Online Supplementary Material

Steffen Grønneberg and Njål Foldnes
BI Norwegian Business School

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 (2019, 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 $\mathrm{Cov}(\xi, \epsilon) = 0$. 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 $\mathrm{Cov}(\xi, \epsilon) = 0$, 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 $\mathrm{Cov}(\xi, \epsilon) = 0$.

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

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

where $m(x, y) = \mathrm{E}[X_k^* | x \leq X_k^* \leq y]$. 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 Höffding formula for covariances (Höffding, 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 $|C_r\left(F_1^*(x_1), F_2^*(x_2)\right) - F_1^*(x_1)F_2^*(x_2)| \leq |C_r\left(F_1^*(x_1), F_2^*(x_2)\right)| + |F_1^*(x_1)F_2^*(x_2)| = |C_r\left(F_1^*(x_1), F_2^*(x_2)\right)| + |F_1^*(x_1)||F_2^*(x_2)| \leq 1 + 1 \cdot 1 = 2$ by the triangle inequality. We may therefore interchange derivation and integration, and we have $\Psi'(r) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{d}{dr} C_r\left(F_1^*(x_1), F_2(x_2^*)\right) dx_1 dx_2$. 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 $C_r\left(F_1^*(x_1), F_2^*(x_2)\right) = C_r(\Phi(z_1), \Phi(z_2)) = \Phi_{2,r}(z_1, z_2) = \Phi_{2,r}(\Phi^{-1}(F_1^*(x_1)), \Phi^{-1}(F_2^*(x_2)))$. As in Olsson (1979), 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. □

Standard errors for cat-LS-adj

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

$$\begin{aligned}\sqrt{n}(\hat{\rho} - \rho_{X^*}) &= \sqrt{n}(\Psi(\hat{\rho}_{\mathrm{NT}}) - \Psi(\rho_{Z^*})) \\ &= \Psi'(\rho_{Z^*})\sqrt{n}(\hat{\rho}_{\mathrm{NT}} - \rho_{Z^*}) + o_P(1) \qquad (2)\end{aligned}$$

where $o_P(1)$ means a quantity converging in probability to zero as $n \to \infty$. 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 $\hat{\mathcal{P}}_{\mathrm{NT}}$, with respective normal theory limits contained in a vector $\mathcal{P}_{\mathrm{NT}}$. Assume

$$\sqrt{n}[\hat{\mathcal{P}}_{\mathrm{NT}} - \mathcal{P}_{\mathrm{NT}}] \xrightarrow[n\to\infty]{d} N(0, \Gamma) \qquad (3)$$

for some matrix Γ, as derived e.g. in Olsson (1979). 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 (2) combined with Equation (3) implies that

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

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 $K \to \infty$, 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 1. *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, Moss, and Foldnes (2020). Since this dichotomous information is derivable from the full distribution of X, the parameter is identified. □

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 $F_1^*, F_2^*, \ldots, F_p^*$ as input, and provided an estimate of the response correlation matrix as output. As $K \to \infty$, 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

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

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 $(\hat{\tau}_{\mathrm{NT},k,j})$ 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 $\hat{\rho}_{\mathrm{thr}}(F_1^*, F_2^*)$, and cat-LS-adj use adjusted polychorics as described above, say $\hat{\rho}_{\mathrm{adj}}(F_1^*, F_2^*)$. The population limits of these estimators are denoted by $\rho_{\mathrm{thr},K}(F_1^*, F_2^*)$ and $\rho_{\mathrm{adj},K}(F_1^*, F_2^*)$ respectively, where we introduce a subscript K to indicate the dependence on the number of thresholds. We wish to estimate

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

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

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

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

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

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

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

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 $K \to \infty$ if the copula is misspecified. In contrast, cat-LS-thr is consistent as $K \to \infty$, 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

$$Y = (\alpha_0 + \alpha_1 Z)I\{Z < 0\} + (\beta_0 + \beta_1 Z)I\{Z \geq 0\} \tag{4}$$

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 $\tilde{Y} = (Y - \mathrm{E}\, Y)/\mathrm{sd}(Y)$ 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 $\tilde{\alpha}_0 = (\alpha_0 - \mathrm{E}\, Y)/\mathrm{sd}(Y), \tilde{\beta}_0 = (\alpha_0 - \mathrm{E}\, Y)/\mathrm{sd}(Y)$ and $\tilde{\alpha}_1 = \alpha_1/\mathrm{sd}(Y), \tilde{\beta}_1 = \beta_1/\mathrm{sd}(Y)$. 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

