Bivariate Distributions Underlying Responses to Ordinal Variables

Laura Kolbe *, Frans Oort and Suzanne Jak

Faculty of Social and Behavioural Sciences, University of Amsterdam, 1012 WX Amsterdam, The Netherlands;
F.J.Oort@uva.nl (F.O.); S.Jak@uva.nl (S.J.)
* Correspondence: L.Kolbe@uva.nl

Abstract: The association between two ordinal variables can be expressed with a polychoric correlation coefficient. This coefficient is conventionally based on the assumption that responses to ordinal variables are generated by two underlying continuous latent variables with a bivariate normal distribution. When the underlying bivariate normality assumption is violated, the estimated polychoric correlation coefficient may be biased. In such a case, we may consider other distributions. In this paper, we aimed to provide an illustration of fitting various bivariate distributions to empirical ordinal data and examining how estimates of the polychoric correlation may vary under different distributional assumptions. Results suggested that the bivariate normal and skew-normal distributions rarely hold in the empirical datasets. In contrast, mixtures of bivariate normal distributions were often not rejected.

Keywords: polychoric correlation; ordinal data; underlying bivariate normality; mixture of bivariate normal distributions

1. Introduction

Data in the social and behavioral sciences commonly include observations from variables with ordinal response scales, such as Likert items. If ordinal variables do not possess metric properties, alternative techniques than those employed with continuous variables are required. The polychoric correlation coefficient proposed by Pearson [1] is a recommended measure of association between two ordinal variables. The polychoric correlation coefficient is based on the assumption that responses to ordinal variables are generated by two latent underlying continuous variables. The underlying variables are conventionally assumed to follow a bivariate normal distribution, also referred to as underlying bivariate normality. In the present study, this assumption will be addressed.

In simulation studies, the polychoric correlation coefficient has been compared to other measures of association, including the product-moment correlation, Spearman's rank correlation, and Kendall's tau coefficient [2,3]. These studies showed that if the underlying bivariate normality held, the polychoric correlation coefficient was generally closer to the correlation between the two underlying continuous variables than other measures of association. However, bivariate normality of the underlying continuous variables has been considered unrealistic [4,5]. Indeed, experience with empirical data suggests that the underlying bivariate normality assumption may be questionable [6–11]. These findings give rise to the question whether alternative distributions may represent features of the underlying continuous latent variables more accurately.

Some studies suggested that the polychoric correlation coefficient is fairly robust against small to moderate departures from underlying bivariate normality [12–14]. However, Grønneberg and Foldnes [15] explained that these simulation studies used a generation method of the ordinal data that is equivalent to discretizing normal data. In recent evaluations using simulated data that are not compatible with underlying normality, it was found that the polychoric correlation coefficient is highly sensitive to underlying

non-normality [16–18]. Similar results have been found by Jin and Yang-Wallentin [19], who examined the robustness of the polychoric correlation against non-normality with data generated from skew-normal [20], skew-t (v) [21], and Pareto [22] distributions.

In accordance with Muthén and Hofacker [8], Jin and Yang-Wallentin [19] suggested that the estimate of the polychoric correlation should only be used if the underlying bivariate normality assumption holds. When the assumption of underlying bivariate normality does not hold, the polychoric correlation coefficient can be based on other distributional assumptions that represent features of the data more accurately than the bivariate normal distribution [6]. Several alternative underlying bivariate distributions have been proposed already, among which are Azzalini and Dalla Valle's skew-normal distribution [19,20,23] and the mixture of normal distributions [24]. Previous research indeed shows that the polychoric correlation coefficient provides an accurate estimate of the correlation between the two underlying continuous variables when the correct underlying distribution is assumed [10,19,23]. In a study by Roscino and Pollice [23], a polychoric correlation coefficient was introduced based on the hypothesis that the underlying latent variables follow Azzalini and Dalla Valle's bivariate skew-normal distribution. The polychoric correlation based on this distribution yielded better estimates of the correlation between the underlying continuous latent variables than the original polychoric correlation when the sample size was large, or the number of categories of the ordinal variables was small, or the skewness parameters were discordant [23]. More recently, Jin and Yang-Wallentin [19] examined the performance of several generalizations in an extensive simulation study, including Azzalini and Dalla Valle's skew-normal distribution. In line with Roscino and Pollice [23], their results suggest that assuming an underlying skew-normal distribution generally produces lower bias in the polychoric correlation estimate than assuming an underlying bivariate normal distribution.

In addition to the skew-normal distribution, the mixture of normal distributions has been considered for the estimation of the polychoric correlation. Uebersax and Grove [25] proposed a latent mixture model in which the distribution of a latent trait measured by an ordinal variable is defined as a combination of two subgroups' probability density functions. Although their proposed model was presented as an approach to analyze rating agreement, it could also be applied for modeling mixtures with observed ordinal variables [24]. That is, a polychoric correlation coefficient can be based on the assumption that the underlying continuous latent variables follow a mixture of two or more normal distributions contingent on the idea that data have been gathered from two or more subpopulations. Because the means and variances of each subgroup's normal distribution are allowed to differ, the mixture distribution can be a flexible tool to account for heterogeneous and asymmetric data. This distribution may therefore be suited to accurately reflect the features of the underlying continuous latent variables. The polychoric correlation based on an underlying mixture of normal distributions has not been studied yet with empirical or simulated data.

Although it is clear that the polychoric correlation coefficient can be accurately estimated as long as the underlying distribution giving rise to the observed ordinal responses is known, it is impossible to identify the correct underlying distribution for empirical data. The Ref. [26] showed that with two binary variables and underlying non-normality, there can exist a very wide range of tetrachoric correlations that are consistent with the observed data, depending on what distribution is assumed for the underlying latent variables. With more than two response options per variable, and by using substantive knowledge to add restrictions to the underlying distributions, the range of possible correlations will get smaller and eventually converge to an identified case. In practice, it can still be informative to test which distributions are consistent with the observed data and which are not, so that some distributions may be ruled out.

The Present Study

The aim of the present study was to examine the fit of the bivariate normal distribution, bivariate skew-normal distribution, and mixture of bivariate normal distributions to a

large number of pairs of ordinal variables from empirical datasets. Knowledge of the fit of the proposed underlying distributions to empirical data may contribute in several ways. First, the results of this study may help to determine the degree of applicability of the fitted distributions. Second, the results could explicate how to generate more realistic data in future simulation studies, including ordinal variables, to increase generalizability. Third, we examined how estimates of the polychoric correlation vary under different distributional assumptions in empirical data. Fourth, we provide R-syntax that other researchers can adapt to fit the distributions to ordinal data.

The remainder of the present paper is organized as follows. The polychoric correlation assuming underlying bivariate normality is described first. The bivariate skew-normal distribution [20] and mixture of bivariate normal distributions distribution [27] are then introduced as alternatives to the bivariate normal distribution for the latent continuous variables underlying the observed ordered variables. Subsequently, we illustrate estimations of the polychoric correlation on the basis of the bivariate normal distribution, the bivariate skew-normal distribution, and the mixture of bivariate normal distributions with an empirical example. This illustration may serve as a guideline for researchers on how to test the underlying distributions when estimating the polychoric correlation between two observed ordinal variables. We then present the results of a more extensive study on the fit of these distributions to a large number of contingency tables, showing which distributions are rejected most often in real data. We explicitly focus on the bivariate case in this article, and we discuss issues around the multivariate case in the discussion.

