detecting marginal non-normality in the response variables

We here explain why and in what sense it is impossible to detect marginal non-normality in the response variables. This influences the interpretation of tests for underlying normality, as all such tests can do is test whether the copula evaluated at certain points is compatible with normality.

Suppose we have access to unlimited repeated observations from X . Without further substantive knowledge, the multivariate cumulative distribution function of X , given by $F_X(x_1, \dots, x_p) = P(X_1 \leq x_1, \dots, X_p \leq x_p)$ encodes everything we may know statistically, and therefore also what we can statistically say about the underlying X^* without substantive knowledge. We assume that X^* is a continuous random vector.

Let the response variables in the vector X^* have marginal distributions F_1^*, \dots, F_p^* , and recall that the copula C^* of X^* is the cumulative distribution of $(F_1^*(X_1^*), \dots, F_p^*(X_p^*))$ (Joe, 1997; Nelsen, 2007). By Equation (1) on p. 3, we have $p_i(x_j) = P(X_i \leq x_j) = P(X_i^* \leq \tau_{i,x_j}) = F_i^*(\tau_{i,x_j})$. The values $p_i(x_j)$ are known from F_X , and hence known. However, if both the thresholds and the marginal distributions are unknown, we cannot conclude anything about either from the relations $p_i(x_j) = F_i^*(\tau_{i,x_j})$, since each F_i^* is a continuous cumulative distribution function, which attains any value between zero and one.

Also by Equation (1), we have

$$\begin{aligned} F_X(x_1, \dots, x_p) &= P(X_1^* \leq \tau_{1,x_1}, \dots, X_p^* \leq \tau_{p,x_p}) \\ &= P(F_1^*(X_1^*) \leq F_1^*(\tau_{1,x_1}), \dots, F_p^*(X_p^*) \leq F_p^*(\tau_{p,x_p})) \\ &= C^*(F_1^*(\tau_{1,x_1}), \dots, F_p^*(\tau_{p,x_p})) \\ &= C^*(p_1(x_1), \dots, p_p(x_p)). \end{aligned}$$

This shows that C^* has known values at the points $(p_1(x_1), \dots, p_p(x_p))$ for varying x_1, \dots, x_p . But even if we were to know the copula C^* exactly, Sklar’s theorem (Nelsen, 2007; Sklar, 1959) implies that this knowledge would not provide us with any information concerning the thresholds or the marginal distributions, as any copula can be connected to any marginal distribution and still result in a valid probability distribution. Our knowledge of X^* is therefore restricted to knowing the value of C^* at some points, and that $p_i(x_j) = F_i^*(\tau_{i,x_j})$, a statement where thresholds and marginal distributions are confounded. We cannot say anything about the marginals F_1^*, \dots, F_p^* or the thresholds, except for this confounded statement.

In study 2, each response variable is a strictly increasing transformation of a corresponding coordinate belonging to a fully normal random vector. This means that the copula of X^* is normal, see Nelsen (2007, Theorem 2.4.3) and Foldnes and Grønneberg (2015). As seen above, our knowledge of X^* from X does give some restrictions on the copula C^* . But in our illustration, C^* is exactly normal. Since the one- and two-factor solutions differ only in their response marginals, which are completely unknown to us if our knowledge comes only from observing X , it is impossible in principle to empirically detect that the fully normal one-factor solution is not correct.

Finally, we mention that while it is impossible to test marginal normality when only knowing the distribution of X , it may be possible to test marginal normality if we know something about the thresholds. We have knowledge contained in the confounded statement $p_i(x_j) = F_i^*(\tau_{i,x_j})$, which opens up for the possibility of using substantive knowledge of the thresholds to test statements of the marginals. Conceptually, this possibility is similar to being able to test the goodness of fit in confirmatory factor models: This test requires that the factor model is over-identified, and not just identified. While the identification of the model requires substantive knowledge, having an over-identified model requires more substantive knowledge than what is required to reach a just identified model. When combined, we may reach testable implications of substantive knowledge, yet the substantive knowledge all in all required to apply a confirmatory factor model cannot be tested or derived from data alone.

References

- Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. *Mplus technical appendix*, 1–8.
- Bartholomew, D. J. (2007). Three faces of factor analysis. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (Vol. 100, pp. 9–21). Erlbaum.
- Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. Chapman and Hall/CRC.

- 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.
- Bentler, P. (2006). *Eqs 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Berger, J. O. (1985). *Statistical decision theory and bayesian analysis*. Springer-Verlag New York. doi: 10.1007/978-1-4757-4286-2
- Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., ... others (1994). An overview of robust bayesian analysis. *Test*, 3(1), 5–124. doi: 10.1007/BF02562676
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. doi: 10.1002/9781118619179
- Braeken, J., Tuerlinckx, F., & Boeck, P. (2007, June). Copula Functions for Residual Dependency. *psychometrika*, 72(3), 393–411. doi: 10.1007/s11336-007-9005-4
- 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.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological methods & research*, 21(2), 230–258.
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40(1), 5–32.
- Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative approaches to structural modeling of ordinal data: A monte carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, 4(4), 261–282. Retrieved from <https://doi.org/10.1080/10705519709540077> doi: 10.1080/10705519709540077
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29. doi: 10.1037/1082-989X.1.1.16
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Falkenström, F., Park, S., & McIntosh, C. N. (2021). Using copulas to enable causal inference from non-experimental data: Tutorial and simulation studies. *Psychological Methods*. doi: 10.1037/met0000414
- 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.
- Flora, D. B., & Flake, J. K. (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 49(2), 78.
- 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. doi: 10.1080/10705511.2019.1673168
- Foldnes, N., & Grønneberg, S. (2021). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*. Retrieved from <https://doi.org/10.1037/met0000385> (Online first)
- Foldnes, N., & Grønneberg, S. (2020). 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. (2021). Non-normal data simulation using piecewise linear transforms. *Structural Equation Modeling: A Multidisciplinary Journal*, 0(0), 1–11. doi: 10.1080/10705511.2021.1949323
- 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.
- Galarza, C. E., Kan, R., & Lachos, V. H. (2021). Momtrunc: Moments of folded and doubly truncated multivariate distributions [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MomTrunc> (R package version 5.97)
- Grønneberg, S., Foldnes, N., & Marcoulides, K. M. (2021). covsim: An r package for simulating non-normal data for structural equation models using copulas. *Journal of Statistical Software*. (forthcoming)
- Grønneberg, S., & Moss, J. (2021). Partial identification of latent correlations with polytomous data. *Psychometrika*. (Submitted)
- Grønneberg, S., Moss, J., & Foldnes, N. (2020). Partial identification of latent correlations with binary data. *Psychometrika*, 85(4), 1028–1051. doi: 10.1007/s11336-020-09737-y
- Höfding, W. (1940). Masstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und Instituts für Angewandte Mathematik der Universität Berlin*, 5, 181–233.
- Insua, D. R., & Ruggeri, F. (2012). *Robust bayesian analysis* (Vol. 152). Springer Science & Business Media. doi: 10.1007/978-1-4612-1306-2
- Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, 82(1), 67–85.
- Joe, H. (1997). *Multivariate models and multivariate dependence concepts* (Vol. 73). Chapman & Hall/CRC.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202.
- Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using lisrel*. Technical report, Scientific Software International, Inc., Lincolnwood, IL.
- Jöreskog, K. G., & Sörbom, D. (2015). *Lisrel 9.20 for windows* [computer software]. Skokie, IL: Scientific Software International, Inc..
- Klaassen, C. A., & Wellner, J. A. (1997). Efficient estimation in

