PSYCHOMETRIKA—VOL. 82, NO. 1, 67–85
MARCH 2017
DOI: 10.1007/s11336-016-9512-2

ASYMPTOTIC ROBUSTNESS STUDY OF THE POLYCHORIC CORRELATION ESTIMATION

SHAOBO JIN AND FAN YANG-WALLENTIN

UPPSALA UNIVERSITY

Asymptotic robustness against misspecification of the underlying distribution for the polychoric correlation estimation is studied. The asymptotic normality of the pseudo-maximum likelihood estimator is derived using the two-step estimation procedure. The *t* distribution assumption and the skew-normal distribution assumption are used as alternatives to the normal distribution assumption in a numerical study. The numerical results show that the underlying normal distribution can be substantially biased, even though skewness and kurtosis are not large. The skew-normal assumption generally produces a lower bias than the normal assumption. Thus, it is worth using a non-normal distributional assumption if the normal assumption is dubious.

Key words: underlying distribution, asymptotic covariance matrix, non-normality, pseudo-maximum likelihood.

1. Introduction

Structural equation models (SEMs) are widely used in social sciences to model latent structures. Typically, normal distributions are assumed for both latent variables and error terms. However, observed measures in surveys are often ordinal. For example, a five-point Likert scale is commonly used in psychometric studies. Conceptually, categorical data should not be incorporated into a SEM by assuming they are continuous. There have been numerous advances in the literature on SEMs with respect to analysing ordinal data as they are. The observed ordinal data are usually assumed to be counterparts of some underlying continuous distributions. A typical choice of the underlying distributions is the standard normal distribution. Olsson (1979) studied the one-step maximum likelihood estimator (MLE) and the two-step MLE of the polychoric correlation coefficient. All parameters (i.e. thresholds and polychoric correlation) are estimated simultaneously for the one-step MLE, whereas the thresholds are estimated from the marginals and the polychoric correlation is computed based on the threshold estimates for the two-step MLE. Olsson showed that under the normality assumption, the one- and the two-step MLEs produce similar polychoric correlation estimates and similar variance estimates. Jöreskog (1994) derived the estimator of the asymptotic covariance matrix of the polychoric correlation estimators for the two-step maximum likelihood procedure (for a more compact expression, see Christoffersson & Gunsjö, 1996, and related references).

The underlying normality assumption is questionable. For example, the underlying normality assumption in the Life Orientation Test dataset (Scheier & Carver, 1985) was rejected by Maydeu-Olivares (2006). In yet another example, income is commonly used in the socio-economic status studies (e.g. Chateau, Metge, Prior, & Soodeen, 2012; Hodge & Treiman, 1968; Scharoun-Lee, Adair, Kaufman, & Gordon-Larsen, 2009). A Pareto distribution is classically used to model

Electronic supplementary material The online version of this article (doi:10.1007/s11336-016-9512-2) contains supplementary material, which is available to authorized users.

Correspondence should be made to Shaobo Jin, Department of Statistics, Uppsala University, 751 20 Uppsala, Sweden. Email: shaobo.jin@statistik.uu.se

© 2016 The Author(s). This article is published with open access at Springerlink.com

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

income (Arnold, 2008). Using a normal distribution to model income is dubious because the income is bounded by a lower limit. The question regarding income, however, is commonly categorized in a questionnaire: for example, see the National Longitudinal Study of Adolescent Health dataset (Carolina Population Center, 2009) used by Scharoun-Lee et al. (2009). Thus, "income" is an ordinal indicator with a non-normal underlying distribution. The consequences of violating the underlying normality assumption have been investigated (e.g. Flora & Curran, 2004; Lee & Lam, 1988; Quiroga, 1992). Flora and Curran (2004) generated non-normal data from the Fleishman–Vale–Maurelli method (Fleishman, 1978; Vale & Maurelli, 1983) in which a standard univariate normal random variable is polynomially transformed to introduce skewness and kurtosis. The authors found that the polychoric correlation estimates are only slightly biased when the underlying distribution has a skewness of 0.75 or 1.25 and a kurtosis of 1.75 or 3.75. They found, however, that the polychoric correlation is not robust against extreme underlying non-normality (e.g. skewness = 5 and kurtosis = 50). Lee and Lam (1988) generated non-normal data from an elliptical t distribution and an elliptical contaminated normal distribution and noted that the polychoric correlation estimates based on the normality assumption are fairly robust against non-normal underlying distributions. The study of Quiroga (1992) was conducted using non-normal data from an underlying bivariate skew-normal distribution and from the Fleishman–Vale–Maurelli method. The author also suggests that the polychoric correlation estimator is robust to non-normality. These studies share two features in common. First, they assume that the underlying distribution is normal to investigate the effect of underlying non-normality. So, a non-normal distribution assumption has not been systematically studied. Second, they are simulation studies. To our knowledge, there are no robustness studies on polychoric correlations from a theoretical standpoint.

Because the polychoric correlation is not distribution-free, tests of the underlying normality assumption are desired. For example, LISREL (Jöreskog & Sörbom, 1996) uses a likelihood ratio test to assess underlying normality, which is equivalent to a Pearson χ^2. Maydeu-Olivares, Forero, Gallardo-Pujol, and Renom (2009) and Maydeu-Olivares and Joe (2005, 2006) introduced a variant of the Pearson's χ^2 that is more suitable for the two-step MLE of the polychoric correlation. LISREL (Jöreskog & Sörbom, 1996) also provides the root-mean-square error of approximation (RMSEA) to assess the underlying normality assumption.

If the normality assumption fails, a new assumption of distribution is needed. Quiroga (1992) studied a new underlying distributional assumption whose marginal distributions are weighted averages of a univariate skew-normal distribution and a standard univariate normal distribution. Through an empirical example, the author showed that the polychoric correlation estimates based on the new assumption of distribution produce a smaller χ^2 test statistic. The normality assumption has also been criticized in the item response theory and alternative distributions have been studied to account for the underlying non-normality (e.g. see Bolfarine & Bazán, 2010; Lucke, 2014; Woods & Thissen, 2006).

The purpose of this paper is twofold. First, we study robustness against misspecification of the underlying distribution from a theoretical perspective. The effect of distributional misspecification under the two-step maximum likelihood procedure is investigated. Because the two-step MLE is computationally easier (Olsson, 1979) and is implemented in LISREL, we focus only on the two-step MLE for its simplicity and popularity. Second, the underlying distribution is not restricted to a standard normal distribution. The t distribution and the skew-normal distribution are used as alternatives in the present study. In particular, the skew-normal distribution has been applied in the item response theory as an alternative to the normality assumption (e.g. see Azevedo, Bolfarine, & Andrade, 2011; Bázan, Branco, & Bolfarine, 2006; Molenaar, 2015; Molenaar, Dolan, & de Boeck, 2012; Santos, Azevedo, & Bolfarine, 2013). Because the underlying distribution cannot be fully determined from ordinal data, we attempt to pinpoint potential alternatives for the bivariate normal distribution assumption.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

The remainder of this paper is organized as follows. General theories are presented, followed by numerical examples to illustrate our ideas. A brief conclusion ends the paper.

2. General Theory

Consider two ordinal variables U and V with m_U and m_V categories, respectively. The classic polychoric correlation estimation method assumes that there are two underlying continuous variables X and Y for U and V, respectively. The values of U and V are defined through X and Y as

$$U = i \Leftrightarrow \tau_{i-1} < X \leq \tau_i \quad i = 1, 2, \ldots, m_U,$$
$$V = j \Leftrightarrow \xi_{j-1} < Y \leq \xi_j \quad j = 1, 2, \ldots, m_V,$$

where $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{m_U-1})'$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{m_V-1})'$ are thresholds such that

$$-\infty = \tau_0 < \tau_1 < \cdots < \tau_{m_U-1} < \tau_{m_U} = \infty,$$
$$-\infty = \xi_0 < \xi_1 < \cdots < \xi_{m_V-1} < \xi_{m_V} = \infty.$$

The true joint distribution function is denoted by $\mathrm{F}(x, y; \rho, \boldsymbol{\zeta})$ with two marginal distributions $\mathrm{F}_1(x)$ and $\mathrm{F}_2(y)$, where ρ is the correlation coefficient and $\boldsymbol{\zeta}$ is the vector of other parameters (e.g. degrees of freedom, location, and scale parameters). The corresponding joint density function is $\mathrm{f}(x, y; \rho)$ with marginal densities $\mathrm{f}_1(x)$ and $\mathrm{f}_2(y)$. Because the true distribution family is unknown, we assume the underlying distribution to be $\mathrm{H}(x, y; \rho)$ with marginal distributions $\mathrm{H}_1(x)$ and $\mathrm{H}_2(y)$. The joint density function is $\mathrm{h}(x, y; \rho)$ with marginal densities $\mathrm{h}_1(x)$ and $\mathrm{h}_2(y)$, respectively. Conventionally, $\mathrm{H}(x, y; \rho)$ is taken to be the distribution function of a standard bivariate normal distribution. The normality assumption will be relaxed in our study. We also allow for different marginal distributions both in true underlying distributions and in the assumed ones.