2. The Polychoric Correlation Assuming Different Underlying Distributions

2.1. The Bivariate Normal Distribution

Consider two observed ordinal variables X_1 and X_2 with response categories $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$. The polychoric correlation coefficient assumes that the responses to X_1 and X_2 are generated by latent underlying continuous variables ξ_1 and ξ_2. The relationship between observed ordinal X_1 and X_2 and underlying continuous ξ_1 and ξ_2 may be written as

$$
\begin{aligned}
X_1 = i &\Leftrightarrow \tau_{i-1}^{(1)} < \xi_1 \leq \tau_i^{(1)} \\
X_2 = j &\Leftrightarrow \tau_{j-1}^{(2)} < \xi_2 \leq \tau_j^{(2)},
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\tau}^{(1)}$ and $\boldsymbol{\tau}^{(2)}$ are the thresholds parameters for X_1 and X_2, respectively. The thresholds represent the bounds of the response categories such that $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{I-1} < \tau_I = \infty$. An item with I categories has $I-1$ threshold parameters.

The maximum likelihood estimate of the polychoric correlation between X_1 and X_2 is the value of ρ that minimizes

$$
G(\gamma) = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \ln[p_{ij}/\pi_{ij}(\gamma)],
\tag{2}
$$

where p_{ij} is the observed proportion and $\pi_{ij}(\gamma)$ is the expected proportion for $X_1 = i$ and $X_2 = j$. The expected proportion may be written as

$$
\pi_{ij}(\gamma) = \int_{\tau_{i-1}^{(1)}}^{\tau_i^{(1)}} \int_{\tau_{j-1}^{(2)}}^{\tau_j^{(2)}} f(\xi_1, \xi_2 | \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\xi_1 d\xi_2,
\tag{3}
$$

where $f(\cdot)$ is a bivariate normal distribution with means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Because the underlying continuous latent variables are not directly observed, their means and variances are usually fixed at zero and one, respectively. Alternatively, the location and scale of the underlying variable could be identified by fixing two thresholds. There exist two approaches to estimating the polychoric correlation coefficient [28]. In the two-step

approach, the threshold parameters are estimated from univariate information first, and the other parameters are estimated in a second step while fixing the thresholds to the values obtained in the first step. Another approach is to estimate all parameters simultaneously. Throughout this manuscript, we use the latter approach.

2.2. The Skew-Normal Distribution

The univariate skew-normal distribution was proposed by Azzalini [29] and extended to the multivariate skew-normal distribution by Azzalini and Dalla Valle [20]. The skew-normal distribution is a natural generalization of the normal distribution with extra shape parameters $\boldsymbol{\alpha}$ that regulate skewness and kurtosis. The bivariate skew-normal distribution involves two shape parameters α_1 and α_2, and simplifies to the bivariate normal distribution when $\alpha_1 = \alpha_2 = 0$. With larger absolute values of α, the skewness and kurtosis of the distribution increase. The distribution is right-skewed and leptokurtic if $\alpha > 0$, and left-skewed and leptokurtic if $\alpha < 0$. According to Azzalini and Dalla Valle [20], the skew-normal distribution is reasonably flexible with regard to empirical data fitting, and maintains some convenient formal properties of the normal distribution. An additional advantage of the bivariate skew-normal distribution is that its marginal distributions are skew-normal.

The density function of the bivariate skew-normal distribution for the underlying continuous latent variables ξ_1 and ξ_2 is given by

$$g(\xi_1, \xi_2) = 2\phi(\xi_1, \xi_2; \omega)\Phi(\alpha_1\xi_1 + \alpha_2\xi_2), \tag{4}$$

where $\phi(\cdot, \cdot; \omega)$ is the density function of the standard normal distribution with correlation ω, and $\Phi(\cdot)$ is the standard normal distribution function. Under the underlying bivariate skew-normal assumption, the expected probability $\pi_{ij}(\boldsymbol{\gamma}) = P(X = i, Y = j)$ may be written as

$$\pi_{ij}(\boldsymbol{\gamma}) = \int_{\tau_{i-1}^{(1)}}^{\tau_i^{(1)}} \int_{\tau_{j-1}^{(2)}}^{\tau_j^{(2)}} g(\xi_1, \xi_2 | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \omega, \boldsymbol{\alpha})d\xi_1 d\xi_2, \tag{5}$$

where $g(\cdot)$ denotes the bivariate skew-normal distribution. The polychoric correlation coefficient under the bivariate skew-normal distribution can be obtained by

$$\rho = \frac{\omega - 2\pi^{-1}\delta_1\delta_2}{((1 - 2\pi^{-1}\delta_1^2)(1 - 2\pi^{-1}\delta_2^2))^{1/2}}, \tag{6}$$

where π is the actual number π. The parameters δ_1 and δ_2 can be computed from α_1 α_2 and ω by

$$\begin{aligned}
\alpha_1 &= \frac{\delta_1 - \delta_2\omega}{((1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega))^{1/2}} \\
\alpha_2 &= \frac{\delta_2 - \delta_1\omega}{((1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega))^{1/2}},
\end{aligned} \tag{7}$$

and vary in $(-1, 1)$.

2.3. Mixture of Normal Distributions

The mixture of normal distributions was first proposed by Pearson [27]. The mixture of normal distributions decomposes the population into a set of subpopulations, often called components. The mixture distribution is composed of a weighted sum of each subpopulation's normal distribution with means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. An advantage of the mixture of normal distributions is that great flexibility can be achieved with only a few subpopulations [25]. In the present study, we only considered mixtures of two subpopulations.

The density function of the mixture of bivariate normal distributions for the underlying continuous latent variables ξ_1 and ξ_2 with two subpopulations can be written as

$$h(\xi_1, \xi_2) = \lambda \phi(\xi_1, \xi_2; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + (1 - \lambda) \phi(\xi_1, \xi_2; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*), \tag{8}$$

where $0 < \lambda < 1$ is the component probability, that is, the prevalence of the subpopulation in which the underlying continuous latent variables ξ_1 and ξ_2 follow a bivariate normal distribution with means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In the subpopulation with a prevalence of $(1 - \lambda)$, the underlying continuous latent variables follow a bivariate normal distribution with means $\boldsymbol{\mu}^*$ and covariance matrix $\boldsymbol{\Sigma}^*$.

The expected proportion π_{ij} under the mixture of bivariate normal distributions in Equation (8) is

$$\pi_{ij}(\boldsymbol{\gamma}) = \int_{\tau_{i-1}^{(1)}}^{\tau_i^{(1)}} \int_{\tau_{j-1}^{(2)}}^{\tau_j^{(2)}} h(\xi_1, \xi_2 | \lambda, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) d\xi_1 d\xi_2, \tag{9}$$

where $h(\cdot)$ denotes the mixture distribution. For identification of the underlying continuous latent variables, the means and variances in the first subpopulation can be fixed at zero and unity, respectively, while constraining the thresholds to be equal across subpopulations. All other parameters can be freely estimated. The polychoric correlation in the first subpopulation is equal to the covariance σ_{12}. The polychoric correlation in the second subpopulation ρ^* is equal to the covariance σ_{12}^* divided by the product of the standard deviations of ξ_1 and ξ_2, $\sqrt{\text{diag}(\boldsymbol{\Sigma}^*)}$.