- the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 55–77. doi: 10.2307/3318652
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Krupskii, P., & Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120, 85–101. doi: 10.1016/j.jmva.2013.05.001
- Leamer, E. E. (1985). Sensitivity analyses would help. *The American Economic Review*, 75(3), 308–313. Retrieved from <https://www.jstor.org/stable/1814801>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949.
- Mair, P. (2018). *Modern psychometrics with r*. Springer. doi: <https://doi.org/10.1007/978-3-319-93177-7>
- Mair, P. (2020). Mpsychor: Modern psychometrics with r [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MPSychoR> (R package version 0.10-8)
- Mair, P., Hofmann, E., Gruber, K., Hatzinger, R., Zeileis, A., & Hornik, K. (2015). Motivation, values, and work design as drivers of participation in the r open source project for statistical computing. *Proceedings of the National Academy of Sciences*, 112(48), 14788–14792. doi: 10.1073/pnas.1506047112
- 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.
- Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, 71(1), 57–77. doi: 10.1007/s11336-005-0773-4
- Monroe, S. (2018). Contributions to estimation of polychoric correlations. *Multivariate behavioral research*, 53(2), 247–266.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132.
- Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika*, 53(4), 563–577.
- Muthén, B., & Muthén, L. (2012). *Mplus version 7: User's guide*. Los Angeles, CA: Muthén & Muthén.
- Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.
- Nikoloulopoulos, A. K., & Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, 80(1), 126–150. doi: 10.1007/s11336-013-9387-4
- Nilsen, F. A., Bang, H., Boe, O., Martinsen, Ø. L., Lang-Ree, O. C., & Røysamb, E. (2020). The multidimensional self-control scale (mscs): Development and validation. *Psychological Assessment*, 32(11), 1057.
- 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.
- 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.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54(1), 131–151.
- 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.
- Sklar, M. (1959). *Fonctions de repartition a n dimensions et leurs marges*. Université Paris 8.
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393–408.
- Tallis, G. (1962). The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, 18(3), 342–353.
- Vaida, F., & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, 18(4), 797–817.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465–471.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.
- Van Erp, S., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default bayesian structural equation modeling. *Psychological Methods*, 23(2), 363. doi: 10.1037/met0000162
- Xia, Y., & Yang, Y. (2019). Rmsea, cfi, and tli in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior research methods*, 51(1), 409–428.

Online Supplementary Material to Article *Factor analyzing ordinal items requires substantive knowledge*

Steffen Grønneberg & Njål Foldnes

The necessity of a discretization model

We here briefly motivate the importance of the discretization framework for ordinal factor models.

There are three main ways to think about factor analysis with ordinal data. Firstly, we may use a discretization model, as discussed above. Secondly, we may use conditional probability assumptions, such as the considerations leading to the multivariate item response theory (IRT) model as discussed in (Bartholomew, Steele, Galbraith, & Moustaki, 2008, Chapters 8 and 9). While we will not discuss the IRT perspective in this article, we note that IRT models may usually be re-written as discretization models, see Takane and De Leeuw (1987) and Foldnes and Grønneberg (2019a, Appendix).

Thirdly, we may attempt to apply a factor analysis model for continuous data directly to the observations. As Foldnes and Grønneberg (2021) showed mathematically, this may work well under certain assumptions, but we need to be careful about how we quantify the degree of success. In Foldnes and Grønneberg (2021), we start out with a random vector X^* following a continuous factor model, and X^* is then discretized into X . We identify conditions for when estimating X via cont-ML will succeed, meaning that it will reach approximately the same model that X^* follows. This argument takes the discretization process as its starting point, and does assume that the the assumptions leading to a covariance structure are fulfilled for the ordinal observations.

Taking a covariance based factor model for an ordinal X as a starting point is not recommended (Bollen, 1989, Chapter 9). Let us briefly review why this is so. Consider the equations of a factor model the form $X = \mu + \Lambda\xi + \epsilon$. Identifying assumptions include $\text{Cov}(\xi, \epsilon) = 0$. If ξ and/or ϵ take on only a finite number of outcomes, we get a conceptually complex interplay between the support and distribution of ξ , ϵ , the attained values of μ , Λ , and the restriction that $\text{Cov}(\xi, \epsilon) = 0$, which induces problems of such a serious character that this option should not be considered. If in contrast, ξ is to be continuous, which is the traditional perspective, there will be a non-linear dependence between ξ , ϵ : if ξ is continuous, ϵ has to convert the continuous vector $\Lambda\xi$ into the strict categories of X , which is a highly non-linear process. The dependence between ξ and ϵ seems difficult to interpret and motivate, as it must be of such a character that we still have the identifying assumption $\text{Cov}(\xi, \epsilon) = 0$.