2.1. Two-Step Estimation

2.1.1. Threshold Estimation Let n_{ij} and p_{ij} be the observed frequency and proportion, respectively, of $U = i$ and $V = j$, for $i = 1, \ldots, m_U$ and $j = 1, \ldots, m_V$. If the true underlying distribution F is different from the assumed distribution H, the MLEs of thresholds will be inconsistent estimators of $\boldsymbol{\tau}_0 = (\tau_{1,0}, \ldots, \tau_{m_U-1,0})'$ and $\boldsymbol{\xi}_0 = (\xi_{1,0}, \ldots, \xi_{m_V-1,0})'$, where the subscript 0 indicates true values. Consider the ordinal variable U first. Denote $\boldsymbol{n}_U = (n_{1\cdot}, \ldots, n_{m_U\cdot})'$, where $n_{i\cdot} = \sum_{j=1}^{m_V} n_{ij}$ is the marginal total for $i = 1, 2, \ldots, m_U$. The corresponding marginal proportion is $\boldsymbol{p}_U = (p_{1\cdot}, \ldots, p_{m_U\cdot})'$. The pseudo-maximum likelihood estimator (PMLE) of $\boldsymbol{\tau}$, denoted as $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_{m_U-1})'$, is obtained by maximizing

$$Q(\boldsymbol{\tau}) = \sum_{i=1}^{m_U} n_{i\cdot} \log \int_{\tau_{i-1}}^{\tau_i} \mathrm{h}_1(x)\mathrm{d}x \ .$$

It is easy to see that $\hat{\boldsymbol{\tau}}$ is a consistent estimator of $\boldsymbol{\tau}^*$, where $\mathrm{H}_1(\boldsymbol{\tau}^*) = \mathrm{F}_1(\boldsymbol{\tau}_0)$, because the observed cell probabilities are consistent estimators of $\mathrm{F}_1(\boldsymbol{\tau}_0)$. Similarly, $\hat{\boldsymbol{\xi}}$ is a consistent estimator of $\boldsymbol{\xi}^*$, where $\mathrm{H}_2(\boldsymbol{\xi}^*) = \mathrm{F}_2(\boldsymbol{\xi}_0)$. Let $\boldsymbol{P}$ be an $m_U \times m_V$ matrix with (i, j)-th entry p_{ij}. Then

$$\frac{\partial \mathrm{Q}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} = n\boldsymbol{B}_U(\boldsymbol{\tau})' \boldsymbol{D}_U^{-1}(\boldsymbol{\tau}) \boldsymbol{p}_U,$$

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

$$\frac{\partial^2 \mathrm{Q}(\boldsymbol{\tau})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} = -n \boldsymbol{B}_U(\boldsymbol{\tau})' \boldsymbol{D}_U^{-1}(\boldsymbol{\tau}) \boldsymbol{D}_{\boldsymbol{p}} \boldsymbol{D}_U^{-1}(\boldsymbol{\tau}) \boldsymbol{B}_U(\boldsymbol{\tau}) + n\boldsymbol{S},$$

where $n = \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} n_{ij}$ is the total number of observations,

$$\boldsymbol{B}_U(\boldsymbol{\tau}) = \begin{pmatrix} \mathrm{h}_1(\tau_1) & 0 & \cdots & 0 \\ -\mathrm{h}_1(\tau_1) & \mathrm{h}_1(\tau_2) & \cdots & 0 \\ 0 & -\mathrm{h}_1(\tau_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathrm{h}_1(\tau_{m_U-1}) \\ 0 & 0 & \cdots & -\mathrm{h}_1(\tau_{m_U-1}) \end{pmatrix},$$

$\boldsymbol{D}_U(\boldsymbol{\tau}) = \mathrm{Diag}\left(\int_{\tau_0}^{\tau_1} \mathrm{h}_1(x)\mathrm{d}x, \ldots, \int_{\tau_{m_U-1}}^{\tau_{m_U}} \mathrm{h}_1(x)\mathrm{d}x\right)$, $\boldsymbol{D}_{\boldsymbol{p}} = \mathrm{Diag}(p_{1\cdot}, \ldots, p_{m_U\cdot})$, $\boldsymbol{p}_U = \boldsymbol{P}\mathbf{1}_{m_V}$ with $\mathbf{1}_{m_V}$ being an $m_V \times 1$ vector of 1's, and $\boldsymbol{S}$ is a diagonal matrix with i-th element

$$\left(\frac{p_{i,\cdot}}{\int_{\tau_{i-1}}^{\tau_i} \mathrm{h}_1(x)\mathrm{d}x} - \frac{p_{i+1,\cdot}}{\int_{\tau_i}^{\tau_{i+1}} \mathrm{h}_1(x)\mathrm{d}x}\right) \frac{\partial \mathrm{h}_1(\tau_i)}{\partial \tau_i},$$

for $i = 1, \ldots, m_U - 1$. The operator $\mathrm{Diag}(\cdot)$ constructs a diagonal matrix using the enclosed vector as diagonal elements. The Taylor expansion of $n^{-1/2}\partial Q(\hat{\boldsymbol{\tau}})/\partial \boldsymbol{\tau}$ around $\boldsymbol{\tau}^*$ is

$$0 = n^{-1/2} \frac{\partial Q\left(\hat{\boldsymbol{\tau}}\right)}{\partial \boldsymbol{\tau}} = n^{-1/2} \frac{\partial Q(\boldsymbol{\tau}^*)}{\partial \boldsymbol{\tau}} + n^{-1/2} \frac{\partial^2 Q(\tilde{\boldsymbol{\tau}})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'} \left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}^*\right) + o_p(1), \tag{1}$$

where $\tilde{\boldsymbol{\tau}}$ lies between $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}^*$. Because both $\hat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\tau}}$ are consistent estimators of $\boldsymbol{\tau}^*$, $n^{-1}\partial^2 Q(\tilde{\boldsymbol{\tau}})/\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'$ is consistent for $-\boldsymbol{B}_U(\boldsymbol{\tau}^*)' \boldsymbol{D}_U^{-1}(\boldsymbol{\tau}^*) \boldsymbol{B}_U(\boldsymbol{\tau}^*)$. So, Eq. (1) implies

$$n^{1/2}\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}^*\right) = n^{1/2}\left[\boldsymbol{B}_U\left(\boldsymbol{\tau}^*\right)' \boldsymbol{D}_U^{-1}\left(\boldsymbol{\tau}^*\right) \boldsymbol{B}_U\left(\boldsymbol{\tau}^*\right)\right]^{-1} \boldsymbol{B}_U\left(\boldsymbol{\tau}^*\right)' \boldsymbol{D}_U^{-1}\left(\boldsymbol{\tau}^*\right) \boldsymbol{p}_U + o_p(1). \tag{2}$$

Similar arguments applying to $\boldsymbol{\xi}$ yield

$$n^{1/2}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right) = n^{1/2}\left[\boldsymbol{B}_V\left(\boldsymbol{\xi}^*\right)' \boldsymbol{D}_V^{-1}\left(\boldsymbol{\xi}^*\right) \boldsymbol{B}_V\left(\boldsymbol{\xi}^*\right)\right]^{-1} \boldsymbol{B}_V\left(\boldsymbol{\xi}^*\right)' \boldsymbol{D}_V^{-1}\left(\boldsymbol{\xi}^*\right) \boldsymbol{p}_V + o_p(1), \tag{3}$$

where $\boldsymbol{p}_V = \boldsymbol{P}'\mathbf{1}_{m_U}$. Here $\boldsymbol{B}_V$ and $\boldsymbol{D}_V$ are defined by substituting h_1 with h_2 in $\boldsymbol{B}_U$ and $\boldsymbol{D}_U$. The PMLEs $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\xi}}$ are inconsistent in the sense that $\boldsymbol{\tau}^*$ and $\boldsymbol{\xi}^*$ are different from the true values $\boldsymbol{\tau}_0$ and $\boldsymbol{\xi}_0$.

2.1.2. Polychoric Correlation Coefficient Estimation Under the distributional assumption H, the assumed cell probability is

$$\pi_{ij,(\mathrm{H})} = \int_{\tau_{i-1}}^{\tau_i} \int_{\xi_{j-1}}^{\xi_j} \mathrm{h}(x, y)\mathrm{d}y\mathrm{d}x,$$

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

while the true cell probability $\pi_{ij,(\mathrm{F})}$ is obtained by substituting $\mathrm{h}(x, y)$ with $\mathrm{f}(x, y)$. Conditionally on $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\xi}}$, the polychoric correlation ρ is estimated by maximizing

$$\mathrm{L}\left(\rho, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) = \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} p_{ij} \log\, \pi_{ij,(\mathrm{H})}.$$

Theorem 2.2 in White (1982) shows that the PMLE is a consistent estimator that minimizes the Kullback–Leibler information (Kullback & Leibler, 1951) under some regularity conditions, one of which is that the absolute value of $\log \pi_{ij,(\mathrm{H})}$ is dominated by a variable with finite expectation. Such a regularity condition is satisfied if $\pi_{ij,(\mathrm{H})} = 0$ implies $\pi_{ij,(\mathrm{F})} = 0$ for all (i, j). Consequently, Theorem 2.2 in White (1982) shows that $\hat{\rho}$ converges to ρ^* that minimizes the Kullback–Leibler information

$$\sum_{i=1}^{m_U} \sum_{j=1}^{m_V} \pi_{ij,(\mathrm{F})} \left(\log\, \pi_{ij,(\mathrm{F})} - \log\, \pi_{ij,(\mathrm{H})}\right).$$

Theorem 1. *Assume* $g\left(\rho, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right) = \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} \pi_{ij,(F)} \log\, \pi_{ij,(H)}$, *as a function of* ρ, *has a unique maximum at* ρ^*. *If* $\pi_{ij,(H)} = 0$ *implies* $\pi_{ij,(F)} = 0$ *for all* (i, j), *then there exists a root* $\hat{\rho}$ *of the equation*

$$\frac{\partial}{\partial \rho} \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} p_{ij} \log\, \pi_{ij,(H)} = 0$$

such that $\hat{\rho}$ *is a consistent estimator of* ρ^*.