The number of parameters to be estimated for the mixture of bivariate normal distributions can be reduced by imposing restrictions on the polychoric correlations, means, or variances. One possibility is to restrict ρ and ρ^* to be equal. With this restriction, the association between ξ_1 and ξ_2 is the same in each subpopulation. Another approach to reduce the number of parameters is to fix ρ^* at zero, resulting in a mixture of bivariate normal distributions in which ξ_1 and ξ_2 are not associated in the second subpopulation. Additionally, one can restrict the means to be equal across the two subpopulations, $\boldsymbol{\mu} = \boldsymbol{\mu}^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. This restriction results in a mixture of bivariate normal distributions with a single mode. Another option is to restrict the variances of the underlying variables to be equal across the subpopulations, $\boldsymbol{\sigma} = \boldsymbol{\sigma}^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

2.4. Testing Underlying Distributions

The fit of underlying distributions on empirical data can be tested for a given pair of ordinal variables with the likelihood ratio test (LRT; [7]). The LRT statistic is given by

$$2N \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \ln[p_{ij} / \hat{\pi}_{ij}] = 2N G(\hat{\boldsymbol{\gamma}}), \tag{10}$$

where $\hat{\pi}_{ij}$ is the estimated expected proportion obtained from the tested distribution with estimated parameter vector $\hat{\boldsymbol{\gamma}}$. When the model holds, the LRT statistic is asymptotically chi-squared distributed with degrees of freedom equal to

$$df = I \times J - 1 - n, \tag{11}$$

where $I \times J$ is the number of response patterns and n is the number of estimated parameters [30].

3. Illustrative Example

Below, we briefly demonstrate estimating the polychoric correlation assuming an underlying bivariate normal distribution, underlying bivariate skew-normal distribution, and underlying mixture of bivariate normal distributions with an empirical example. The purpose of this illustration is to show how one can test whether an underlying distribution

may be suitable or not when estimating the polychoric correlation. The R scripts of the example are available in Appendixes A–C.

Example: Type D Personality Data

The Ref. [31] gathered observations from 541 respondents on the Type D personality, which can be described as a tendency to experience negative affectivity and social inhibition. The DS14 was used as an instrument to measure these two tendencies, and observations on Item 2 from the negative affectivity subscale and Item 6 from the social inhibition subscale are presented in a contingency table given in Table 1. Assuming underlying bivariate normality, the estimated polychoric correlation between negative affectivity and social inhibition was $\hat{\rho} = 0.29$. The bivariate normal distribution, however, did not fit the data at a 5% level of significance, $\chi^2(15) = 52.69$, $p < 0.001$. The bivariate skew-normal distribution with estimated correlation $\hat{\rho} = 0.29$ and shape parameters $\alpha_1 = 0.07$ and $\alpha_2 = 1.67$ did not fit the data either, $\chi^2(13) = 52.23$, $p < 0.001$. In addition, the increase in fit compared to the bivariate normal distribution was not significant, $\Delta\chi^2(2) = 0.50$, $p = 0.780$. Figure 1 shows the estimated underlying bivariate skew-normal distribution and its marginal distributions. The marginal distribution underlying the observed social inhibition was skewed to the right (see Figure 1b) as indicated by the positive shape parameter.

Table 1. Contingency table for the Type D personality data.

Negative Affectivity	Social Inhibition				
	False	Rather False	Neutral	Rather True	True
False	67	15	16	8	3
Rather false	41	28	30	4	2
Neutral	34	48	39	11	1
Rather true	35	22	34	28	5
True	24	10	11	11	9

(**a**) Univariate distribution of negative affectivity

(**b**) Univariate distribution of social inhibition

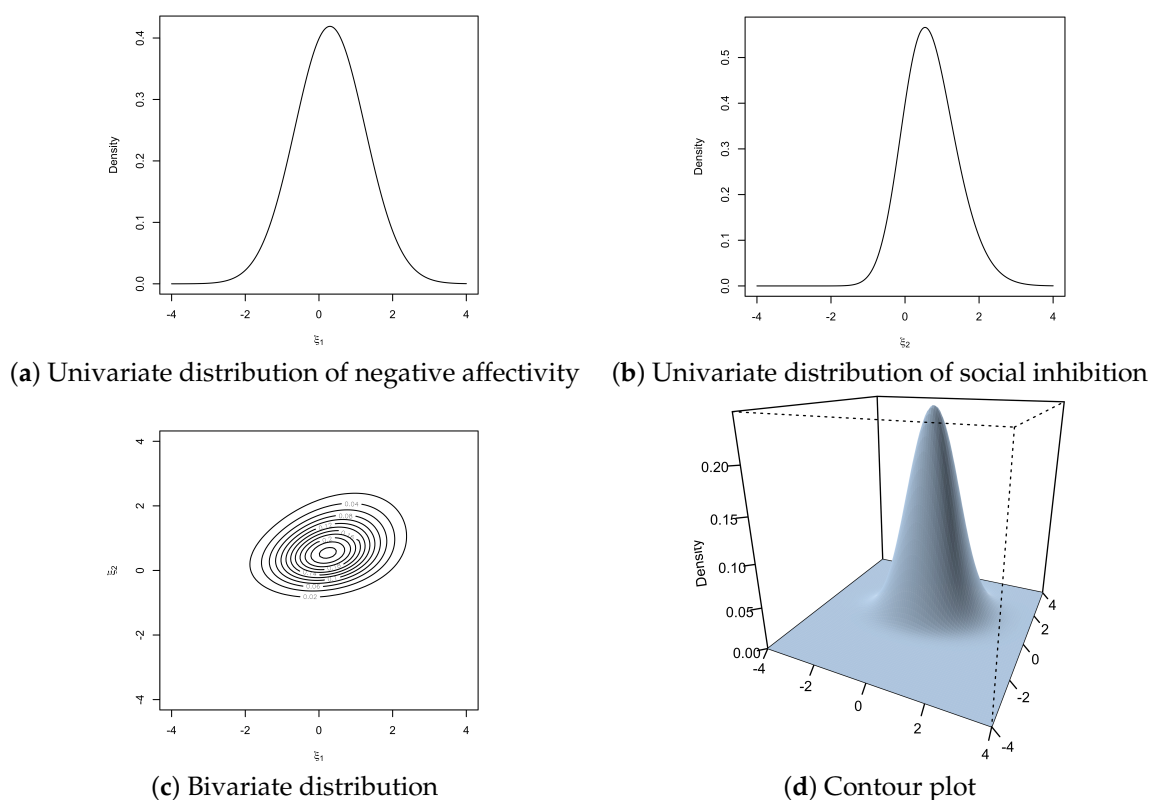
(**c**) Bivariate distribution

(**d**) Contour plot

Figure 1. The univariate distributions (**a**,**b**), bivariate (**c**) and contour plot (**d**) of the estimated mixture of bivariate normal distributions for the Type D personality data.

The mixture of bivariate normal distributions with freely estimated means, variances, and correlation in the second subsample was not rejected at a 5% level of significance, $\chi^2(9) = 9.36$, $p = 0.405$. The estimated component probability was $\hat{\lambda} = 0.50$, reflecting equally large subsamples. In the first subsample, the underlying continuous variables followed a bivariate standard normal distribution with means of 0, variances of 1, and a polychoric correlation of $\hat{\rho} = 0.85$. In the other subsample, the bivariate normal distribution of the underlying variables had means of $\hat{\boldsymbol{\mu}}^* = \begin{pmatrix} 0.34 \\ -0.10 \end{pmatrix}$, a covariance matrix of $\hat{\boldsymbol{\Sigma}}^* = \begin{pmatrix} 1.14 & -0.30 \\ -0.30 & 0.72 \end{pmatrix}$, and a polychoric correlation of $\hat{\rho}^* = -0.33$. This mixture of bivariate normal distributions fitted significantly better than a mixture distribution with the correlation in the second subsample fixed at zero, $\Delta\chi^2(1) = 4.88$, $p = 0.027$, indicating that the correlation in the second subsample significantly differs from zero.

The estimated mixture of bivariate normal distributions is illustrated in Figure 2. Figure 2a,b shows that the distribution contained a single mode, and Figure 2c,d shows that the correlation estimates in the subsamples had opposite signs. In half of the sample, negative affectivity was positively associated with social inhibition.