Implementing the adjustments in the illustrative case for Study 1 & 2

We here consider some computational details for the implementation of cont-ML-adj and cat-ls-adj for Study 1 & 2.

The cont-ML-adj of Foldnes and Grønneberg (2021) adjusts the observable variables encoding the data using

$$\hat{x}_{k,j} = m(\hat{\tau}_{k,j-1}, \hat{\tau}_{k,j}) \quad (11)$$

where $m(x, y) = E[X_k^* | x \leq X_k^* \leq y]$. In the later section “Mathematical results for implementing cont-ML-adj in the illustration” (p. 3) we provide exact formulas for computing m for the distributions we consider in the illustration, and these are implemented in R code provided as supplementary material.

Using the cat-LS-adj for estimation and inference requires the calculation of Ψ and Ψ' . While calculating Ψ can always be done using numerical integration of the integral given in Proposition 1, our illustration consists of simple transformations of normal variables, and there exist well-established formulas for moments of truncated multivariate normal variables which can be used to calculate Ψ directly. In our implementation given in the supplementary material, we use the R package MomTrunc (Galarza, Kan, & Lachos, 2021), which is based on recursive formulas from Vaida and Liu (2009), to calculate Ψ . Due to the resulting fast evaluation of Ψ , our implementation use numerical differentiation of Ψ to calculate Ψ' . In the case at hand, this is quicker than numerically evaluating the integral definition of Ψ' in Proposition 1.

Proof of Proposition 1

Proof of Proposition 1. The formula for Ψ is the Höfdding formula for covariances (Höfdding, 1940), using that F_1^*, F_2^* are standardized. Since $r \mapsto C_r(u, v)$ is strictly increasing for any $0 < u, v < 1$ (Joe, 1997, Section 5.1), Ψ is strictly increasing. Finally, since all cumulative distribution functions are in probabilities, and hence contained within the interval $[0, 1]$, also $|C_r(F_1^*(x_1), F_2^*(x_2)) - F_1^*(x_1)F_2^*(x_2)| \leq |C_r(F_1^*(x_1), F_2^*(x_2))| + |F_1^*(x_1)F_2^*(x_2)| = |C_r(F_1^*(x_1), F_2^*(x_2))| + |F_1^*(x_1)||F_2^*(x_2)| \leq 1 + 1 \cdot 1 = 2$ by the triangle inequality. We may therefore interchange derivation and integration, and we have $\Psi'(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{d}{dr} C_r(F_1^*(x_1), F_2^*(x_2)) dx_1 dx_2$. Letting $\Phi_{2,r}$ be the cumulative distribution function of a bivariate normal random vector with standardized marginals and correlation r , we recall $\Phi_{2,r}(z_1, z_2) = C_r(\Phi(z_1), \Phi(z_2))$. Therefore, with $z_1 = \Phi^{-1}(F_1^*(x_1))$ and $z_2 = \Phi^{-1}(F_2^*(x_2))$, we have $C_r(F_1^*(x_1), F_2^*(x_2)) = C_r(\Phi(z_1), \Phi(z_2)) = \Phi_{2,r}(z_1, z_2) = \Phi_{2,r}(\Phi^{-1}(F_1^*(x_1)), \Phi^{-1}(F_2^*(x_2)))$. As in Olsson (1979a), we use that $\frac{d}{dr} \Phi_{2,r}(z_1, z_2) = \phi_{2,r}(z_1, z_2)$ from Tallis (1962, p. 344). The result then follows. \square

Standard errors for cat-LS-adj

By the delta method (e.g. Van der Vaart, 2000, Chapter 3), we have that

$$\begin{aligned}\sqrt{n}(\hat{\rho} - \rho_{X^*}) &= \sqrt{n}(\Psi(\hat{\rho}_{\text{NT}}) - \Psi(\rho_{Z^*})) \\ &= \Psi'(\rho_{Z^*}) \sqrt{n}(\hat{\rho}_{\text{NT}} - \rho_{Z^*}) + o_P(1)\end{aligned}\quad (12)$$

where $o_P(1)$ means a quantity converging in probability to zero as $n \rightarrow \infty$. This forms the basis for computing the asymptotic covariance matrix of a vector of adjusted polychoric correlations using a simple formula: Let us gather $p(p-1)/2$ normal theory polychoric correlations in a vector $\hat{\mathcal{P}}_{\text{NT}}$, with respective normal theory limits contained in a vector \mathcal{P}_{NT} . Assume

$$\sqrt{n}(\hat{\mathcal{P}}_{\text{NT}} - \mathcal{P}_{\text{NT}}) \xrightarrow[n \rightarrow \infty]{d} N(0, \Gamma) \quad (13)$$

for some matrix Γ , as derived e.g. in Olsson (1979a). Let the respective adjusted estimators and limits be contained in vectors $\hat{\mathcal{P}}$ and \mathcal{P} . Let Λ be a $q \times q$ matrix, where $q = d(d-1)/2$, containing zeros except for its diagonal, which contain elements of the form $\Psi'(\rho_{Z^*})$ for the Ψ' function of the corresponding coordinate as deduced above. Then Equation (12) combined with Equation (13) implies that

$$\begin{aligned}\sqrt{n}(\hat{\mathcal{P}} - \mathcal{P}) &= \Lambda \sqrt{n}(\hat{\mathcal{P}}_{\text{NT}} - \mathcal{P}_{\text{NT}}) + o_P(1) \\ &\xrightarrow[n \rightarrow \infty]{d} N(0, \Lambda \Gamma \Lambda').\end{aligned}$$

The asymptotic covariance matrix of the adjusted polychoric estimators is therefore given by $\Lambda \Gamma \Lambda'$. Alternatively, statistical inference can be done using the parametric bootstrap (Efron & Tibshirani, 1994).

What if the copula is non-normal?