That is, $\hat{\rho}$ is a consistent estimator of ρ^* that minimizes the probabilistic divergence between H and F (Kullback, 1959) in the sense of the Kullback–Leibler information. This minimized divergence implies similarities of H and F in terms of cell probabilities.

The assumption in Theorem 1 requires uniqueness of the maximum. In so doing, we rule out all cases with local maxima. If we have several stationary points, we can then only conclude that one of the stationary points minimizes the Kullback–Leibler information.

2.1.3. Asymptotic Variance of Polychoric Correlations Let $\mathrm{L}_\rho\left(\rho, \boldsymbol{\tau}, \boldsymbol{\xi}\right)$ denote the first order partial derivative of $\mathrm{L}\left(\rho, \boldsymbol{\tau}, \boldsymbol{\xi}\right)$ with respect to ρ. Similar symbols are used to represent other partial derivatives and higher order partial derivatives. $\mathrm{L}_\rho\left(\hat{\rho}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right)$ can be expanded around ρ^* for a sufficiently large n,

$$0 = n^{1/2}\mathrm{L}_\rho\left(\hat{\rho}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) = \underbrace{n^{1/2}\mathrm{L}_\rho\left(\rho^*, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right)}_{(i)} + \underbrace{n^{1/2}\left(\hat{\rho} - \rho^*\right)\mathrm{L}_{\rho\rho}\left(\tilde{\rho}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right)}_{(ii)} + o_p(1), \quad (4)$$

where $\tilde{\rho}$ lies between ρ^* and $\hat{\rho}$. Term (i) in Eq. (4) is equivalent to

$$\begin{aligned} n^{1/2}\mathrm{L}_\rho\left(\rho^*, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) =& n^{1/2}\mathrm{L}_\rho\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right) + n^{1/2}\mathrm{L}_{\rho\boldsymbol{\tau}}\left(\rho^*, \tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\xi}}\right)'\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}^*\right) \\ &+ n^{1/2}\mathrm{L}_{\rho\boldsymbol{\xi}}\left(\rho^*, \tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\xi}}\right)'\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right) + o_p(1), \end{aligned} \quad (5)$$

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

where $\tilde{\boldsymbol{\tau}}$ lies between $\hat{\boldsymbol{\tau}}$ and $\boldsymbol{\tau}^*$ and $\tilde{\boldsymbol{\xi}}$ lies between $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}^*$. Hence, if $\pi_{ij,(\mathrm{H})} = 0$ implies $\pi_{ij,(\mathrm{F})} = 0$ for all (i, j) in a neighbourhood of $(\boldsymbol{\tau}^*, \boldsymbol{\xi}^*)$ given the correlation ρ^*, $\mathrm{L}_{\rho\boldsymbol{\tau}}\left(\rho^*, \tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\xi}}\right)$ is consistent for $\mathrm{g}_{\rho\boldsymbol{\tau}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)$ and $\mathrm{L}_{\rho\boldsymbol{\xi}}\left(\rho^*, \tilde{\boldsymbol{\tau}}, \tilde{\boldsymbol{\xi}}\right)$ is consistent for $\mathrm{g}_{\rho\boldsymbol{\xi}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)$. Thus, Eq. (5) is equivalent to

$$\begin{aligned} n^{1/2}\mathrm{L}_\rho\left(\rho^*, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) = {} & n^{1/2}\mathrm{L}_\rho\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right) + n^{1/2}\mathrm{g}_{\rho\boldsymbol{\tau}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)'\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}^*\right) \\ & + n^{1/2}\mathrm{g}_{\rho\boldsymbol{\xi}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)'\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right) + o_p(1). \end{aligned}$$

Likewise, Term (ii) can be written as

$$n^{1/2}\left(\hat{\rho} - \rho^*\right)\mathrm{L}_{\rho\rho}\left(\tilde{\rho}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) = n^{1/2}\left(\hat{\rho} - \rho^*\right)\mathrm{g}_{\rho\rho}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right) + o_p(1),$$

provided that $\pi_{ij,(\mathrm{H})} = 0$ implies $\pi_{ij,(\mathrm{F})} = 0$ in a neighbourhood of $\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)$. Hence, combining with Eqs. (2) and (3), (4) is equivalent to

$$\begin{aligned} n^{1/2}\left(\hat{\rho} - \rho^*\right) = {} & -\frac{n^{1/2}}{\mathrm{g}_{\rho\rho}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)}\left\{\mathrm{L}_\rho\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right) + \mathrm{g}_{\rho\boldsymbol{\tau}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)'\left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}^*\right)\right. \\ & \left. + \mathrm{g}_{\rho\boldsymbol{\xi}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)'\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^*\right)\right\} + o_p(1) \\ = {} & -\frac{n^{1/2}}{\mathrm{g}_{\rho\rho}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)}\mathrm{tr}\left(\boldsymbol{\Lambda}'\boldsymbol{P}\right) + o_p(1), \end{aligned}$$

where $\boldsymbol{\Lambda} = \boldsymbol{A} + \boldsymbol{E}_\tau \mathrm{g}_{\rho\boldsymbol{\tau}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)\mathbf{1}'_{m_V} + \mathbf{1}_{m_U}\mathrm{g}_{\rho\boldsymbol{\xi}}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)'\boldsymbol{E}'_\xi$ with $\boldsymbol{A}$ being an $m_U \times m_V$ matrix with (i, j)-th element $\left(\partial\pi_{ij}/\partial\rho\right)/\pi_{ij}$, $\boldsymbol{E}_\tau = \boldsymbol{D}_\tau^{-1}\boldsymbol{B}_\tau\left(\boldsymbol{B}'_\tau\boldsymbol{D}_\tau^{-1}\boldsymbol{B}_\tau\right)^{-1}$, and $\boldsymbol{E}_\xi = \boldsymbol{D}_\xi^{-1}\boldsymbol{B}_\xi\left(\boldsymbol{B}'_\xi\boldsymbol{D}_\xi^{-1}\boldsymbol{B}_\xi\right)^{-1}$. Note that $\mathrm{tr}\left(\boldsymbol{\Lambda}'\boldsymbol{P}\right) = \mathrm{vec}(\boldsymbol{\Lambda})'\mathrm{vec}(\boldsymbol{P})$, where $\mathrm{vec}(\cdot)$ stacks the columns of the enclosed matrix and

$$\sqrt{n}\left(\mathrm{vec}(\boldsymbol{P}) - \mathrm{vec}(\boldsymbol{\pi}_{(\mathrm{F})})\right) \xrightarrow{d} N\left(\mathbf{0}, \mathrm{Diag}(\boldsymbol{P}) - \mathrm{vec}(\boldsymbol{P})\mathrm{vec}(\boldsymbol{P})'\right)$$

with $\boldsymbol{\pi}_{(\mathrm{F})}$ being an $m_U \times m_V$ matrix with (i, j)-th entry $\pi_{ij,(\mathrm{F})}$. The arguments above establish the following theorem.

Theorem 2. *Let $\hat{\rho}$ be the consistent root of $L_\rho\left(\rho, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right) = 0$ given $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\xi}}$. Assume $\pi_{ij,(H)} = 0$ implies $\pi_{ij,(F)} = 0$ in a neighbourhood of $\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)$, then*

$$n^{1/2}\left(\hat{\rho} - \rho^*\right) \xrightarrow{d} N\left(0, \sigma^2\right),$$

where $\sigma^2 = \left[tr\left((\boldsymbol{\Lambda} \odot \boldsymbol{\Lambda})'\boldsymbol{\pi}_{(F)}\right) - \left(tr\left(\boldsymbol{\Lambda}'\boldsymbol{\pi}_{(F)}\right)\right)^2\right] / \left(g_{\rho\rho}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)\right)^2$. Here, matrix $\boldsymbol{\Lambda}$ is evaluated under $\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)$. The operator $\odot$ implies element-wise multiplication.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

2.1.4. Estimating the Asymptotic Covariance Matrix Following Theorem 2, the asymptotic variance of $\hat{\rho}$ can be consistently estimated by

$$\frac{1}{\left[\mathrm{L}_{\rho\rho}\left(\hat{\rho}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\xi}}\right)\right]^2}\left[\sum_{i=1}^{m_U}\sum_{j=1}^{m_V}\hat{\lambda}_{ij}^2 p_{ij} - \left(\sum_{i=1}^{m_U}\sum_{j=1}^{m_V}\hat{\lambda}_{ij} p_{ij}\right)^2\right],$$

where $\hat{\lambda}_{ij}$ is the (i, j)-th element in $\hat{\boldsymbol{\Lambda}}$. The polychoric correlation between variables U and V satisfies

$$n^{1/2}\left(\hat{\rho}^{(UV)} - \rho^{*(UV)}\right) = \frac{-n^{1/2}}{\mathrm{g}_{\rho\rho}^{(UV)}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)}\mathrm{tr}\left(\boldsymbol{\Lambda}^{(UV)'}\boldsymbol{P}^{(UV)}\right) + o_p(1),$$

where the superscript (UV) emphasizes that all quantities are evaluated under the distributional assumption for U and V. Similarly, the polychoric correlation between variables K and Z satisfies

$$n^{1/2}\left(\hat{\rho}^{(KZ)} - \rho^{*(KZ)}\right) = \frac{-n^{1/2}}{\mathrm{g}_{\rho\rho}^{(KZ)}\left(\rho^*, \boldsymbol{\tau}^*, \boldsymbol{\xi}^*\right)}\mathrm{tr}\left(\boldsymbol{\Lambda}^{(KZ)'}\boldsymbol{P}^{(KZ)}\right) + o_p(1).$$

The underlying distributional assumption for U and V can either be the same as that for K and Z or different. Thus, the asymptotic covariance between $\hat{\rho}^{(UV)}$ and $\hat{\rho}^{(KZ)}$ is consistently estimated by

$$\frac{\sum_{a=1}^{m_U}\sum_{b=1}^{m_V}\sum_{c=1}^{m_K}\sum_{d=1}^{m_Z}\hat{\lambda}_{ab}^{(UV)}\left(p_{abcd}^{(UVKZ)} - p_{ab}^{(UV)}p_{cd}^{(KZ)}\right)\hat{\lambda}_{cd}^{(KZ)}}{\mathrm{L}_{\rho\rho}\left(\hat{\rho}^{(UV)}, \hat{\boldsymbol{\tau}}^{(UV)}, \hat{\boldsymbol{\xi}}^{(UV)}\right)\mathrm{L}_{\rho\rho}\left(\hat{\rho}^{(KZ)}, \hat{\boldsymbol{\tau}}^{(KZ)}, \hat{\boldsymbol{\xi}}^{(KZ)}\right)}, \tag{6}$$

where $p_{abcd}^{(UVKZ)}$ is the sample proportion of observing $U = a$, $V = b$, $K = c$, and $Z = d$. Under the assumption that the underlying distribution is normal and correctly specified, Eq. (6) reduces to the estimator in Jöreskog (1994).