(**a**) Univariate distribution of negative affectivity   (**b**) Univariate distribution of social inhibition

(**c**) Bivariate distribution   (**d**) Contour plot

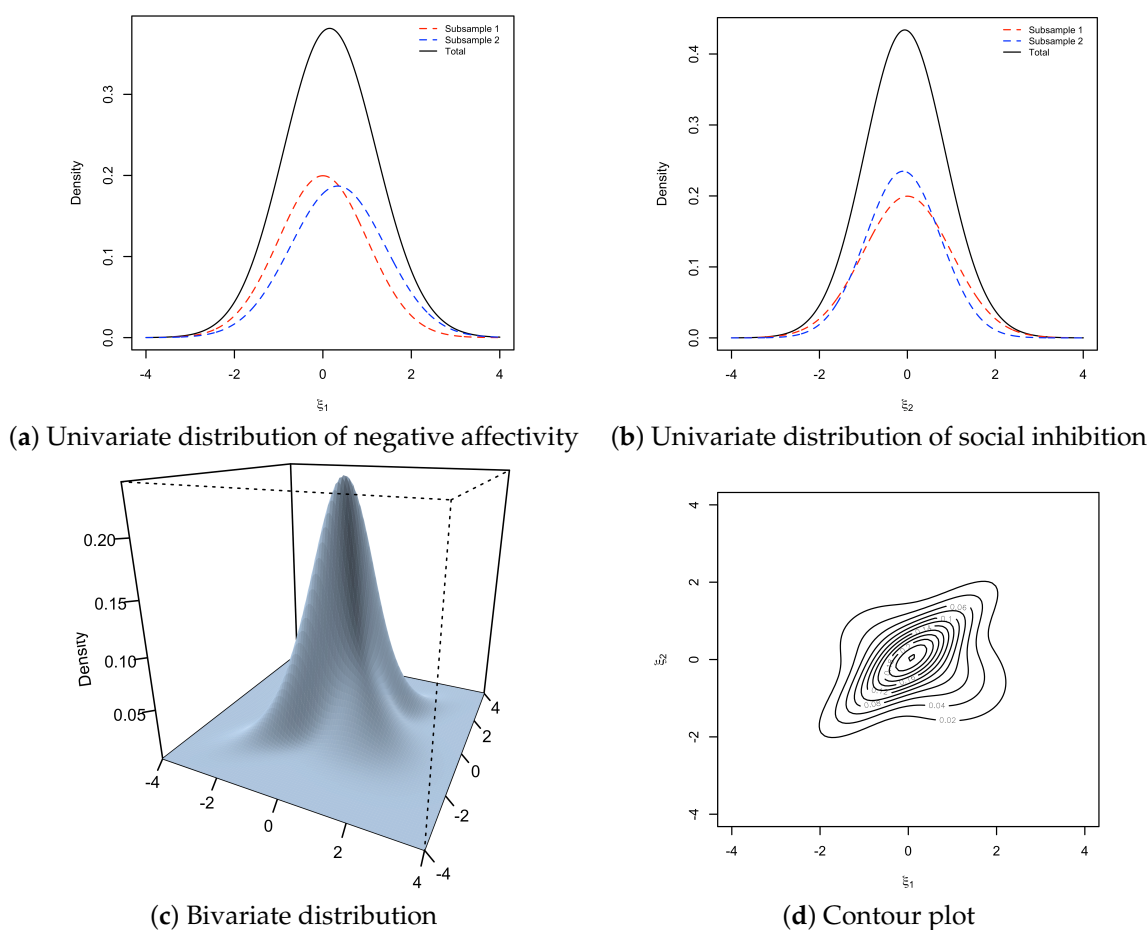



Figure 2. The univariate distributions (**a**,**b**), bivariate (**c**) and contour plot (**d**) of the estimated mixture of bivariate normal distributions for the Type D personality data.

4. Empirical Study on the Fit of Different Distributions

We now turn to our empirical study in which the fit of the bivariate normal distribution, bivariate skew-normal distribution, and the mixture of bivariate normal distributions was tested. These distributions were fitted to 700 contingency tables stemming from two datasets described below.

4.1. Empirical Data

4.1.1. Type D Personality Data

The first empirical dataset that was used in the present study was gathered by [31] to examine whether negative affectivity and social inhibition predict cardiovascular events in 541 patients with coronary artery disease. The sample consisted of 541 patients, of which 473 were male. Negative affectivity and social inhibition were measured using the DS14 [32]. The DS14 is a widely used instrument for the assessment of the Type D personality, which is described as a tendency towards negative affectivity (i.e., experiencing negative emotions) and social inhibition (i.e., experiencing social discomfort, reticence, and lack of social poise). The DS14 contains 14 items with five ordered response categories each (1 = *false*, 2 = *rather false*, 3 = *neutral*, 4 = *rather true*, 5 = *true*). Negative affectivity and social inhibition were both assessed with seven items. Two social inhibition items, items 1 and 3, were negatively worded and were therefore recoded in this study such that a higher score indicates a higher level of social inhibition. With 14 items, there are 91 pairs of items to be analyzed.

4.1.2. Health Status Data

The second dataset concerned a nationwide health status survey conducted by the Netherlands Organisation of Applied Scientific Research TNO [33]. The Dutch version [33] of the SF-36 health survey [34] was used to assess the health status of 1742 adults. The SF-36 health survey consists of 36 items with ordered response categories, organized into eight different aspects of health: physical functioning (PF; ten items with three response categories each), role limitations due to physical health (RP; four items with two response categories each), bodily pain (BP; two items with five and six response categories each, respectively), general health perceptions (GH; five items with five response categories each), vitality (VT; four items with six response categories each), social functioning (SF; two items with five response categories each), role limitations due to emotional health (RE; three items with two response categories each), and general mental health (MH; five items with six response categories each). In addition, there is one item with five response categories to assess Health Comparison (HC). The observed response categories were coded such that higher scores indicate higher levels of functioning or well-being. The dataset consisted of 630 contingency tables in total. When both observed ordinal variables were dichotomous ($I = J = 2$), the null hypothesis of underlying bivariate normality could not be tested; therefore, the 21 contingency tables with two dichotomous variables were excluded from the analysis.

4.2. Fitted Distributions

The possible bivariate distributions that could be tested varied as a function of the number of response categories of each ordinal variable in the contingency table (see Table 2). This is because a distribution can only be tested for a given pair of ordinal variables if the degrees of freedom that partly depend on the number of response categories are positive. Consider, for example, a pair of ordinal variables with $I \times J = 2 \times 3 = 6$ possible response patterns. The total number of thresholds to be estimated for this pair of variables is $I + J - 2 = 2 + 3 - 2 = 3$. If the bivariate normal distribution is fitted, ρ is additionally free to be estimated. Hence, the degrees of freedom when fitting the bivariate normal distribution to a 2×3 contingency table are $I \times J - n - 1 = 6 - 4 - 1 = 1$. However, we would not be able to fit the bivariate skew-normal distribution to a 2×3 contingency table, because with the two additional parameters α_1 and α_2 the degrees of freedom become negative, $I \times J - n - 1 = 6 - 6 - 1 = -1$.

The bivariate normal distribution was fitted to each pair of ordinal variables, where the means μ were fixed at zero and the variances σ were fixed at unity. The polychoric correlation ρ and all thresholds $\tau^{(1)}$ and $\tau^{(2)}$ were free to be estimated. When the bivariate skew-normal distribution was fitted to the data, the location and scale parameters were fixed at zero and unity, respectively. The polychoric correlation ρ and all thresholds $\tau^{(1)}$

and $\boldsymbol{\tau}^{(2)}$ were free to be estimated. Moreover, the shape parameters α_1 and α_2 were both estimated. When the mixture of bivariate normal distributions was fitted to the data, the means and variances in the first subpopulation were fixed at zero and unity, respectively, and the polychoric correlation ρ, the thresholds $\boldsymbol{\tau}^{(1)}$ and $\boldsymbol{\tau}^{(2)}$, and the component probability λ were free parameters. Thresholds were constrained to be equal across subpopulations. The correlation in the second subpopulation was either fixed at zero ($\rho^* = 0$), constrained to be equal to the correlation in the first subpopulation ($\rho^* = \rho$), or set free to be estimated. Moreover, the second subpopulation's mean and variance of both underlying continuous variables were freely estimated.