While Assumption 1.1 is a weak assumption, Assumption 1.2 is a strong assumption, which will likely often not be fulfilled. An investigation of non-normal copulas in the context of ordinal factor models is found in Foldnes and Grønneberg (2021), and we consider the subject as partly outside the scope of the present article. Surprisingly, the requirement of knowing the copula is less absolute than knowing the marginals: Foldnes and Grønneberg (2021, Lem. 1) considered a modification of the normal theory polychoric correlation which takes into account the marginal information in a slightly different way than the adjusted polychoric correlation considered in the present article. For that alternative polychoric correlation, as $K \rightarrow \infty$, we consistently estimate the Pearson correlation of the response variables as long as the marginals are correctly specified, but the copula of the response variables need not be normal nor even known.

For completeness, we here briefly discuss what happens if we know that $(X_1^*, X_2^*)'$ has a non-normal copula that fulfills

some regularity conditions specified shortly. The full development of this problem is considered outside the scope of the present paper.

Assumption 1.2 can be extended as follows

Assumption 2. 2.' We assume that $(X_1^*, X_2^*)'$ has a copula with cumulative distribution function C_θ for $\theta \in \Theta$ where Θ is an interval of real numbers with possibly infinite length. Here, C_θ is such that for all $u, v \in (0, 1)$ the function $\theta \mapsto C_\theta(u, v)$ is strictly increasing.

Lemma 1. Under 2.2', θ is identified.

Proof. The θ parameter is identified by just dichotomous knowledge, using the argument just above Theorem 1 in Grønneberg et al. (2020). Since this dichotomous information is derivable from the full distribution of X , the parameter is identified. \square

The parameter θ may now be estimated by standard ML, or some variant of it as in Jin and Yang-Wallentin (2017). From the estimated copula parameters, and the known marginals F_1^*, F_2^* , we may compute the Pearson correlation of the response distribution. The result of this calculation is our proposed estimator for response correlations. Inference theory then follows by standard asymptotics for ML estimators, or inference theory that takes into account e.g. two step estimation such as in Jin and Yang-Wallentin (2017). Since we have a fully specified parametric model, inference may also follow from the parametric bootstrap (Efron & Tibshirani, 1994). These correlation estimates, and their asymptotic covariance matrix, are then used in cat-LS, replacing the normal theory polychoric correlations and their asymptotic covariance matrix.

A review of cat-LS-thr, and a comparison to cat-LS-adj

Foldnes and Grønneberg (2021) suggested an adjustment to normal theory polychoric correlations, which we may call threshold adjusted polychorics, took response marginals $F_1^*, F_2^*, \dots, F_p^*$ as input, and provided an estimate of the response correlation matrix as output. As $K \rightarrow \infty$, this estimate is consistent as long as the response marginals are correctly specified. This property is shared with cont-ML-adj, and holds irrespective of the true underlying response copula.

When replacing standard polychorics in cat-LS estimation with threshold adjusted polychorics, we get an estimator which we call cat-LS-thr.

In normal theory polychorics, the thresholds are estimated using

$$\hat{\tau}_{\text{NT},k,j} = \Phi^{-1}(\hat{P}(X_k \leq j))$$

where $\hat{P}(X_k \leq j)$ is an empirical probability based on a given sample, and Φ is the cumulative distribution function of the

standard normal distribution. In threshold adjusted polychorics, the thresholds are instead estimated based on the relation $\tau_{k,j} = F_k^{*-1}(P(X_k \leq j))$, and set to

$$\hat{\tau}_{k,j} = F_k^{*-1}(\hat{P}(X_k \leq j)),$$

where $\hat{P}(X_k \leq j)$ is the empirical probability of this event. The estimator $\hat{\tau}_{k,j}$ will be consistent as long as the marginals are (F_k^*) . Code to implement the threshold adjusted polychorics is given in the online supplementary material of Foldnes and Grønneberg (2021). Standard errors are presently only available for normal marginals, in which case the cat-LS-thr is the standard cat-LS. Standard errors may be computed using bootstrap methods, and the derivation of analytical formulas is left for future research.

The threshold adjusted polychorics of Foldnes and Grønneberg (2021) is then the normal theory polychorics, but instead of using $(\hat{\tau}_{NT,k,j})$ as threshold estimates, $(\hat{\tau}_{k,j})$ is rather used. Of course, if response marginals are fixed to standard normal, the standard normal theory polychoric estimates reappear.

We now compare cat-LS-thr with the cat-LS-adj suggested in the present paper. Without loss of generality, we assume $p = 2$. The only difference between these methods is that cat-LS-thr use threshold adjusted polychoric correlations, say $\hat{\rho}_{thr}(F_1^*, F_2^*)$, and cat-LS-adj use adjusted polychorics as described above, say $\hat{\rho}_{adj}(F_1^*, F_2^*)$. The population limits of these estimators are denoted by $\rho_{thr,K}(F_1^*, F_2^*)$ and $\rho_{adj,K}(F_1^*, F_2^*)$ respectively, where we introduce a subscript K to indicate the dependence on the number of thresholds. We wish to estimate

$$\rho_{X^*} = \text{Cov}(X_1^*, X_2^*).$$

As shown in Foldnes and Grønneberg (2021, Lem. 1), we have

$$\lim_{K \rightarrow \infty} \rho_{thr,K}(F_1^*, F_2^*) = \rho_{X^*}$$

if the marginals of (X_1^*, X_2^*) indeed are F_1^*, F_2^* . When the marginals are misspecified, expressions for the limit of the threshold adjusted polychoric was also derived in Foldnes and Grønneberg (2021, see the discussion right after Lem. 1), which we may use to see that

$$\lim_{K \rightarrow \infty} \rho_{thr,K}(\Phi, \Phi) = \text{Cov}(\Phi^{-1}(F_1^*(X_1^*)), \Phi^{-1}(F_2^*(X_2^*)))$$

where F_1^*, F_2^* are the true marginals. We therefore have that

$$\begin{aligned} \hat{\rho}_{adj}(F_1^*, F_2^*) &= \Psi(\hat{\rho}_{thr}(\Phi, \Phi)) \\ &\xrightarrow{n \rightarrow \infty} \Psi(\rho_{thr}(\Phi, \Phi)) \\ &\xrightarrow{K \rightarrow \infty} \Psi(\text{Cov}(\Phi^{-1}(F_1^*(X_1^*)), \Phi^{-1}(F_2^*(X_2^*)))) \end{aligned}$$

Now unless X_1^*, X_2^* happens to have a normal copula, this limit is not equal to ρ_{X^*} , showing that cat-LS-adj and the

briefly mentioned cat-LS-ext which extends cont-ML-adj to non-normal copulas, are both inconsistent as $K \rightarrow \infty$ if the copula is misspecified. In contrast, cat-LS-thr is consistent as $K \rightarrow \infty$, as long as the marginals are correctly specified.