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

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* (4)*, with $Z \sim N(0,1)$ and $\alpha_1, \beta_1 > 0$. Then $F_Y(y) = P(Y \leq y) = \Phi\left(\alpha_1^{-1}(y - \alpha_0)\right) I\{y < \alpha_0\} + I\{y \geq \beta_0\}\left[\Phi(\beta_1^{-1}(y - \beta_0))\right] + \Phi(0)\left[I\{y \geq \alpha_0\} - I\{y \geq \beta_0\}\right]$, and $f(y) = F_Y'(y) = \alpha_1^{-1}\phi\left(\alpha_1^{-1}(y - \alpha_0)\right) I\{(y < \alpha_0\} + \beta_1^{-1}\phi(\beta_1^{-1}(y - \beta_0))I\{y \geq \beta_0\}$. 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

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

(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 $\Phi\left([\alpha_1^{-1}(y - \alpha_0)]^-\right) = \Phi\left(\alpha_1^{-1}(y - \alpha_0)]I\{y < \alpha_0\}\right) = \Phi\left(\alpha_1^{-1}(y - \alpha_0)\right) I\{y < \alpha_0\} + \Phi(0)I\{y \geq \alpha_0\}$.

The density of Y is therefore

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

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

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

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 $F_Y^{-1}(x) = (\beta_1 I\{1/2 \leq x \leq 1\} + \alpha_1 I\{0 \leq x < 1/2\})\,\Phi^{-1}(x) + \alpha_0$. □

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

$$\mathrm{E}\left(Z\middle|x \leq Z \leq y\right) = [\phi(x) - \phi(y)]/[\Phi(y) - \Phi(x)], \tag{5}$$

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

Proposition 1. *Let Y be defined as in Equation* (4)*, where Z has finite mean and is a continuous distribution. Suppose α_1, β_1 are both positive. We then have that*

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

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

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

We also have

$$\begin{aligned}
&\mathrm{E}[YI\{x \le Y \le y\}] = \mathrm{E}(\alpha_0 + \alpha_1 Z)I\{Z < 0, x \le Y \le y\} \\
&\quad + \mathrm{E}(\beta_0 + \beta_1 Z)I\{Z \ge 0, x \le Y \le y\} \\
&\overset{(a)}{=} \mathrm{E}(\alpha_0 + \alpha_1 Z)I\{[\alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-\} \\
&\quad + \mathrm{E}(\beta_0 + \beta_1 Z)I\{[\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+\} \\
&\overset{(b)}{=} \alpha_0 P([\alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-) \\
&\quad + \alpha_1 P([\alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-) \\
&\quad \cdot \mathrm{E}[Z \big| [\alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-] \\
&\quad + \beta_0 P([\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+) \\
&\quad + \beta_1 P([\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+) \\
&\quad \cdot \mathrm{E}[Z \big| [\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+] \\
&= P([\alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-) \\
&\quad \cdot \Big[\alpha_0 + \alpha_1\,\mathrm{E}[Z \big| \alpha_1^{-1}(x-\alpha_0)]^- \le Z \le [\alpha_1^{-1}(y-\alpha_0)]^-]\Big] \\
&\quad + P([\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+) \\
&\quad \cdot \Big[\beta_0 + \beta_1\,\mathrm{E}[Z \big| [\beta_1^{-1}(x-\beta_0)]^+ \le Z \le [\beta_1^{-1}(y-\beta_0)]^+]\Big]
\end{aligned}$$

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

□

Monte Carlo results

Study 3

	Polychoric			Adjusted polychoric		
Margins	ρ = .2	ρ = .4	ρ = .7	ρ = .2	ρ = .4	ρ = .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 1

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

	Polychoric			Adjusted polychoric		
Margins	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	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 2
Study 3: Coverage rates at the 95% level of confidence, sample size $n = 100$.

	Polychoric			Adjusted polychoric		
Margins	$\rho = .2$	$\rho = .4$	$\rho = .7$	$\rho = .2$	$\rho = .4$	$\rho = .7$
$\Gamma_1\Gamma_1$	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 3
Study 3: Coverage rates at the 95% level of confidence, sample size $n = 300$.

References

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. Chapman and Hall/CRC.

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

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

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

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)

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., 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

Höffding, W. (1940). Masstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und Instituts fur Angewandte Mathematik der Universitat Berlin*, *5*, 181–233.

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.

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

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.

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