Table 2. The degrees of freedom to test underlying bivariate distributions among different contingency tables.

Distributions	$I \times J$									C
	2×3	3×3	2×5	2×6	3×5	3×6	5×5	5×6	6×6	
Normal	1	3	3	4	7	9	17	19	24	700
Skew-normal	−1	1	1	2	5	7	15	17	22	630
Mixture (ρ^* fixed)	−4	−2	−2	−1	2	4	12	14	19	452
Mixture (ρ^* free)	−5	−3	−3	−2	1	3	11	13	18	452

Note. C reflects the total number of contingency tables that could be analyzed in this study. The degrees of freedom are equal to $I \times J - n - 1$, where I and J are numbers of response categories and n is the number of estimated parameters.

4.3. Analysis

The underlying bivariate distributions were fitted to the pairs of ordinal variables by minimizing Equation (10). The minimization was solved using a one-step procedure [28]. That is, all parameters in γ were simultaneously estimated. With k items, $k(k-1)/2$ tests are conducted in each dataset. Therefore, the goodness of fit was not only tested at 0.05, but also at a Bonferroni-adjusted level of significance $\frac{2 \times 0.05}{k(k-1)}$ to avoid inflated family-wise Type I error rates (note that this procedure is different from Raykov and Marcoulides [35] who applied a Benjamini-Hochberg procedure). The null hypothesis was rejected when the LRT statistic was significant. This indicated that the underlying continuous latent variables did not follow the bivariate distribution being tested.

In order to minimize Equation (10), the R function `nlminb()` from the PORT library [36] was used. This function uses a quasi-Newton algorithm that can be subjected to box constraints. Substantial non-convergence rates were observed when fitting the bivariate skew-normal and mixture of bivariate normal distributions with the default `nlminb()` method (see Appendix D). Inspection of the convergence messages suggested that the stopping tolerances were too tight. We therefore used default stopping tolerances, but adjusted the relative tolerance when we encountered non-convergence. As we also observed convergence to local minima, we used multiple starting values for the bivariate skew-normal distribution. For the mixture of bivariate normal distributions, we imposed lower and/or upper constraints on the polychoric correlations, the component probability, and the variances of the underlying continuous latent variables in order to avoid inadmissible values of these parameters.

We calculated the percentage of contingency tables for which each tested underlying distribution was rejected in each of the datasets. In addition, we evaluated the absolute difference in polychoric correlation estimates averaged across all contingency tables in both datasets as outcome variables. The absolute difference was defined as $| \hat{\rho}_A - \hat{\rho}_N |$, where $\hat{\rho}_N$ is the polychoric correlation estimate assuming underlying bivariate normality, and $\hat{\rho}_A$ is the estimate of the polychoric correlation assuming an alternative underlying bivariate distribution (i.e., skew-normal distribution or mixture of bivariate normal distributions). For the two mixtures of bivariate normal distributions in which ρ^* is allowed to differ from ρ, the correlation of the largest subsample was used for the calculation of the absolute difference. Note that the absolute differences do not reflect differences with regard to population values and therefore cannot be interpreted as reflecting estimation bias. Instead,

the absolute differences in polychoric correlation estimates provide information about the range of estimates obtained using different distributions with empirical data. Results were analyzed with R version 3.4.3 [37].

4.4. Results

4.4.1. Rejections of the Distributions

Table 3 shows the rejection percentages of the bivariate distributions among all pairs of ordinal variables. In the Type D personality dataset, the bivariate normal distribution was rejected for 83.52% of the variable pairs when the level of significance was 0.05. With a Bonferroni-adjusted significance level, the bivariate normal distribution was rejected for 42.86% of the contingency tables. For most pairs of ordinal variables, the assumption of underlying bivariate normality is thus violated. The bivariate skew-normal distribution obtained comparably high percentages of rejection. With a significance level of 0.05, the bivariate skew-normal distribution was rejected for 79.55% of the variable pairs. The distributions that obtained the lowest rejection percentages were the mixtures of bivariate normal distributions. With a Bonferroni-adjusted significance level, the mixture of bivariate normal distributions with ρ^* fixed at zero was not rejected for any pair of variables.

Table 3. Rejection percentages of each distribution across all pairs of variables.

Distributions	Type D Personality		Health Status	
	Unadjusted	**Bonferroni Adjusted**	**Unadjusted**	**Bonferroni Adjusted**
Normal	83.52	42.86	71.43	35.96
Skew-normal	79.55	44.09	63.09	20.88
Mixture ($\rho^* = 0$)	14.94	0.00	20.78	4.99
Mixture ($\rho^* = \rho$)	34.07	6.53	34.63	12.74
Mixture (ρ^* free)	18.68	3.30	20.22	3.60

Note. The unadjusted level of significance was 0.05 for each of the distributions. The Bonferroni-adjusted significance level was 0.05/91 in the Type D personality dataset. In the health status dataset, the Bonferroni-adjusted significance level was 0.05/609 for the bivariate normal distribution, 0.05/539 for the bivariate skew-normal distribution, and 0.05/361 for the mixtures of normal distributions.

In the health status dataset, the bivariate normal distribution was rejected for 71.43% of the pairs of variables using an unadjusted significance level. The null hypothesis of underlying bivariate normality was rejected for 35.06% pairs of variables when a Bonferroni-adjusted level of significance was used. The percentages of rejection of the bivariate skew-normal distribution were slightly lower. Again, the mixtures of bivariate normal distributions showed substantially lower rejection percentages. For instance, with a Bonferroni-adjusted significance level, the mixture of bivariate normal distributions in which ρ^* is fixed at zero and the means and variances of both underlying variables are freely estimated was rejected for only 4.99% of contingency tables. When ρ^* was additionally estimated, the rejection percentage was 3.60%.

4.4.2. Absolute Difference in ρ

The average absolute differences for the bivariate skew-normal distribution and mixtures of bivariate normal distributions are presented in Table 4. Compared to the other distributions, the skew-normal distribution and mixture of bivariate normal distributions with $\rho^* = \rho$ generally produced polychoric correlation estimates that were close to the polychoric correlation estimate assuming underlying bivariate normality $\hat{\rho}_N$. These distributions obtained comparable average absolute differences (i.e., 0.03 and 0.04). The largest average absolute differences were found for the mixture of bivariate normal distributions with ρ^* as a free parameter and the mixture of bivariate normal distributions with ρ^* fixed at zero. The average absolute differences of these distributions ranged from 0.26 to 0.31, and were substantially larger than those obtained by the bivariate skew-normal distribution and the mixture of bivariate normal distributions with $\rho^* = \rho$.

Table 4. The average absolute differences in ρ.

Distributions	Type D Personality	Health Status
Skew-normal	0.03	0.04
Mixture ($\rho^* = 0$)	0.26	0.30
Mixture ($\rho^* = \rho$)	0.03	0.04
Mixture (ρ^* free)	0.26	0.31

Note. The absolute differences were averaged across the contingency tables in the dataset. In the Type D personality dataset, there was a total of 91 contingency tables. In the health status dataset, the total number of contingency tables was 609 for the bivariate normal distribution, 539 for the bivariate skew-normal distribution, and 361 for the mixtures of normal distributions.