Mathematical results for implementing cont-ML-adj in study 1 & 2

The following results should already be available in the literature. Since we do not know a reference that derive all of these results and state them in a simple form, and since it has some value to illustrate the process required to derive the conditional expectation used in the cont-ML-adj, we provide complete calculations of the following results for the reader's convenience. We do not aim at generality, but instead aim at providing enough explanation to reproduce our illustrations.

Let

$$Y = (\alpha_0 + \alpha_1 Z)I\{Z < 0\} + (\beta_0 + \beta_1 Z)I\{Z \geq 0\} \quad (14)$$

where we usually assume $Z \sim N(0, 1)$.

Let us first consider how to standardize these distributions. Suppose given Y in the above form. Then also $\tilde{Y} = (Y - E Y)/\text{sd}(Y)$ is of the same algebraic form, but with new coefficients $\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\beta}_0, \tilde{\beta}_1$. To see this, notice firstly that

$$aY = (a\alpha_0 + a\alpha_1 Z)I\{Z < 0\} + (a\beta_0 + a\beta_1 Z)I\{Z \geq 0\},$$

and secondly that

$$\begin{aligned} Y + a &= Y + a \cdot 1 = Y + a(I\{Z < 0\} + I\{Z \geq 0\}) \\ &= (a + \alpha_0 + \alpha_1 Z)I\{Z < 0\} + (a + \beta_0 + \beta_1 Z)I\{Z \geq 0\}. \end{aligned}$$

To standardize Y , we therefore only need to compute its expectation and variance, and then use updated coefficients $\tilde{\alpha}_0 = (\alpha_0 - E Y)/\text{sd}(Y)$, $\tilde{\beta}_0 = (\beta_0 - E Y)/\text{sd}(Y)$ and $\tilde{\alpha}_1 = \alpha_1/\text{sd}(Y)$, $\tilde{\beta}_1 = \beta_1/\text{sd}(Y)$. Code to compute the expectation and variance of Y when $Z \sim N(0, 1)$ is given in the online supplementary material.

We here consider the conditional distribution and expectation of Y given knowledge of Y being contained in an interval $[x, y]$. We only consider the case when α_1, β_1 have the same sign. The general case follows by the same arguments, but will not be useful for our illustrations in the present article.

The following results assumes that α_1, β_1 are positive. When they are both negative, we may use the presented results as follows. Notice that

$$\tilde{Y} := -Y = (-\alpha_0 + (-\alpha_1)Z)I\{Z < 0\} + (-\beta_0 + (-\beta_1)Z)I\{Z \geq 0\}$$

is of the form

$$\tilde{Y} = (\tilde{\alpha}_0 + \tilde{\alpha}_1 Z)I\{Z < 0\} + (\tilde{\beta}_0 + \tilde{\beta}_1 Z)I\{Z \geq 0\}$$

with $\tilde{\alpha}_1, \tilde{\beta}_1$ both positive.

Let $P(Y \leq y) = F_Y(y)$ and $f_Y(y) = (d/dx)F_Y(y)$. We have $P(\tilde{Y} \leq y) = P(-Y \leq y) = P(Y \geq -y) = 1 - P(Y \leq -y) =$

$1 - F_Y(-y)$, with a density given by $f_{\tilde{Y}}(y) = (d/dx)P(\tilde{Y} \leq y) = f_Y(-y)$.

Finally, consider $F_{\tilde{Y}}(y) = x$. Then $1 - F_Y(-y) = x$, so that $F_Y(-y) = 1 - x$ and so $-y = F_Y^{-1}(1 - x)$, which means $F_{\tilde{Y}}^{-1}(x) = -F_Y^{-1}(1 - x)$.

For the conditional mean, notice that since $x \leq Y \leq y \iff -y \leq -Y \leq -x \iff -y \leq \tilde{Y} \leq -x$, we have

$$E[Y|x \leq Y \leq y] = -E[-Y|-y \leq \tilde{Y} \leq -x] = E[\tilde{Y}|-y \leq \tilde{Y} \leq -x].$$

We may therefore reduce the case where α_1, β_1 are both negative to the case where α_1, β_1 are both positive, using the above equations.

The following results allow $x = -\infty, y = \infty$. We derive the density to draw exact plots of the densities in the article.

Lemma 2. Suppose Y follows Equation (14), with $Z \sim N(0, 1)$ and $\alpha_1, \beta_1 > 0$. Then $F_Y(y) = P(Y \leq y) = \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + I\{y \geq \beta_0\}[\Phi(\beta_1^{-1}(y - \beta_0))] + \Phi(0)[I\{y \geq \alpha_0\} - I\{y \geq \beta_0\}]$, and $f(y) = F'_Y(y) = \alpha_1^{-1}\phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + \beta_1^{-1}\phi(\beta_1^{-1}(y - \beta_0))I\{y \geq \beta_0\}$. If also $\alpha_0 = \beta_0$, then $F_Y^{-1}(x) = (\beta_1 I\{1/2 \leq x \leq 1\} + \alpha_1 I\{0 \leq x < 1/2\})\Phi^{-1}(x) + \alpha_0$.