2.2. A Variant of Two-Step Estimation

The above two-step estimation is applicable to bivariate distributions whose marginal distributions do not depend on unknown parameters. For example, the mean and variance of a bivariate normal distribution are unknown parameters and assuming a standard normal distribution fixes those parameters to known values. In many other distributions, unknown parameters are included in the marginal distributions. Consequently, the above two-step MLE cannot be obtained unless the unknown parameters are prefixed. In such a case, a variant of the two-step MLE can be obtained instead. The MLE maximizes

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{m_U}\sum_{j=1}^{m_V} p_{ij}\log\ \pi_{ij,(\mathrm{H})}$$

with respect to the vector $\boldsymbol{\theta}$ that consists of free unknown parameters. Not all parameters in ρ and $\boldsymbol{\zeta}$ are free parameters. The mean and variance of an ordinal variable are not identified. Thus, the scale

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

and location parameters that do not contribute to the correlation coefficient are not identified. In some distributions (e.g. the skew-normal distribution introduced later), the correlation coefficient is also determined by $\boldsymbol{\zeta}$ and, therefore, is not a free parameter. If $L(\boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} p_{ij} \log \pi_{ij,(\mathrm{H})} = \mathbf{0}$$

is solved to obtain $\hat{\boldsymbol{\theta}}$. By standard calculation,

$$n^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) = -n^{1/2} E\left(L_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}^*)\right)^{-1} \boldsymbol{C}(\boldsymbol{\theta}^*)\mathrm{vec}(\boldsymbol{P}) + o_p(1),$$

where $\boldsymbol{\theta}^* = (\rho^*, \boldsymbol{\zeta}^{*'})'$ and the k-th row in $\boldsymbol{C}$ is $\mathrm{vec}(\boldsymbol{C}_k)'$ with the (i, j)-th element being $\frac{1}{\pi_{ij,(H)}} \frac{\partial \pi_{ij,(H)}}{\partial \boldsymbol{\theta}_k}$, provided that $E\left(L_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}^*)\right)$ is invertible. Assume that the correlation coefficient satisfies $\rho = \rho(\boldsymbol{\theta})$ in which $\rho_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$ is nonzero. The delta method (Ferguson, 1996) indicates

$$n^{1/2}\left(\hat{\rho} - \rho^*\right) \xrightarrow{d} N\left(0, \rho_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*)' E\left(L_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}^*)\right)^{-1} \boldsymbol{C}(\boldsymbol{\theta}^*)\boldsymbol{\Omega}\boldsymbol{C}(\boldsymbol{\theta}^*)' E\left(L_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}^*)\right)^{-1} \rho_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*)\right),$$

where $\boldsymbol{\Omega}$ is the asymptotic covariance matrix of $\mathrm{vec}(\boldsymbol{P})$. Hence, the asymptotic covariance between $\hat{\rho}^{(UV)}$ and $\hat{\rho}^{(KZ)}$ can be consistently estimated in a similar manner to Eq. (6). Let the matrix $\boldsymbol{\Upsilon}$ be constructed through $\mathrm{vec}(\boldsymbol{\Upsilon})' = \rho_{\boldsymbol{\theta}}(\boldsymbol{\theta})' E\left(L_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{C}(\boldsymbol{\theta})$. Then the asymptotic covariance between $\hat{\rho}^{(UV)}$ and $\hat{\rho}^{(KZ)}$ is consistently estimated by

$$\sum_{a=1}^{m_U} \sum_{b=1}^{m_V} \sum_{c=1}^{m_K} \sum_{d=1}^{m_Z} \hat{\boldsymbol{\Upsilon}}_{ab}^{(UV)} \left(p_{abcd}^{(UVKZ)} - p_{ab}^{(UV)} p_{cd}^{(KZ)}\right) \hat{\boldsymbol{\Upsilon}}_{cd}^{(KZ)}. \tag{7}$$

In Olsson (1979), thresholds $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ are parameters for the one-step MLE. However, the thresholds are not always directly estimated for the one-step MLE for other distributions. For example, the density function of a bivariate skew-normal distribution in Azzalini and Valle (1996) is

$$\mathrm{f}(x, y) = 2\phi_2(x, y; \omega)\, \Phi(\alpha_1 x + \alpha_2 y)\ , \tag{8}$$

where $\phi_2(\cdot, \cdot; \omega)$ is the density function of the bivariate standard normal distribution with correlation coefficient ω, $\Phi(\cdot)$ is the distribution function of a standard normal distribution, and α_1 and α_2 control skewness and kurtosis. The covariance matrix of X and Y is

$$\boldsymbol{W} - \frac{2}{\pi\left(1 + \alpha_1^2 + 2\omega\alpha_1\alpha_2 + \alpha_2^2\right)} \boldsymbol{R}\ , \tag{9}$$

where

$$\boldsymbol{W} = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{R} = \begin{pmatrix} (\alpha_1 + \omega\alpha_2)^2 & (\alpha_1 + \omega\alpha_2)(\alpha_2 + \omega\alpha_1) \\ (\alpha_1 + \omega\alpha_2)(\alpha_2 + \omega\alpha_1) & (\alpha_2 + \omega\alpha_1)^2 \end{pmatrix}.$$

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

Thus, the correlation coefficient is affected by ω, α_1 and α_2. The marginal distributions are univariate skew-normal distributions with densities

$$\mathrm{f}(x) = 2\phi(x)\,\Phi(\alpha x), \tag{10}$$

where $\alpha = (\alpha_1 + \omega\alpha_2)/\left[1 + (1-\omega^2)\alpha_2^2\right]^{1/2}$ for X and $\alpha = (\alpha_2 + \omega\alpha_1)/\left[1 + (1-\omega^2)\alpha_1^2\right]^{1/2}$ for Y. Bazán et al. (2006), Molenaar (2015), and Molenaar et al. (2012) have applied the univariate skew-normal distribution to the item response theory. The marginal distributions are affected by ω, α_1, and α_2, as are the thresholds. Therefore, the thresholds are not free parameters. The vector of free parameters in the variant of two-step estimation is $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \omega)'$.

3. Numerical Examples

A numerical study is conducted in this section to examine the asymptotic bias under different distributional assumptions. Asymptotic limits of PMLE for polychoric correlation coefficients are numerically computed.

3.1. Distributional Assumption

Four experiments are conducted in which different true underlying distributions are investigated.

3.1.1. Experiment 1: Elliptical Distribution In probability and statistics, an elliptical distribution belongs to a broad family of probability distributions. The bivariate joint density function of an elliptical distribution is of the form

$$\frac{1}{2\pi\sigma_{11}\sigma_{22}\left(1-\rho^2\right)^{1/2}}\mathrm{q}(z), \tag{11}$$

where $\mathrm{q}(\cdot)$ is a univariate function and

$$z = \frac{1}{1-\rho^2}\left[\frac{(x-\mu_1)^2}{\sigma_{11}^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_{11}\sigma_{22}} + \frac{(y-\mu_2)^2}{\sigma_{22}^2}\right]$$

with σ_{11} being the variance of X and σ_{22} being the variance of Y. An elliptical distribution generalizes the normal distribution and keeps some properties (e.g. Balakrishnan & Lai, 2009; Fang, Kotz, & Ng, 1990; Kelker, 1970). Some examples of the bivariate elliptical distributions that will be used later are

1. Normal distributions: $q(z) = \exp(-z/2)$;
2. $t(v)$ distributions with degrees of freedom v: $q(z) = (1 + z/v)^{-(v+2)/2}$;
3. Bivariate uniform distributions: $q(z) = 2\mathrm{I}_{\{z\leq 1\}}$ with I being an indicator function;
4. Bivariate Logistic distributions: $q(z) = 4\exp(-z)/\left[1+\exp(-z)\right]^2$;
5. Bivariate exponential power distributions: $q(z) = 2\exp\left(-z^{\beta}/2\right)/\left(2^{1/\beta}\Gamma(1+1/\beta)\right)$.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

The elliptical distribution family plays a very important role in robustness studies (e.g. Kano, Berkane, & Bentler, 1993). In the context of Pearson correlation estimation, Hampel, Ronchetti, Rousseeuw, and Stahel (1986) showed that the PMLE of the covariance matrix is proportional to the MLE under the true distributional assumption, provided that continuous data have been acquired. This result enables us to use any member of the family to estimate the correlation matrix, having the same estimates as if the true distribution were used. Likewise, Berkane, Kano, and Bentler (1994) claimed that "there is practically no cost in treating the distribution as multivariate t with specified (possibly small) degrees of freedom" (Berkane et al., 1994, p. 266) when the true distribution is normal and continuous data are observed. It only slightly inflates the variance of the resulting estimator. Thus, it is worth investigating the effect of an underlying elliptical distribution. Because we have only categorical data, the mean and variance are not identified. But then only the correlation coefficient ρ is the parameter of interest, so we can assume $\mu_1 = \mu_2 = 0$ and $\sigma_{11} = \sigma_{22} = 1$.