5. Discussion

In line with existing literature [6–11], this study showed that the underlying bivariate normal distribution seldom holds in empirical data. This may indicate that in this study, the polychoric correlation based on underlying bivariate normality is an under- or over-estimation of the correlation between the underlying continuous variables for most pairs of ordinal variables [19]. The bivariate skew-normal distribution was also frequently rejected in empirical data.

A possible explanation for the high rejection percentages of the bivariate skew-normal distribution may involve the dependency between skewness and kurtosis. In Azzalini and Dalla Valle's skew-normal distribution [20], skewness and kurtosis were regulated with the same parameter. Hence, a bivariate skew-normal distribution with high skewness cannot have low kurtosis, or the other way around. This might not be realistic, as in practice, data can be highly skewed with low kurtosis, or the other way around. Moreover, similar to Jin and Yang-Wallentin [19], we encountered non-convergence and local optima when estimating the parameters of the bivariate skew-normal distribution. Multiple starting values and adjusted stopping tolerances for minimizing the LRT statistic were used in this study to overcome these problems.

The mixture of bivariate normal distributions was not often rejected in the empirical datasets. Specifically, the mixture of bivariate normal distributions with a free or fixed-at-zero correlation in the second subpopulation was often found to be consistent with the data. Although the mixture with a freely estimated correlation in the second subpopulation is less restrictive, the mixture with a correlation in the second subpopulation fixed at zero was rejected less often in one of the datasets. This was against our expectations, because in theory, a less restrictive distribution (with more parameters to be estimated) will always fit the data better than a nested, more restrictive distribution (with fewer parameters to be estimated). The higher rejection percentages of the less restrictive mixture distribution with a free correlation in the second subpopulation may be caused by an increase in fit of the distribution that is too small relative to the decrease in degrees of freedom, or by local optima. Overall, our results support Uebersax's [24] suggestion that the mixture distribution is a flexible tool to model non-normal distributions, but there is a need for further studies in which its performance is investigated. One of the issues of the mixture distribution is the risk of overfitting because of the large number of parameters. Mixture distributions may therefore not generalize easily to other samples.

5.1. Future Research and Limitations

The current paper only considers testing bivariate distributions. A logical next step would be to extend the procedures to a multivariate approach, and estimate a matrix of polychoric correlation coefficients based on all variable pairs, with various underlying distributions. The extension to the multivariate case brings about multiple issues to be considered. For example, it would make sense to add constraints to the thresholds, since it would be undesirable if a different set of thresholds would be estimated for each pair of ordinal variables [28]. Additionally, with the mixture distributions, it may be needed to constrain the component probabilities for the subpopulations to be equal across variable

pairs, and to avoid estimating a free correlation coefficient in each subpopulation (but instead constrain them to be equal across subpopulations, or to fix one of the correlations to zero). In order to fit structural equation models to polychoric correlation matrices using weighted least squares estimators, the asymptotic covariance matrix of the estimated parameters will be needed. Currently, there are no straightforward methods to obtain the asymptotic covariance matrix based on other underlying distributions than the bivariate normal, although as a reviewer suggested, the approach by Monroe [38] might be extended. Correspondingly, the estimated polychoric correlation matrix should be positive semi-definite in order to fit structural equation models, which may not be the case without the appropriate constraints, or with a misspecified underlying distribution. Moreover, although the LRT has been evaluated to test for underlying multivariate normality [35,39], as well as a parametric bootstrap procedure [17], to our knowledge, there exist no simulation studies investigating the statistical performance of these tests when the tested distribution is non-normal. Such a simulation study would be useful to get information about the Type 1 and Type 2 errors, estimation bias in correlation coefficients, and the needed sample sizes for adequate performance of the tests.

A limitation of the present study is that although a large number of contingency tables were analyzed, they only stemmed from two datasets. This study provides a first insight into the distributions underlying the responses to ordinal variables, but more data must be investigated in order to verify the results. Moreover, the present study examined only a few non-normal bivariate distributions. It would be interesting to evaluate the fit of other non-normal bivariate distributions to empirical data as well. For example, Timofeeva and Khailenko [10] proposed a polychoric correlation assuming a generalized lambda distribution for the underlying continuous variables. The generalized lambda distribution is a non-symmetric extension of Tukey's lambda distribution and is known for its high flexibility in physical and social science settings, among others [40]. Research shows that the bivariate generalized lambda distribution is rejected less often for empirical data than the bivariate normal distribution [10]. It would be interesting to examine the rejection percentages of the generalized lambda distributions in other, possibly larger, empirical datasets that consist of ordinal responses.

5.2. Conclusions

Overall, this study showed that the bivariate normal and skew-normal distribution were often rejected when tested against empirical data. The results of this study also showed that the polychoric correlation estimates based on the skew-normal distribution and mixture distributions can be substantially different from the estimates assuming underlying bivariate normality. Hence, the present study underlines the importance of testing the assumed underlying distribution for the estimation of the polychoric correlation coefficient.

Author Contributions: Conceptualization, F.O., S.J. and L.K.; methodology, F.O., S.J. and L.K.; software, F.O., S.J. and L.K.; formal analysis, L.K.; investigation, F.O., S.J. and L.K.; resources, F.O.; writing—original draft preparation, L.K.; writing—review and editing, L.K., S.J. and F.O.; visualization, L.K.; supervision, S.J. and F.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Dutch Research Council under Grant NWO-VENI-451-16-001 awarded to S. Jak.

Data Availability Statement: The data are available in the R script presented in Appendix C and upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

```
## Script: Functions for the bivariate (skew -)normal distribution

# Fit function
```

```
LR_skew <- function(params) {
nthresholds1 = length(params[1:nthresholds1])
nthresholds2 = length(params[(nthresholds1+1):(nthresholds1+nthresholds2)])
upperlimit1 = 10 + params[nthresholds1]
upperlimit2 = 10 + params[nthresholds1+nthresholds2]
limits1 = c(params[1:nthresholds1], upperlimit1)
limits2 = c(params[(nthresholds1+1):(nthresholds1+nthresholds2)], upperlimit2)
if(is.na(params["alpha1"])){alpha1 = 0} else {alpha1 = params["alpha1"]}
if(is.na(params["alpha2"])){alpha2 = 0} else {alpha2 = params["alpha2"]}
cumul = matrix(0, ncats1, ncats2)
expp = matrix(0, ncats1, ncats2)
for (i in 1:ncats1) {
for (j in 1:ncats2) {
cumul[i,j] = sn::pmsn(c(limits1[i], limits2[j]), c(0,0),
matrix(c(1, params["corr"], params["corr"], 1),2,2),
c(alpha1, alpha2))
}
}
expp[1,1] = cumul[1,1]
for (i in 2:ncats1) { expp[i,1] = cumul[i,1] - cumul[i-1,1] }
for (j in 2:ncats2) { expp[1,j] = cumul[1,j] - cumul[1,j-1] }
for (i in 2:ncats1) {
for (j in 2:ncats2) {
expp[i,j] = cumul[i,j] - cumul[i-1,j] - cumul[i,j-1] + cumul[i-1,j-1]
}
}
pi = ifelse(expp > 0, expp, 0.0000000001)
p = ifelse (obsp > 0, obsp, 0.0000000001)
return(2*ntot*sum(obsp*log(p/pi)))
}

# Optimization
fit_skewnorm <- function(parameters){
results_skew = nlminb(parameters, LR_skew, control = list(rel.tol = 1e-3))
out = data.frame(matrix(NA, 1, 1))
colnames(out) = c("chisq")
out$chisq = results_skew$objective
out$df = ncats1 * ncats2 - 1 - length(results_skew$par)
out$p = 1 - pchisq(results_skew$objective, out$df)
out$corr = results_skew$par["corr"]
if(!is.na(results_skew$par["alpha1"])) {
out$alpha1 = results_skew$par["alpha1"]
}
if(!is.na(results_skew$par["alpha2"])) {
out$alpha2 = results_skew$par["alpha2"]
}
options(scipen=999)
return(list("results" = results_skew, "output" = out))
}
```