Proof. Using the notation $x^- = \min(x, 0)$ we get

$$\begin{aligned} P(Y \leq y) &\stackrel{(a)}{=} P(Y \leq y, Z < 0) + P(Y \leq y, Z \geq 0) \\ &= P(\alpha_0 + \alpha_1 Z \leq y, Z < 0) + P(\beta_0 + \beta_1 Z \leq y, Z \geq 0) \\ &\stackrel{(b)}{=} P(Z \leq \alpha_1^{-1}(y - \alpha_0), Z < 0) + P(Z \leq \beta_1^{-1}(y - \beta_0), Z \geq 0) \\ &\stackrel{(c)}{=} P(Z < [\alpha_1^{-1}(y - \alpha_0)]^-) \\ &\quad + I\{\beta_1^{-1}(y - \beta_0) \geq 0\}P(0 \leq Z \leq \beta_1^{-1}(y - \beta_0)) \\ &= \Phi([\alpha_1^{-1}(y - \alpha_0)]^-) \\ &\quad + I\{\beta_1^{-1}(y - \beta_0) \geq 0\}[\Phi(\beta_1^{-1}(y - \beta_0)) - \Phi(0)] \\ &= \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + \Phi(0)I\{y \geq \alpha_0\} \\ &\quad + I\{y \geq \beta_0\}[\Phi(\beta_1^{-1}(y - \beta_0)) - \Phi(0)] \\ &= \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + I\{y \geq \beta_0\}[\Phi(\beta_1^{-1}(y - \beta_0))] \\ &\quad + \Phi(0)[I\{y \geq \alpha_0\} - I\{y \geq \beta_0\}]. \end{aligned}$$

(a) For disjoint A, B we have $P(C) = P(C \cap A) + P(C \cap B)$. (b) We assume $\alpha_1, \beta_1 > 0$. (c) Comma is short hand for intersection, so $\{Z \leq a_1, Z < a_2\} = \{Z < \min(a_1, a_2)\}$. Also, if $\beta_1^{-1}(y - \beta_0) < 0$, we have $P(Z \leq \beta_1^{-1}(y - \beta_0), Z \geq 0) = 0$, but otherwise, we have $P(Z \leq \beta_1^{-1}(y - \beta_0), Z \geq 0) = P(0 \leq Z \leq \beta_1^{-1}(y - \beta_0))$. (d) Since $\alpha_1, \beta_1 > 0$, we have $I\{\beta_1^{-1}(y - \beta_0) \geq 0\} = I\{y \geq \beta_0\}$. We also have $\alpha_1^{-1}(y - \alpha_0)^- = \alpha_1^{-1}(y - \alpha_0)I\{y < \alpha_0\}$ and therefore $\Phi([\alpha_1^{-1}(y - \alpha_0)]^-) = \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + \Phi(0)I\{y \geq \alpha_0\}$.

The density of Y is therefore

$$\begin{aligned} f(y) &= (d/dy)P(Y \leq y) \\ &= \alpha_1^{-1}\phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} \\ &\quad + \beta_1^{-1}\phi(\beta_1^{-1}(y - \beta_0))I\{y \geq \beta_0\}. \end{aligned}$$

We now show the second statement of the lemma, and additionally assume $\alpha_0 = \beta_0$. We then have

$$\begin{aligned} F_Y(y) &= \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + I\{y \geq \alpha_0\}[\Phi(\beta_1^{-1}(y - \alpha_0))] \\ &\quad + \Phi(0)[I\{y \geq \alpha_0\} - I\{y \geq \beta_0\}] \\ &= \Phi(\alpha_1^{-1}(y - \alpha_0))I\{y < \alpha_0\} + [\Phi(\beta_1^{-1}(y - \alpha_0))]I\{y \geq \alpha_0\}. \end{aligned}$$

Let $F_Y(y) = x$. If $0 \leq x < 1/2 = \Phi(0)$, then we must have $y < \alpha_0$, and so $\alpha_1^{-1}(y - \alpha_0) = x$, which means $\alpha_1^{-1}(y - \alpha_0) = \Phi^{-1}(x)$ and so $y = \alpha_1\Phi^{-1}(x) + \alpha_0$. If $x \geq 1/2$, we have $y \geq \alpha_0$, and so by the same argument, $y = \beta_1\Phi^{-1}(x) + \alpha_0$. Therefore, we have the claimed $F_Y^{-1}(x) = (\beta_1 I\{1/2 \leq x \leq 1\} + \alpha_1 I\{0 \leq x < 1/2\})\Phi^{-1}(x) + \alpha_0$. \square

In the following proposition, we may use that when $Z \sim N(0, 1)$, we have

$$E(Z|x \leq Z \leq y) = [\phi(x) - \phi(y)]/[\Phi(y) - \Phi(x)], \quad (15)$$

and that $P(x \leq Y \leq y)$ can be computed using the cumulative distribution function identified in Lemma 2.

Proposition 2. Let Y be defined as in Equation (14), where Z has finite mean and is a continuous distribution. Suppose α_1, β_1 are both positive. We then have that

$$\begin{aligned} m(x, y) &= E[Y|x \leq Y \leq y] \\ &= \frac{P([\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-)}{P(x \leq Y \leq y)} \\ &\quad \cdot [\alpha_0 + \alpha_1 E[Z|\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-] \\ &\quad + \frac{P([\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+)}{P(x \leq Y \leq y)} \\ &\quad \cdot [\beta_0 + \beta_1 E[Z|\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+] \end{aligned}$$

Proof. The assumption that Z has a continuous distribution is only used to not have to keep track of sharp inequalities.