For some members of the elliptical distribution family, the marginal distribution is still elliptical but not of the same type (Gómez, Gómez-villegas, & Marín, 2003). The bivariate uniform distribution, the logistic distribution, and the exponential power distribution possess such properties. The support of the bivariate uniform distribution is not the whole Cartesian plane, whereas the other distributions have the whole Cartesian plane as their support. The exponential power distribution includes the normal distribution ($\beta = 1$) and the Laplace distribution ($\beta = 1/2$) as special cases.

3.1.2. Experiment 2: Skew-Normal Distribution An elliptical distribution is symmetric. Quiroga (1992) reported that kurtosis does not have strong effects on the polychoric correlation but that skewness increases the bias. The above elliptical distributions examine various values of kurtosis. The following distributions introduce nonzero values of skewness.

A natural generalization of a standard normal distribution is the univariate skew-normal distribution proposed by Azzalini (1985) and extended by Azzalini and Valle (1996) to a multivariate skew-normal distribution. The bivariate density function, covariance matrix and marginal density function are shown in Eqs. (8), (9), and (10), respectively. The ranges of the skewness and excess kurtosis are $(-0.9953, 0.9953)$ and $[0, 0.8692)$, respectively (Azzalini & Capitanio, 2014, p. 32). This range is close to the low skewness and low kurtosis case in Flora and Curran (2004). The reader can refer to Azzalini (2005) for an overview of the skew-normal distribution and to Azzalini and Capitanio (2014) for the expressions of skewness and excess kurtosis. Note that the bivariate skew-normal distribution proposed by Azzalini and Valle (1996) is different from the skew-normal distribution in Quiroga (1992). The specification in Azzalini and Valle (1996) is used in the present study for its connection with the skew-t(υ) distribution in the next experiment.

3.1.3. Experiment 3: Skew-$t(\upsilon)$ Distribution Skewness can also be introduced to the t(υ) distribution. Azzalini and Capitanio (2003) proposed a multivariate skew-$t(\upsilon)$ distribution whose bivariate density function is

$$\mathrm{f}(x, y) = 2\mathrm{t}(x, y; \omega, \upsilon)\,\mathrm{T}\left((\alpha_1 x + \alpha_2 y)\left(\frac{\upsilon + 2}{\upsilon + (x^2 - 2\omega xy + y^2)/(1 - \rho^2)}\right)^{1/2}; \upsilon + 2\right),$$

where $\mathrm{t}(\cdot, \cdot; \omega, \upsilon)$ is the density function of a standard bivariate t distribution with correlation w and degrees of freedom υ and $\mathrm{T}(\cdot; \upsilon + 2)$ is the distribution function of a univariate t distribution with degrees of freedom $\upsilon + 2$. The covariance matrix of X and Y is

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

$$\frac{v}{v-2}\boldsymbol{W} - \frac{2}{\pi\left(1+\alpha_1^2+2\omega\alpha_1\alpha_2+\alpha_2^2\right)}\boldsymbol{R},$$

provided that $v > 2$. Both marginal distributions are univariate skew-t distributions with density function

$$\mathrm{f}(x) = 2\mathrm{t}(x; v)\,\mathrm{T}\left(\alpha x\left(\frac{v+1}{v+x^2}\right)^{1/2}; v+1\right).$$

The reader is directed to Azzalini and Capitanio (2003) for the expressions of skewness and excess kurtosis.

3.1.4. Experiment 4: Other Distributions The skew-normal and t(v) distributions are special cases of the skew-t(v) distribution family. There are many distributions that are not members of the skew-elliptical distribution family. In addition, the underlying distribution cannot be truly determined from the observed ordinal data. It is therefore important to investigate the effect of distributional misspecification using the distributions that do not belong to the skew-t distribution family. A Pareto distribution is commonly used to model income (Arnold, 2008) and income is commonly used as an indicator of socio-economic status. Mardia (1962) proposed a multivariate Pareto distribution in which the bivariate density function is

$$\mathrm{f}(x, y; a, \theta_1, \theta_2) = (a+1)a(\theta_1\theta_2)^{a+1}(\theta_2 x + \theta_1 y - \theta_1\theta_2)^{-(a+2)},$$

and the marginal density function is $\mathrm{f}(x) = a\theta_i^a x^{-(a+1)}$, with $x \geq \theta_1 > 0$, $y \geq \theta_2 > 0$, and $a > 0$. The correlation coefficient between X and Y is $1/a$, which is always positive.

3.2. Numerical Design

Three combinations of categories are used. First, both U and V have five categories with cell probabilities (0.1, 0.2, 0.4, 0.2, 0.1) and (0.1, 0.1, 0.3, 0.3, 0.2), respectively. Second, both U and V have three categories with cell probabilities (0.2, 0.5, 0.3) and (0.1, 0.3, 0.6), respectively. Third, U has three categories with cell probabilities (0.2, 0.5, 0.3) and V has five categories with cell probabilities (0.1, 0.1, 0.3, 0.3, 0.2).

In Experiment 1, β in the exponential power distribution is $\beta = 0.3, 0.4, 0.5, 0.6$. In Experiments 2 and 3, three values of α_1 are considered ($\alpha_1 = 0.1, 0.5, 1$) and 20 evenly spaced values of α_2 are considered ranging from 0.5 to 10 for both skew-normal and skew-t(v) distributions. Thus, different combinations of univariate skewness and kurtosis are investigated. In Experiments 1, 2 and 3, the degrees of freedom for the t(v) and skew-t(v) distributions are 4, 6, 8, and 10. In Experiment 4, parameters for the Pareto distribution are $\theta_1 = \theta_2 = 3$.

For all experiments, two values of ρ are used: $\rho_0 = 0.4, 0.6$. For the purpose of illustration, the assumed underlying distributions are bivariate normal, skew-normal, and t(v) distributions. The normal assumption consists of only one unknown parameter of interest, ρ. The skew-normal assumption consists of three parameters: α_1, α_2, and ω that determine the correlation coefficient. The degrees of freedom in the t(v) are prefixed to be 4, 6, 8, 10 and the correlation coefficient is the only parameter of interest. The expressions of the partial derivatives of the skew-normal distribution and t distribution can be found in the supplementary materials.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

3.3. Numerical Results

To assess the bias of polychoric correlation estimates, the relative bias (RB) is computed, which is defined as $RB = 100 \times \left(\hat{\rho} - \rho_0\right)/\rho_0$. Following the definition in Flora and Curran (2004), $RB \leq 5, 5 < RB \leq 10$, and $RB \geq 10$ indicate slight, moderate, and large bias, respectively. To assess the closeness of the fit, the limit value of RMSEA

$$RMSEA = \left(\max \left[\frac{2 \sum_{i=1}^{m_U} \sum_{j=1}^{m_V} \pi_{ij,(\mathrm{F})} \log \left(\pi_{ij,(\mathrm{F})} / \pi_{ij,(\mathrm{H})} \right)}{m_U m_V - m_U - m_V}, 0 \right] \right)^{1/2},$$

is computed. Owing to space limitations, here only some main results are presented and discussed in this subsection. Complete results can be found in the supplementary materials.

3.3.1. Experiment 1 Figures 1 displays the RB and RMSEA values when the true correlation is 0.4. As expected, assuming a wrong underlying distribution generally biases the polychoric correlation. Observe that the skew-normal distribution contains the normal distribution as a special case. Thus, both the normal and skew-normal assumptions consistently estimate the polychoric correlation when the true underlying distribution is normal. When the true underlying distribution is a normal distribution or a t distribution, all distributional assumptions produce a low RB (less than 5 %). When the true underlying distribution is a uniform distribution or logistic distribution, the normal assumption generally produces a low-biased correlation estimate. However, the t assumption may produce a high RB (Figure 1). The normal and skew-normal assumptions can produce moderately biased polychoric correlations when the underlying distribution is the exponential power distribution with $\beta = 0.3$ that corresponds to a distribution with high kurtosis (Figure 1). As the kurtosis in the exponential power family decreases, the magnitude of RB concomitantly decreases. When the underlying distribution is non-normal, the normal and skew-normal assumptions may produce different correlation estimates. Thus, the skew-normal distribution adjusts the underlying non-normality by introducing some degree of skewness. Consequently, the magnitude of RB may become higher but the RMSEA may become lower (Figure 1), which occurs when the number of categories is three for both ordinal variables. The polychoric correlation based on the underlying normal assumption generally underestimates the true correlation coefficient. The $t(4)$ and $t(6)$ assumptions sometimes outperform the normal assumption in Experiment 1.