Appendix B

```
## Script: Functions for the mixture of bivariate distributions

# Bivariate normal distribution
biv <- function(thresholds1, thresholds2, muvar, covma) {
nthresholds1 = length(thresholds1)
nthresholds2 = length(thresholds2)
ncats1 = nthresholds1 + 1
ncats2 = nthresholds2 + 1
upperlimit1 = 10 + thresholds1[nthresholds1]
upperlimit2 = 10 + thresholds2[nthresholds2]
limits1 = c(thresholds1, upperlimit1)
limits2 = c(thresholds2, upperlimit2)
cumul = matrix(0, ncats1, ncats2)
expp = matrix(0, ncats1, ncats2)
for (i in 1:ncats1) {
for (j in 1:ncats2) {
cumul[i,j] = mnormt::pmnorm(c(limits1[i], limits2[j]), muvar, covma)
}
}
expp[1,1] = cumul[1,1]
for (i in 2:ncats1) { expp[i,1] = cumul[i,1] - cumul[i-1,1] }
for (j in 2:ncats2) { expp[1,j] = cumul[1,j] - cumul[1,j-1] }
for (i in 2:ncats1) {
for (j in 2:ncats2) {
expp[i,j] = cumul[i,j] - cumul[i-1,j] - cumul[i,j-1] + cumul[i-1,j-1]
}
}
return(expp)
}
```

```r
# Fit function
LR_mixed <- function(params) {
thresholds1 = params[1:nthresholds1]
thresholds2 = params[(nthresholds1+1):(nthresholds1+nthresholds2)]
nthresholds1 = length(params[1:nthresholds1])
nthresholds2 = length(params[(nthresholds1+1):(nthresholds1+nthresholds2)])
if(is.na(params["sigmastar1"])){sigmastar1 = 1} else {sigmastar1 = params["sigmastar1"]}
if(is.na(params["sigmastar2"])){sigmastar2 = 1} else {sigmastar2 = params["sigmastar2"]}
if(is.na(params["mustar1"])){mustar1 = 0} else {mustar1 = params["mustar1"]}
if(is.na(params["mustar2"])){mustar2 = 0} else {mustar2 = params["mustar2"]}
if(is.na(params["corrstar"])){corrstar = params["corr"]}
else {corrstar = params["corrstar"]}
covstar = corrstar*sqrt(sigmastar1*sigmastar2)
expp = biv(thresholds1, thresholds2, c(0,0), matrix(c(1, params["corr"],
params["corr"], 1),2,2))
if (params["prop"] > 0) {
exppstar = biv(thresholds1, thresholds2, c(mustar1, mustar2),
matrix(c(sigmastar1, covstar, covstar, sigmastar2),2,2))
expp = (params["prop"]*expp) + ((1-params["prop"])*exppstar)
}
pi = ifelse(expp > 0, expp, 0.0000000001)
p = ifelse (obsp > 0, obsp, 0.0000000001)
return(2*ntot*sum(obsp*log(p/pi)))
}

# Optimization
fit_mix <- function(parameters, ll, uu){
if (missing(ll)) ll = -100
if (missing(uu)) uu = 100
results_mixed = nlminb(parameters, LR_mixed, lower = ll, upper = uu)
out = data.frame(matrix(NA, 1, 1))
colnames(out) = c("chisq")
out$chisq = results_mixed$objective
out$df = ncats1 * ncats2 - 1 - length(results_mixed$par)
out$p = round(1 - pchisq(results_mixed$objective, out$df), 3)
out$corr = results_mixed$par["corr"]
if(!is.na(results_mixed$par["mustar1"])) {
out$mustar1 = results_mixed$par["mustar1"]
}
if(!is.na(results_mixed$par["mustar2"])) {
out$mustar2 = results_mixed$par["mustar2"]
}
out$prop = results_mixed$par["prop"]
if(!is.na(results_mixed$par["sigmastar1"])) {
out$sigstar1 = results_mixed$par["sigmastar1"]
}
if(!is.na(results_mixed$par["sigmastar2"])) {
out$sigstar2 = results_mixed$par["sigmastar2"]
}
if(!is.na(results_mixed$par["corrstar"])) {
out$corrstar = results_mixed$par["corrstar"]
}
options(scipen=999)
return(list("results" = results_mixed, "output" = out))
}
```

Appendix C

```r
## Script: Fitting distributions to Type D personality data

# Required packages
library(polycor) # we used version 3.0.6
library(sn)      # we used version 2.0.0
library(mnormt)  # we used version 0.7-10
library(mokken)  # we used version 2.0.2

########################
#### Initialization ####
########################

data("DS14")

obsn = table(DS14[,"Na2"], DS14[,"Si6"])

ncats1 = nrow(obsn)
ncats2 = ncol(obsn)

ntot = sum(obsn)
obsp = obsn/ntot
```

```
proportions2 = matrix(colSums(obsp), 1, ncats2)
proportions1 = matrix(rowSums(obsp), ncats1, 1)

premultiplier = matrix(0, ncats1, ncats1)
for (i in 1:ncats1) for (j in 1:i) premultiplier[i, j] = 1

postmultiplier = matrix(0, ncats2, ncats2)
for (i in 1:ncats2) for (j in i:ncats2) postmultiplier[i, j] = 1

cumulprops2 = proportions2 %*% postmultiplier
cumulprops1 = premultiplier %*% proportions1

nthresholds1 = ncats1 - 1
nthresholds2 = ncats2 - 1

thresholds1 = matrix(0, 1, nthresholds1)
for (i in 1:nthresholds1) thresholds1[i] = qnorm(cumulprops1[i])

thresholds2 = matrix (0, 1, nthresholds2)
for (i in 1:nthresholds2) thresholds2[i] = qnorm(cumulprops2[i])

corr = polycor::polychor(obsn)

###########################
#### Fit distributions ####
###########################

# Fit bivariate normal distribution
results_norm = fit_skewnorm(c("th1" = thresholds1, "th2" = thresholds2,
"corr" = corr))
results_norm

# Fit skew-normal distribution
results_skew = fit_skewnorm(c("th1" = thresholds1, "th2" = thresholds2,
"corr" = corr, "alpha" = c(2,2)))
results_skew

# Calculate polychoric correlation assuming a skew-normal
dp = list(xi = c(0,0), Omega = matrix(c(1, results_skew$output$corr,
results_skew$output$corr, 1),2,2), alpha = c(results_skew$output$alpha1,
results_skew$output$alpha2))
sn1 = sn::makeSECdistr(dp, family = "SN")
summary(sn1)

polcorr = (results_skew$output$corr -2*pi^(-1)*0.3816442*0.8618373) /
(((1-2*pi^(-1)*0.3816442^2)*(1-2*pi^(-1)*0.8618373^2))^0.5)

# Fit mixture distribution
param = c("th1" = thresholds1, "th2" = thresholds2, "corr" = corr, "prop" = 0.7,
"corrstar" = corr, "sigmastar1" = 1, "sigmastar2" = 1, "mustar1" = 0,
"mustar2" = 0)
results_mix = fit_mix(param, c(rep(-10, nthresholds1+nthresholds2), -1, 0, -1,
0.001, 0.001, -10, -10), c(rep(10, nthresholds1+nthresholds2),
1, 1, 1, 10, 10, 10, 10))
results_mix
```

Appendix D

In the `nlminb()` function, the relative tolerance defaults to 1e-10. Relative tolerance works as follows. The minimization stops if the algorithm is unable to reduce the value of the objective to be minimized by a factor of the sum of the absolute value of the objective and the relative tolerance. We prevented non-convergence by using `nlminb()` with a relative tolerance of 1e-5 or 1e-3. We found similar rejection percentages but less convergence problems with the adjusted relative tolerances. In the table below, the non-convergences rates under the different relative tolerances are presented.