We assume α_1, β_1 are both positive. We have that

$$\begin{aligned} m(x, y) &= E[Y|x \leq Y \leq y] \\ &= P(x \leq Y \leq y)^{-1} E[YZI\{x \leq Y \leq y\}]. \end{aligned}$$

We also have

$$\begin{aligned} E[YZI\{x \leq Y \leq y\}] &= E(\alpha_0 + \alpha_1 Z)I\{Z < 0, x \leq Y \leq y\} \\ &\quad + E(\beta_0 + \beta_1 Z)I\{Z \geq 0, x \leq Y \leq y\} \\ &\stackrel{(a)}{=} E(\alpha_0 + \alpha_1 Z)I\{[\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-\} \\ &\quad + E(\beta_0 + \beta_1 Z)I\{[\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+\} \\ &\stackrel{(b)}{=} \alpha_0 P([\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-) \\ &\quad + \alpha_1 P([\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-) \\ &\quad \cdot E[Z | [\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-] \\ &\quad + \beta_0 P([\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+) \\ &\quad + \beta_1 P([\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+) \\ &\quad \cdot E[Z | [\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+] \\ &= P([\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-) \\ &\quad \cdot [\alpha_0 + \alpha_1 E[Z | [\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-]] \\ &\quad + P([\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+) \\ &\quad \cdot [\beta_0 + \beta_1 E[Z | [\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+]] \end{aligned}$$

(a) Since $Y = \alpha_0 + \alpha_1 Z$ if $Z < 0$, we have $I\{Z < 0, x \leq Y \leq y\} = I\{Z < 0, x \leq \alpha_0 + \alpha_1 Z \leq y\} = I\{Z < 0, \alpha_1^{-1}(x - \alpha_0) \leq Z \leq \alpha_1^{-1}(y - \alpha_0)\} = I\{[\alpha_1^{-1}(x - \alpha_0)]^- \leq Z \leq [\alpha_1^{-1}(y - \alpha_0)]^-\}$. Similarly, $I\{Z \geq 0, x \leq Y \leq y\} = I\{[\beta_1^{-1}(x - \beta_0)]^+ \leq Z \leq [\beta_1^{-1}(y - \beta_0)]^+\}$. (b) We use again that for any random variable Y with finite mean, we have $E[Y|x \leq Y \leq y] = P(x \leq Y \leq y)^{-1} E[YZI\{x \leq Y \leq y\}]$. We also use that $E I\{A\} = P(A)$. \square

Monte Carlo results**Study 3**

Margins	Polychoric			Adjusted polychoric		
	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	1.0	0.7	0.7	-0.6	-0.6	0.1
$\tilde{\Gamma}_1\Gamma_1$	2.0	3.0	4.0	-0.7	-0.2	0.0
$\tilde{\Gamma}_1\tilde{\Gamma}_1$	2.0	1.0	0.9	0.4	-0.3	0.2
$\Gamma_2\Gamma_1$	3.6	3.5	2.1	-0.8	-0.1	-0.2
$\Gamma_2\tilde{\Gamma}_1$	5.6	7.3	9.3	-1.1	-0.3	-0.0
$\Gamma_2\Gamma_2$	6.8	4.5	2.4	0.2	-0.5	0.0
$\tilde{\Gamma}_2\Gamma_1$	6.4	7.1	9.4	-0.3	-0.6	0.1
$\tilde{\Gamma}_2\tilde{\Gamma}_1$	5.1	2.7	2.2	0.7	-0.9	-0.1
$\tilde{\Gamma}_2\Gamma_2$	10.8	14.1	17.4	-0.6	0.3	-0.2
$\tilde{\Gamma}_2\tilde{\Gamma}_2$	6.7	4.9	2.2	0.1	-0.0	-0.2
$\Gamma_3\Gamma_1$	7.2	5.9	4.3	0.4	0.1	0.2
$\Gamma_3\tilde{\Gamma}_1$	9.4	10.8	13.1	-0.6	-0.5	-0.4
$\Gamma_3\Gamma_2$	10.0	7.1	3.6	1.2	0.5	0.3
$\Gamma_3\tilde{\Gamma}_2$	14.7	18.0	24.3	-0.6	-0.4	0.0
$\Gamma_3\Gamma_3$	12.7	8.6	3.7	1.8	0.7	-0.0
$\tilde{\Gamma}_3\Gamma_1$	9.8	10.9	13.2	-0.2	-0.4	-0.4
$\tilde{\Gamma}_3\tilde{\Gamma}_1$	6.8	5.8	4.3	-0.1	0.0	0.2
$\tilde{\Gamma}_3\Gamma_2$	15.1	18.4	24.2	-0.2	-0.1	-0.1
$\tilde{\Gamma}_3\tilde{\Gamma}_2$	9.9	6.2	3.7	1.0	-0.4	0.4
$\tilde{\Gamma}_3\Gamma_3$	20.0	24.3	33.0	0.1	0.0	-0.1
$\tilde{\Gamma}_3\tilde{\Gamma}_3$	11.2	8.8	3.9	0.3	0.9	0.1
$N(0,1)\Gamma_1$	0.5	0.6	1.3	-0.6	-0.5	0.2
$N(0,1)\tilde{\Gamma}_1$	-0.3	1.1	1.3	-1.2	-0.0	0.2
$N(0,1)\Gamma_2$	3.2	4.3	4.6	-1.1	-0.1	0.2
$N(0,1)\tilde{\Gamma}_2$	4.1	4.0	4.6	-0.2	-0.4	0.2
$N(0,1)\Gamma_3$	6.1	7.0	7.1	-1.1	-0.3	-0.2
$N(0,1)\tilde{\Gamma}_3$	7.6	7.4	7.2	0.3	0.2	-0.0
$N(0,1)N(0,1)$	0.1	-0.4	-0.1	0.1	-0.4	-0.1

Table 6

Relative bias for the polychoric estimator and its adjusted version. The results are aggregated over sample sizes $n = 100, 300, 1000$.