3.3.2. Experiment 2 As expected, the polychoric correlation is consistently estimated when the true and assumed underlying distributions are both skew-normal (Figure 2). The normal assumption produces negatively biased correlation estimates. It can be moderately or strongly biased unless both α_1 and α_2 are small. Recall that α_1 and α_2 control the skewness and kurtosis of the underlying distribution. Small values of α_1 and α_2 only introduce a small departure from the bivariate normal distribution. All the t distribution assumptions produce similar RBs relative to the normal assumption. Under both the normal and t(υ) distribution assumptions, three categories in both ordinal variables generally lead to a higher magnitude of the RB value than five categories in both variables. For example, the RB with five-category variables does not exceed -15 when $\alpha_1 = 1$ and $\rho_0 = 0.4$, whereas the RB with three-category variables frequently exceeds -25 under the same condition (Figure 2). As the true value of the correlation increases while the other conditions remain the same, the RB generally becomes smaller (see Figures 7, 8, and 9 in the supplementary materials).

In Experiment 2, the RMSEA can be misleading when the number of categories is three in both variables. Consider the normal assumption as an example. The magnitude of RB may exceed 10 when $\alpha_1 = 0.1$, $\rho_0 = 0.4$, and both ordinal variables have three categories (Figure 2), whereas

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

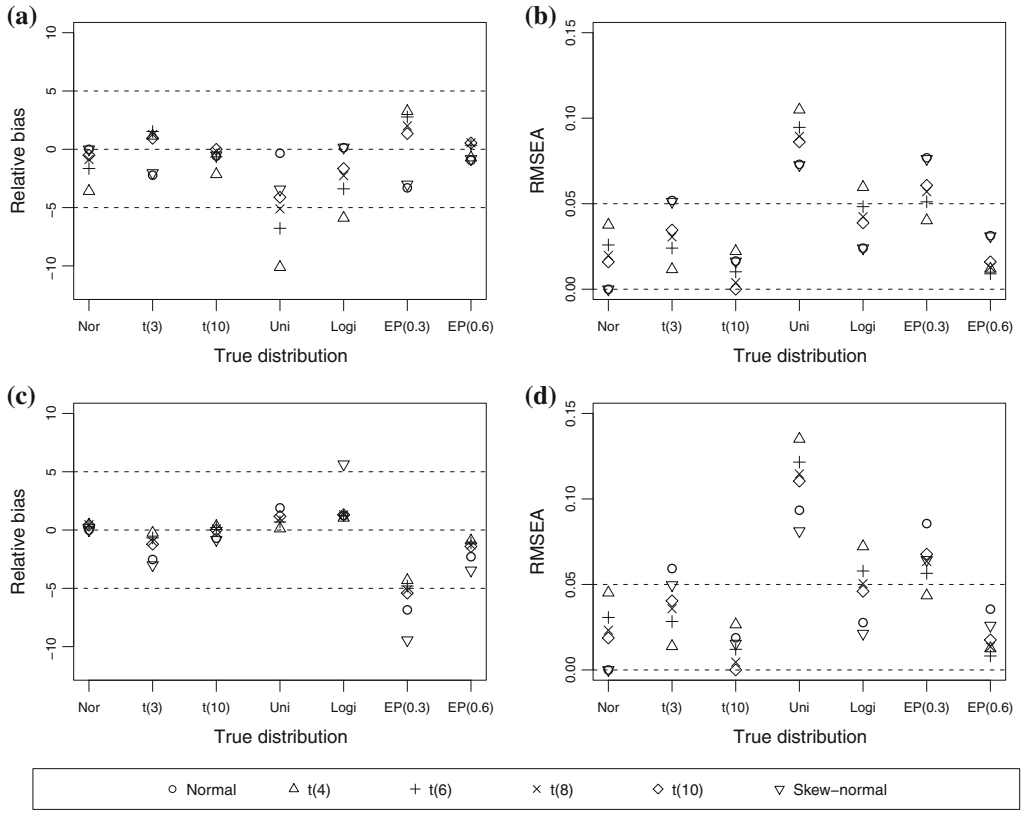


FIGURE 1.
Relative bias (RB) and root-mean-square error of approximation (RMSEA) of correlation estimates when the true underlying distribution belongs to the elliptical distribution family. The true correlation coefficient is 0.4, **a** RB when both ordinal variables have five categories. **b** RMSEA when both ordinal variables have five categories. **c** RB when both ordinal variables have three categories. **d** RMSEA when both ordinal variables have three categories. *Note Nor* normal, *Uni* uniform, *Logi* logistic, EP(·)=exponential power distribution with the enclosed value of β.

the RMSEA is still below 0.05 (Figure 3). The pattern is more dramatic when $\alpha_1 = 1$. The RB is almost -20 when $\alpha_2 = 1.5$ and $\rho_0 = 0.4$, but the RMSEA is slightly below 0.05. Thus, the estimated probabilities can be rather close to the true probabilities but the polychoric correlation can be largely biased. This event occurs because RMSEA only measures the closeness between the estimated and true category probabilities, and is not a direct measure of the correlation estimate.

On the other hand, although the skew-normal assumption consistently estimates the polychoric correlation in Experiment 2, the numerical difficulties (such as non-convergence and local maximizer) are encountered in the present study. The fit function $L(\boldsymbol{\theta})$ can be fairly flat (see Figure 13 in the supplementary materials as an illustration). A bad choice of the starting value for the numerical optimization process can lead to the aforementioned issues. Thus, 20 starting values are employed. As a result, the skew-normal assumption is computationally much more intensive than the normal assumption.

3.3.3. Experiment 3 When the true underlying distribution is a skew-t(4) distribution, the normal and t(v) underlying distributional assumptions lead to a largely biased polychoric correlation, except when both α_1 and α_2 are small (Figure 4). A small pair of (α_1, α_2) only introduces a small skewness and kurtosis to the underlying distribution, which is similar to a t(4) distribution.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

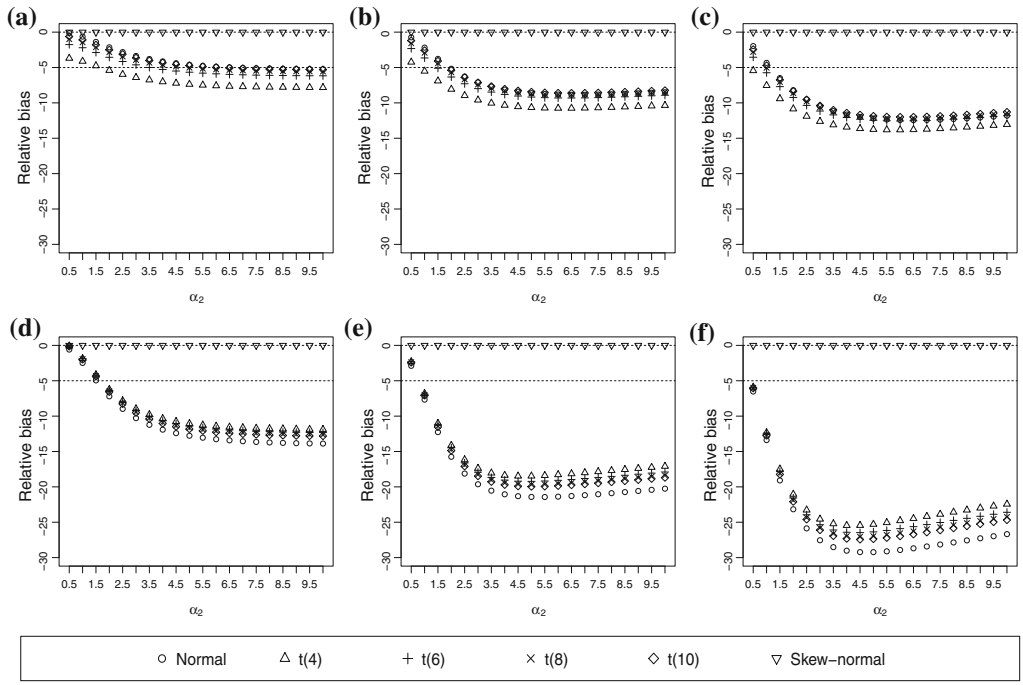


FIGURE 2.
Relative bias (RB) of correlation estimates when the true underlying distribution is skew-normal. The true correlation coefficient is 0.4. **a** $\alpha_1 = 0.1$ and both ordinal variables have five categories. **b** $\alpha_1 = 0.5$ and both ordinal variables have five categories. **c** $\alpha_1 = 1$ and both ordinal variables have five categories. **d** $\alpha_1 = 0.1$ and both ordinal variables have three categories. **e** $\alpha_1 = 0.5$ and both ordinal variables have three categories. **f** $\alpha_1 = 1$ and both ordinal variables have three categories.

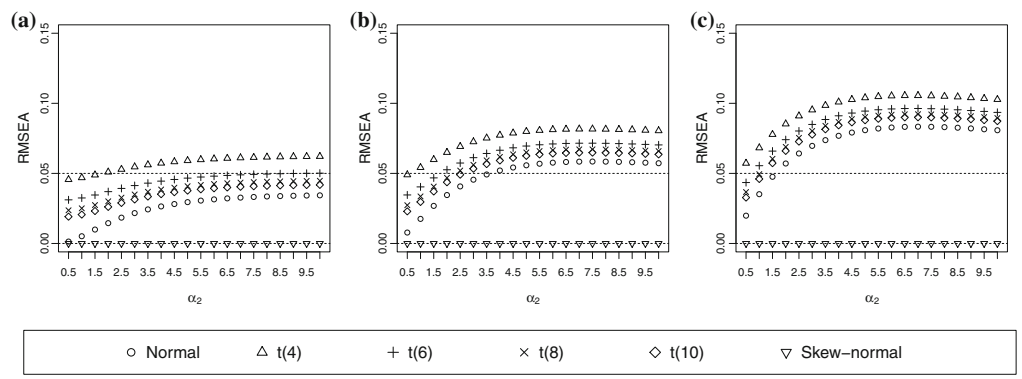


FIGURE 3.
Root-mean-square error of approximation (RMSEA) of correlation estimates when the true underlying distribution is skew-normal. The true correlation coefficient is 0.4. Both ordinal variables have three categories. **a** $\alpha_1 = 0.1$. **b** $\alpha_1 = 0.5$. (c) $\alpha_1 = 1$.