Table A1. Nonconvergence rates of each of the distributions under different relative tolerances.

Distributions	Type D Personality			Health Status		
	1e-10	1e-5	1e-3	1e-10	1e-5	1e-3
Normal	0.00	0.00	0.00	0.82	0.00	0.00
Skew-normal	81.32	36.26	5.49	62.15	7.79	7.79
Mixture ($\rho^* = 0$)	24.18	23.08	17.58	14.13	6.09	6.09
Mixture ($\rho^* = \rho$)	20.88	1.10	0.00	54.85	0.55	0.55
Mixture (ρ^* free)	42.86	25.27	7.69	41.55	4.16	4.16

References

1. Pearson, K. Mathematical Contributions to the Theory of Evolution. VII. On the Correlation of Characters not Quantitatively Measurable. *Philos. Trans. R. Soc. London. Ser. A Contain. Pap. Math. Phys. Character* **1900**, *195*, 1–47. [CrossRef]
2. Jöreskog, K.G.; Sörbom, D. *PRELIS 2 User's Reference Guide*; Scientific Software International: Chicago, IL, USA, 1996.
3. Babakus, E.; Ferguson, C.E. On choosing the appropriate measure of association when analyzing rating scale data. *J. Acad. Mark. Sci.* **1988**, *16*, 95–102. [CrossRef]
4. Yule, G.U. On the methods of measuring association between two attributes. *J. R. Stat. Soc.* **1912**, *75*, 579–652. [CrossRef]
5. Robitzsch, A. Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Front. Educ.* **2020**, *5*, 177. [CrossRef]
6. Ekström, J. *A Generalized Definition of the Polychoric Correlation Coefficient*; Department of Statistics, UCLA: Los Angeles, CA, USA, 2011. Available online: https://escholarship.org/uc/item/583610fv (accessed on 1 September 2021).
7. Jöreskog, K.G. *Structural Equation Modeling with Ordinal Variables Using LISREL*; Scientific Software International: Chicago, IL, USA, 2005.
8. Muthén, B.; Hofacker, C. Testing the assumptions underlying tetrachoric correlations. *Psychometrika* **1988**, *53*, 563–577. [CrossRef]
9. Şimşek, G.G.; Noyan, F. Structural equation modeling with ordinal variables: A large sample case study. *Qual. Quant.* **2012**, *46*, 1571–1581. [CrossRef]
10. Timofeeva, A.Y.; Khailenko, E.A. Generalizations of the polychoric correlation approach for analyzing survey data. In Proceedings of the 2016 11th International Forum on Strategic Technology (IFOST), Novosibirsk, Russia, 1–3 June 2016; pp. 254–258. [CrossRef]
11. Yamamoto, K.; Murakami, H. Model based on skew normal distribution for square contingency tables with ordinal categories. *Comput. Stat. Data Anal.* **2014**, *78*, 135–140. [CrossRef]
12. Flora, D.B.; Curran, P.J. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* **2004**, *9*, 466. [CrossRef]
13. Lee, S.Y.; Lam, M.L. Estimation of polychoric correlation with elliptical latent variables. *J. Stat. Comput. Simul.* **1988**, *30*, 173–188. [CrossRef]
14. Quiroga, A.M. Studies of the Polychoric Correlation and other Correlation Measures for Ordinal Variables. Ph.D. Thesis, Acta Universitatis Upsaliensis, Univsersity of Uppsala, Uppsala, Sweden, 1992.
15. Grønneberg, S.; Foldnes, N. A Problem with discretizing Vale–Maurelli in simulation studies. *Psychometrika* **2019**, *84*, 554–561. [CrossRef]
16. Foldnes, N.; Grønneberg, S. The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychol. Methods* **2021**. [CrossRef]
17. Foldnes, N.; Grønneberg, S. Pernicious polychorics: The impact and detection of underlying non-normality. *Struct. Equ. Model. Multidiscip. J.* **2020**, *27*, 525–543. [CrossRef]
18. Foldnes, N.; Grønneberg, S. On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika* **2019**, *84*, 1000–1017. [CrossRef]
19. Jin, S.; Yang-Wallentin, F. Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika* **2017**, *82*, 67–85. [CrossRef] [PubMed]
20. Azzalini, A.; Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726. [CrossRef]
21. Azzalini, A.; Capitanio, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2003**, *65*, 367–389. [CrossRef]
22. Mardia, K.V. Multivariate pareto distributions. *Ann. Math. Stat.* **1962**, *33*, 1008–1015. [CrossRef]
23. Roscino, A.; Pollice, A. A generalization of the polychoric correlation coefficient. In *Data Analysis, Classification and the Forward Search*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 135–142. [CrossRef]
24. Uebersax, J.S. Latent Correlation with Skewed Latent Distributions: A Generalization of the Polychoric Correlation Coefficient and a Computer Program for Estimation. Available online: https://www.john-uebersax.com/stat/skewed.htm (accessed on 17 September 2021).
25. Uebersax, J.S.; Grove, W.M. A latent trait finite mixture model for the analysis of rating agreement. *Biometrics* **1993**, *49*, 823–835. [CrossRef] [PubMed]

26. Grønneberg, S.; Moss, J.; Foldnes, N. Partial Identification of Latent Correlations with Binary Data. *Psychometrika* **2020**, *85*, 1028–1051. [CrossRef] [PubMed]

27. Pearson, K. Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **1894**, *185*, 71–110. [CrossRef]

28. Olsson, U. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **1979**, *44*, 443–460. [CrossRef]

29. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.

30. Agresti, A. *Categorical Data Analysis*; Wiley-Interscience: New York, NY, USA, 2002; Volume 482.

31. Denollet, J.; Pedersen, S.S.; Vrints, C.J.; Conraads, V.M. Predictive value of social inhibition and negative affectivity for cardiovascular events and mortality in patients with coronary artery disease: The type D personality construct. *Psychosom. Med.* **2013**, *75*, 873–881. [CrossRef] [PubMed]

32. Denollet, J. DS14: Standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosom. Med.* **2005**, *67*, 89–97. [CrossRef] [PubMed]

33. Aaronson, N.K.; Muller, M.; Cohen, P.D.; Essink-Bot, M.L.; Fekkes, M.; Sanderman, R.; Sprangers, M.A.; Te Velde, A.; Verrips, E. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J. Clin. Epidemiol.* **1998**, *51*, 1055–1068. [CrossRef]

34. Ware, J.E.; Snow, K.K.; Kosinski, M.; Gandek, B. *SF-36 Health Survey: Manual and Interpretation Guide*; The Health Institute, New England Medical Center: Boston, MA, USA, 1993.

35. Raykov, T.; Marcoulides, G.A. On examining the underlying normal variable assumption in latent variable models with categorical indicators. *Struct. Equ. Model. Multidiscip. J.* **2015**, *22*, 581–587. [CrossRef]

36. Gay, D.M. Usage summary for selected optimization routines. *Comput. Sci. Tech. Rep.* **1990**, *153*, 1–21.

37. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

38. Monroe, S. Contributions to Estimation of Polychoric Correlations. *Multivar. Behav. Res.* **2018**, *53*, 247–266. [CrossRef]

39. Maydeu-Olivares, A. Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika* **2006**, *71*, 57–77. [CrossRef]

40. Karian, Z.A.; Dudewicz, E.J. Fitting the generalized lambda distribution to data: A method based on percentiles. *Commun.-Stat.-Simul. Comput.* **1999**, *28*, 793–819. [CrossRef]