Margins	Polychoric			Adjusted polychoric		
	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	92.3	93.3	91.8	92.5	93.1	92.2
$\tilde{\Gamma}_1\Gamma_1$	91.1	92.3	86.5	91.3	92.4	92.5
$\tilde{\Gamma}_1\tilde{\Gamma}_1$	91.9	92.8	89.9	92.1	92.6	90.1
$\Gamma_2\Gamma_1$	91.4	92.3	89.2	91.8	92.7	92.4
$\Gamma_2\tilde{\Gamma}_1$	92.9	88.6	63.1	93.5	91.7	91.8
$\Gamma_2\Gamma_2$	91.1	91.2	89.0	91.8	92.7	93.3
$\tilde{\Gamma}_2\Gamma_1$	91.6	89.6	62.0	91.8	91.0	92.5
$\tilde{\Gamma}_2\tilde{\Gamma}_1$	91.9	92.9	89.7	91.7	91.5	92.3
$\tilde{\Gamma}_2\Gamma_2$	90.9	82.5	16.6	93.0	91.4	93.6
$\tilde{\Gamma}_2\tilde{\Gamma}_2$	90.1	91.0	90.3	90.9	92.8	93.5
$\Gamma_3\Gamma_1$	91.2	89.7	84.3	91.7	90.8	92.9
$\Gamma_3\tilde{\Gamma}_1$	90.7	85.1	41.6	91.2	91.7	93.7
$\Gamma_3\Gamma_2$	91.6	89.1	88.1	92.4	92.5	92.2
$\Gamma_3\tilde{\Gamma}_2$	88.9	79.4	0.3	91.4	93.0	93.1
$\Gamma_3\Gamma_3$	90.3	88.3	85.5	92.3	91.3	92.7
$\tilde{\Gamma}_3\Gamma_1$	91.0	86.4	40.3	91.8	92.2	92.7
$\tilde{\Gamma}_3\tilde{\Gamma}_1$	92.7	88.9	83.0	93.0	92.2	90.4
$\tilde{\Gamma}_3\Gamma_2$	89.9	77.1	0.9	91.5	94.0	93.9
$\tilde{\Gamma}_3\tilde{\Gamma}_2$	91.2	91.9	85.8	91.9	92.8	92.3
$\tilde{\Gamma}_3\Gamma_3$	87.7	69.2	0.0	91.4	92.7	93.8
$\tilde{\Gamma}_3\tilde{\Gamma}_3$	90.5	90.0	86.9	92.4	93.9	93.0
$N(0, 1)\Gamma_1$	91.9	92.5	91.9	91.7	92.5	92.8
$N(0, 1)\tilde{\Gamma}_1$	91.4	91.1	90.2	91.6	91.2	91.2
$N(0, 1)\Gamma_2$	92.9	90.1	82.8	92.4	91.1	92.4
$N(0, 1)\tilde{\Gamma}_2$	91.3	91.0	84.7	92.1	92.6	93.0
$N(0, 1)\Gamma_3$	93.4	88.5	74.2	93.8	91.3	92.9
$N(0, 1)\tilde{\Gamma}_3$	93.1	90.0	73.5	93.7	93.1	92.2
$N(0, 1)N(0, 1)$	90.7	93.4	93.5	90.7	93.4	93.5
Mean	91.3	88.5	69.8	92.0	92.3	92.6

Table 7

Study 3: Coverage rates at the 95% level of confidence, sample size $n = 100$.

Margins	Polychoric			Adjusted polychoric		
	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	92.2	94.1	93.6	92.3	94.2	94.6
$\tilde{\Gamma}_1\Gamma_1$	94.0	92.4	80.4	93.5	93.5	95.0
$\tilde{\Gamma}_1\tilde{\Gamma}_1$	94.8	93.5	91.7	95.2	93.8	92.9
$\Gamma_2\Gamma_1$	94.0	91.8	87.4	94.2	93.4	92.7
$\Gamma_2\tilde{\Gamma}_1$	94.4	88.1	32.9	95.3	93.4	93.6
$\Gamma_2\Gamma_2$	92.5	92.3	91.1	93.7	94.9	94.9
$\tilde{\Gamma}_2\Gamma_1$	93.5	90.4	31.2	94.9	95.5	92.5
$\tilde{\Gamma}_2\tilde{\Gamma}_1$	94.6	91.5	88.8	95.3	93.6	94.4
$\tilde{\Gamma}_2\Gamma_2$	91.1	76.4	0.4	93.3	93.2	95.0
$\tilde{\Gamma}_2\tilde{\Gamma}_2$	91.7	91.5	88.4	92.1	94.1	94.8
$\Gamma_3\Gamma_1$	94.5	91.2	77.9	94.3	94.4	93.1
$\Gamma_3\tilde{\Gamma}_1$	92.4	81.4	6.0	93.8	93.6	95.5
$\Gamma_3\Gamma_2$	92.3	88.9	81.0	94.2	94.5	93.0
$\Gamma_3\tilde{\Gamma}_2$	89.1	63.5	0.0	94.2	95.1	95.5
$\Gamma_3\Gamma_3$	93.2	86.5	81.1	95.1	94.3	93.2
$\tilde{\Gamma}_3\Gamma_1$	91.6	82.6	7.0	94.3	93.9	94.3
$\tilde{\Gamma}_3\tilde{\Gamma}_1$	93.3	91.5	79.8	94.2	93.9	94.4
$\tilde{\Gamma}_3\Gamma_2$	90.2	62.7	0.0	94.0	92.2	93.9
$\tilde{\Gamma}_3\tilde{\Gamma}_2$	92.3	89.8	85.5	94.8	93.0	95.3
$\tilde{\Gamma}_3\Gamma_3$	88.6	46.4	0.0	94.0	94.5	94.8
$\tilde{\Gamma}_3\tilde{\Gamma}_3$	91.8	87.2	80.5	94.2	95.2	93.2
$N(0, 1)\Gamma_1$	94.0	94.2	91.0	94.2	94.8	93.4
$N(0, 1)\tilde{\Gamma}_1$	94.5	94.5	91.8	94.1	94.7	93.5
$N(0, 1)\Gamma_2$	93.7	90.5	75.9	93.5	93.7	94.7
$N(0, 1)\tilde{\Gamma}_2$	92.8	93.8	75.5	94.1	94.9	93.8
$N(0, 1)\Gamma_3$	93.4	86.5	54.8	94.0	93.0	94.2
$N(0, 1)\tilde{\Gamma}_3$	92.4	88.4	53.1	94.5	94.6	95.2
$N(0, 1)N(0, 1)$	93.6	94.7	92.9	93.6	94.7	92.9
Mean	92.7	86.3	61.4	94.1	94.1	94.1

Table 8

Study 3: Coverage rates at the 95% level of confidence, sample size $n = 300$.