As known from Experiment 1, the normal and t(v) underlying distributional assumptions are only slightly biased when the true underlying distribution is a t distribution. The skew-normal assumption may produce not so biased correlations when both ordinal variables have three categories and α_1 is small (Figure 4). In general, the skew-normal assumption is less biased than the normal and

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

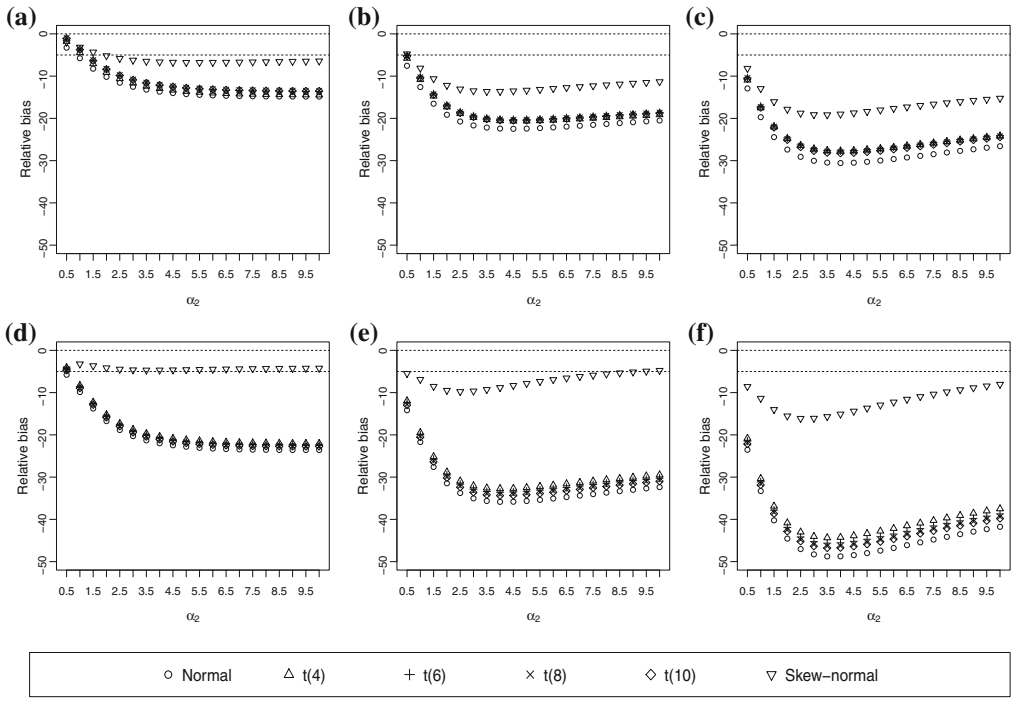


FIGURE 4.
Relative bias (RB) of correlation estimates when the true underlying distribution is skew-t(4). The true correlation coefficient is 0.4. **a** $\alpha_1 = 0.1$ and both ordinal variables have five categories. **b** $\alpha_1 = 0.5$ and both ordinal variables have five categories. **c** $\alpha_1 = 1$ and both ordinal variables have five categories. **d** $\alpha_1 = 0.1$ and both ordinal variables have three categories. **e** $\alpha_1 = 0.5$ and both ordinal variables have three categories. **f** $\alpha_1 = 1$ and both ordinal variables have three categories.

$t(v)$ assumptions. As the degrees of freedom of the skew-$t(v)$ distribution increases, all distributional assumptions become less biased, and the skew-normal assumption in particular is often robust (See the figures in the supplementary materials). This effect is expected from the fact that the skew-normal distribution corresponds to the skew-$t(\infty)$ distribution. Nevertheless, the normal and t(v) assumptions still produce moderately or largely biased polychoric correlations. Similar to the conclusions from the underlying skew-normal distribution, a wrong distributional assumption tends to underestimate the polychoric correlation. Three categories in both ordinal variables generally lead to a higher RB in magnitudes than five categories in both ordinal variables; and a higher value of the true correlation coefficient generally leads to less biased estimates. Similar to the case in Experiment 2, the RMSEA can be misleading as well. A low RMSEA does not necessarily indicate a low RB (e.g. see Figure 26 in the supplementary materials).

3.3.4. Experiment 4 Table 1 shows that all the underlying distributional assumptions tend to be extremely biased when the true underlying distribution is a Pareto distribution. Similar to Experiments 2 and 3, the polychoric correlation tends to be underestimated across all conditions in Experiment 4. The skew-normal assumption produces a lower RB than the normal and t(v) assumptions, although all assumptions generally produce a large RB. The value of RMSEA tends to be small despite the heavily biased polychoric correlation. In particular, the RMSEA produced by the skew-normal assumption is always low. Note that the Pareto distribution is skewed. Thus, the skew-normal distribution assumption mimics the skewed pattern, although the true correlation coefficient is inconsistently estimated.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

TABLE 1.
Relative bias (RB) and root-mean-squared error of approximation (RMSEA) of polychoric correlations in Experiment 4.

m_U	m_V	ρ_0	Assumed distribution					
			Normal	$t(4)$	$t(6)$	$t(8)$	$t(10)$	Skew-normal
RB								
3	3	0.4	−47.18	−45.98	−46.29	−46.47	−46.59	−32.89
		0.6	−51.23	−50.17	−50.42	−50.57	−50.68	−38.78
3	5	0.4	−38.23	−37.83	−37.60	−37.60	−37.65	−32.12
		0.6	−43.50	−43.12	−42.92	−42.93	−42.97	−37.99
5	5	0.4	−36.26	−38.19	−36.94	−36.48	−36.28	−31.60
		0.6	−41.96	−43.35	−42.36	−42.01	−41.86	−37.41
RMSEA								
3	3	0.4	0.04	0.07	0.05	0.05	0.04	0.00
		0.6	0.05	0.08	0.06	0.06	0.06	0.01
3	5	0.4	0.04	0.05	0.04	0.04	0.04	0.01
		0.6	0.05	0.06	0.05	0.05	0.05	0.01
5	5	0.4	0.03	0.04	0.04	0.03	0.03	0.01
		0.6	0.04	0.05	0.04	0.04	0.04	0.01

3.4. Asymptotic Variance

In this subsection, the asymptotic variance is illustrated in Figure 5 when the true underlying distribution is a skew-normal distribution and both ordinal variables have five categories. The skew-normal assumption produces a lower asymptotic variance than do the other assumptions of distribution. The normal assumption often produces a similar asymptotic variance to the t assumption when $\rho_0 = 0.4$. Otherwise, the normal assumption tends to be slightly less variable than the t assumption. However, Figure 5 shows that the asymptotic variances under the skew-normal assumption can be substantially higher than the asymptotic variances of other assumptions of distribution when both ordinal variables have three categories. Recall that the normal assumption is asymptotically biased (Figure 2); however, a lower variance may lead to a lower mean squared error than the skew-normal assumption. Thus, although the skew-normal assumption is asymptotically unbiased, the correlation estimate is likely to have a larger departure from the true value than the normal assumption because of the large variation.

4. Conclusion and Discussion

In this paper, we study robustness of polychoric correlation estimation against misspecification of underlying distributions. The asymptotic polychoric correlation and its asymptotic (co)variance are derived under the conditions of the support of assumed distributions. Unlike the continuous case, the correlation structure is not asymptotically unbiased any more. Although the bias is sometimes small, a large bias can occur, especially when the true underlying distribution is skewed but a bivariate normal or t distribution is assumed. It is seen from the numerical example that the skew-normal assumption performs as well as the conventional normal assumption when the true underlying distribution is a t distribution and improves the normal assumption when skewness exists.

Both Flora and Curran (2004) and Quiroga (1992) found that the normal assumption is robust against non-normal data generated from the Fleishman–Vale–Maurelli method. For example,

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

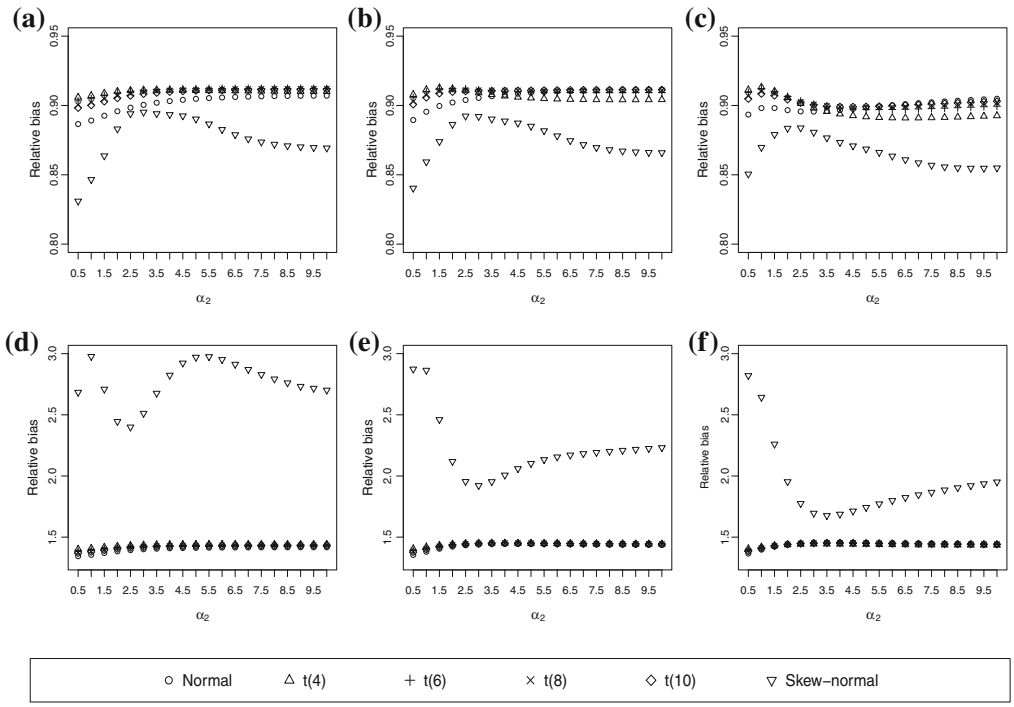


FIGURE 5.
Asymptotic variances of correlation estimators when the true underlying distribution is skew-normal. The true correlation coefficient is 0.4. **a** $\alpha_1 = 0.1$ and both ordinal variables have five categories. **b** $\alpha_1 = 0.5$ and both ordinal variables have five categories. **c** $\alpha_1 = 1$ and both ordinal variables have five categories. **d** $\alpha_1 = 0.1$ and both ordinal variables have three categories. **e** $\alpha_1 = 0.5$ and both ordinal variables have three categories. **f** $\alpha_1 = 1$ and both ordinal variables have three categories.

the largest skewness and kurtosis considered in Flora and Curran (2004) are 1.25 and 3.75, respectively. The RB is lower than 10 in most conditions and is lower than 5 when the number of categories is five and $\rho_0 = 0.49$ (Flora & Curran, 2004, Table 2). Our results show that the polychoric correlation can be largely underestimated using the normal assumption when the true underlying distribution is a skew-normal distribution skewness and kurtosis of which are bounded by some small values. The bias becomes even higher when the true underlying distribution is skew-$t(4)$ or a Pareto distribution in which cases the kurtosis is not well defined. Although the skew-normal assumption is also largely biased sometimes, it greatly improves the conventional normal assumption. Still, the skew-normal assumption has a much higher variance than the normal assumption when the number of categories is small. Thus, the volatility is high under the skew-normal assumption. Obviously, more studies are needed to investigate small sample volatility in order to provide suggestions for practice.

Lee and Lam (1988) suggested using the correct underlying distributional assumption to estimate more accurately the polychoric correlation if the ordinal data are asymmetric. Because the ordinal data indicate the loss of information when comparing with continuous data, we cannot have visual inspections of the underlying distribution. If the tests of the underlying distribution were rejected, the underlying distributional assumption is questionable, and an alternative distributional assumption should be used. In practice, several assumptions of underlying distribution can be tested and then the most plausible one chosen.

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

The normal distribution is a special case of the skew-normal distribution. We have shown that both distributions consistently estimate the polychoric correlation when the true distribution is normal. Thus, the skew-normal assumption, which is able to model skewness and kurtosis, is a natural extension to the conventional normal assumption and frequently outperforms the normal assumption. However, three parameters are simultaneously estimated in the skew-normal distribution. Because the thresholds are determined through α_1, α_2, and ω, the gradient and Hessian matrix involve derivatives of the thresholds with respect to α_1, α_2, and ω. Accordingly, it is computationally more difficult than the normal assumption. Besides, non-convergence and local optimizers are encountered in the present study and multiple starting values are used to obtain the correlation estimate.

Although only the t and skew-normal assumptions are illustrated as non-normal alternatives in the present study, other distributions that are differentiable with respect to unknown parameters can be used to estimate the correlation coefficient by the aid of Theorem 1 or Eq. (6). Its asymptotic variance and covariance can be estimated using Theorem 2 or Eq. (7). For example, the logistic distribution can be assumed in the two-step estimation and the skew-*t* distribution can be assumed in the variant of the two-step estimation. It will be of interest to derive analytical expressions for the skew-elliptical distribution family that consists of the skew-normal and skew-*t* distributions. Our numerical results demonstrate that the skew-normal assumption generally improves the conventional normal assumption in the imaginary case where n is infinite. It is worthy to conduct a simulation study to investigate the small sample bias in estimating the correlation coefficient and its effects on the bias of parameters in a SEM with ordinal data.

Acknowledgments

The research reported in this article has been supported by the Swedish Research Council (VR) under the program: Structural Equation Modeling with Ordinal Variables, 421-2011-1727.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Arnold, B. C. (2008). Pareto and generalized pareto distributions. In D. Chotikapanich (Ed.), *Modeling income distributions and Lorenz curves* (pp. 119–145). New York: Springer.

Azevedo, C. L. N., Bolfarine, H., & Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Computational Statistics & Data Analysis*, *55*, 353–365.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, *32*, 159–188.

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 367–389.

Azzalini, A., & Capitanio, A. (2014). *The skew-normal and related families* (Vol. 3)., Institute of Mathematical Statistics Monographs Cambridge: Cambridge University Press.

Azzalini, A., & Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.

Balakrishnan, N., & Lai, C. D. (2009). *Continuous bivariate distributions* (2nd ed.). New York, NY: Springer.

Bazán, J. L., Branco, M. D., & Bolfarine, H. (2006). A skew item response model. *Bayesian Analysis*, *1*, 861–892.

Berkane, M., Kano, Y., & Bentler, P. M. (1994). Pseudo maximum likelihood estimation in elliptical theory: Effects of misspecification. *Computational Statistics & Data Analysis*, *18*, 255–267.

Bolfarine, H., & Bazán, J. L. (2010). Bayesian estimation of the logistic positive exponent irt model. *Journal of Educational and Behavioral Statistics*, *35*, 693–713.

Carolina Population Center. (2009). *National Longitudinal Study of Adolescent to Adult Health (Add Health) [Data file and code book]*. http://www.cpc.unc.edu/projects/addhealth

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press

Chateau, D., Metge, C., Prior, H., & Soodeen, R. A. (2012). Learning from the census: The Socio-economic Factor Index (SEFI) and health outcomes in Manitoba. *Canadian Journal of Public Health*, *4*, 23–27.
Christoffersson, A., & Gunsjö, A. (1996). A short note on the estimation of the asymptotic covariance matrix for polychoric correlations. *Psychometrika*, *61*, 173–175.
Fang, K. T., Kotz, S., & Ng, K. W. (1990). *Symmetric multivariate and related distributions*. New York, NY: Chapman and Hall.
Ferguson, T. S. (1996). *A course in large sample theory*. New York, NY: Chapman and Hall.
Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521–532.
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*, 446–491.
Gómez, E., Gómez-villegas, M. A., & Marín, J. M. (2003). An survey on continuous elliptical vector distributions. *Revista Matemática Complutense*, *16*, 345–361.
Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York, NY: Wiley.
Hodge, R. W., & Treiman, D. J. (1968). Social participation and social status. *American Sociological Review*, *33*, 722–740.
Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.
Jöreskog, K. G., & Sörbom, D. (1996). *Lisrel 8: User's reference guide*. Chicago, IL: Scientific Software International.
Kano, Y., Berkane, M., & Bentler, P. M. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations. *Journal of the American Statistical Association*, *88*, 135–143.
Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, *32*, 419–430.
Kullback, S. (1959). *Information theory and statistics*. New York, NY: Wiley.
Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.
Lee, S. Y., & Lam, M. L. (1988). Estimation of polychoric correlation with elliptical latent variables. *Journal of Statistical Computation and Simulation*, *30*, 173–188.
Lucke, J. F. (2014). Positive trait item response models. In R. E. Millsap, L. A. van der Ark, D. M. Bolt & C. M. Woods (Eds.), *New Developments in Quantitative Psychology: Presentations from the 77th Annual Psychometric Society Meeting* (Vol. 66, pp. 199–213). New York: Springer.
Mardia, K. V. (1962). Multivariate pareto distributions. *The Annals of Mathematical Statistics*, *33*, 1008–1015.
Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*, *71*, 57–77.
Maydeu-Olivares, A., Forero, C. A., Gallardo-Pujol, D., & Renom, J. (2009). Testing categorized bivariate normality with two-stage polychoric correlation estimates. *Methodology*, *5*, 131–136.
Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2×n contingency tables: A unified framework. *Journal of the American Statistical Association*, *100*, 1009–1020.
Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*, 713–732.
Molenaar, D. (2015). Heteroscedastic latent trait models for dichotomous data. *Psychometrika*, *80*, 625–644.
Molenaar, D., Dolan, C. V., & de Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: Testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika*, *77*, 455–478.
Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460.
Quiroga, A. M. (1992). *Studies of the polychoric correlation and other correlation measures for ordinal variables*. Unpublished Doctoral dissertation, Uppsala University, Uppsala.
Santos, J. R. S., Azevedo, C. L. N., & Bolfarine, H. (2013). A multiple group item response theory model with centered skew-normal latent trait distributions under a bayesian framework. *Journal of Applied Statistics*, *40*, 2129–2149.
Scharoun-Lee, M., Adair, L. S., Kaufman, J. S., & Gordon-Larsen, P. (2009). Obesity, race/ethnicity and the multiple dimensions of socioeconomic status during the transition to adulthood: A factor analysis approach. *Social Science & Medicine*, *68*, 708–716.
Scheier, F. M., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*, 219–247.
Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 1–25.
Woods, C. M., & Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika*, *71*, 281–301.

Manuscript Received: 6 APR 2014
Final Version Received: 17 MAY 2016
Published Online Date: 22 SEP 2016

https://doi.org/10.1007/s11336-016-9512-2 Published online by Cambridge University Press