FACTOR COPULA MODELS FOR ITEM RESPONSE DATA

ARISTIDIS K. NIKOLOULOPOULOS

UNIVERSITY OF EAST ANGLIA

HARRY JOE

UNIVERSITY OF BRITISH COLUMBIA

Factor or conditional independence models based on copulas are proposed for multivariate discrete data such as item responses. The factor copula models have interpretations of latent maxima/minima (in comparison with latent means) and can lead to more probability in the joint upper or lower tail compared with factor models based on the discretized multivariate normal distribution (or multidimensional normal ogive model). Details on maximum likelihood estimation of parameters for the factor copula model are given, as well as analysis of the behavior of the log-likelihood. Our general methodology is illustrated with several item response data sets, and it is shown that there is a substantial improvement on existing models both conceptually and in fit to data.

Key words: conditional independence; factor model dependence structure; latent variable model; limited information; partial correlation.

1. Introduction

Latent variable or factor models are a unified tool for the analysis of high-dimensional response data with dependence coming from latent (unobservable) variables/factors so that the number of dependence parameters is $O(d)$ rather than $O(d^2)$, where d is the number of observed variables; see for example, Bartholomew et al. (2011). For example, a questionnaire or instrument, used in psychometrics to assess abstract concepts, such as the quality of life, conservatism, and general intelligence, may have $d \geq 50$ items or questions, but many questions overlap or are correlated by design. Theoretically, the abstract concept is low-dimensional, so a latent variable model is reasonable.

In this paper, the main new contribution is the construction of factor or conditional independence models based on copula functions (distributions with uniform $U(0,1)$ margins on the unit interval) for item response variables, $Y_1, \ldots, Y_d$ for d items, where the items (questions) are measured on an ordinal scale; $Y_j \in \{0, \ldots, K-1\}$ for $j = 1, \ldots, d$. The p-factor model assumes that $Y_1, \ldots, Y_d$ are conditionally independent given latent

variables $X_1, \ldots, X_p$; and, hence, the joint probability mass function (pmf) is

$$\pi_d(\mathbf{y}) = \Pr(Y_1 = y_1, \ldots, Y_d = y_d)$$

$$= \int \prod_{j=1}^{d} \Pr(Y_j = y_j | X_1 = x_1, \ldots, X_p = x_p) \, dF_{X_1, \ldots, X_p}(x_1, \ldots, x_p),$$

where $F_{X_1, \ldots, X_p}$ is the distribution of the latent variable. We will use a general copula construction, based on a set of bivariate copulas that link observed to latent variables, to specify $\Pr(Y_j = y_j | X_1 = x_1, \ldots, X_p = x_p)$ and arrive at a very general conditional independence or factor model. Discretized multivariate normal (MVN) models with p-factor correlation matrices (Maydeu-Olivares, 2006) are special cases of our general construction when all the above bivariate linking copulas are bivariate normal.

A copula approach in psychometrics was recently proposed by Braeken et al. (2007) and Braeken (2011) who explored the use of Archimedean copulas or a mixture of the independence and comonotonicity copulas to capture the residual dependence of the Rasch model. The multivariate probit or discretized MVN model has been in use for a considerable length of time in psychometrics (Muthén, 1978), and it can be considered as a special case of the MVN copula with univariate probit marginals.

If a discretized MVN model with factor dependence structure provides a poor fit, it could be due to too few factors or because of departure from the assumption that the (conditional) cumulative response function is the normal distribution (Jöreskog and Moustaki, 2001). In our general construction with conditional independence, we modify discretized MVN-based models by replacing bivariate normal copulas between observed and latent variables with other choices of bivariate copulas. Other choices of copulas are better if (a) Y_j's have more probability in joint upper or lower tail than would be expected with a discretized MVN, or (b) Y_j's can be considered as discretized maxima/minima or mixtures of discretized means rather than discretized means.

In this paper, we mainly emphasize that if the ordinal variables in item response can be thought of as discretization of latent random variables that are maxima or minima or high/low quantiles, then the use of factor models based on copulas can be an improvement both conceptually and in fit to data. In the context of item response data, latent maxima, minima and means can arise depending on how a respondent considers specific items. An item might make the respondent think about M past events which, say, have values $W_1, \ldots, W_M$. In answering the item, the subject might take the average, maximum or minimum of $W_1, \ldots, W_M$ and then convert to the ordinal scale depending on the magnitude. Less extreme than the maximum or minimum would be high or low quantiles of values from relevant past events. The case of a latent maxima/minima can occur if the response is based on a best or worst case. Concrete examples are given in Section 5. For different dependent items based on latent maxima or minima, multivariate extreme value and copula theory can be used to select suitable copulas that link observed to latent variables. Copulas that arise from extreme value theory have more probability in one joint tail (upper or lower) than expected with a discretized MVN distribution or an MVN copula with discrete margins. For latent variables that match high or low quantiles of past events, the same comparison with a discretized MVN can be expected.

If item responses are based on discretizations of latent variables that are means, then

it is possible that there can be more probability in both the joint upper and joint lower tail, compared with discretized MVN models. This happens if the respondents consist of a "mixture" population (e.g., different locations or genders). From the theory of elliptical distributions and copulas, it is known that some scale mixtures of MVN have more dependence in the tails.

References that have theory for copulas, extreme value and elliptical distributions are Joe (1997) and McNeil et al. (2005). The theory and application of copulas have become important in finance, insurance and other areas, in order to deal with dependence in the joint tails. Here, we indicate that this can also be important in psychometrics.

The remainder of the paper proceeds as follows. Section 2 has a brief overview of relevant copula theory and then introduces the factor copula model for discrete ordinal responses and discusses its relationship with existing models. Section 3 summarizes the assessment of goodness-of-fit of these models using the statistic of Maydeu-Olivares and Joe (2006), which is based on a quadratic form of the deviations of sample and model-based proportions over all bivariate margins. Estimation techniques and computational details are provided in Section 4; included is an analysis of the behavior of the log-likelihood. Section 5 presents applications of our methodology to three item response data sets. In these examples, it turns out that our model, with linking copulas selected according to the items being plausibly based on latent maxima/minima or mixtures of means, provides better fit than the discretized MVN model with factor dependence structure (also known as the multidimensional normal ogive model) and also captures the factor dependence structure. We conclude with some discussion in Section 6. Additional materials, including details of computational implementation and pseudo-code for the evaluation of the joint log-likelihood and the statistic of Maydeu-Olivares and Joe (2006), are provided as appendices.

2. The factor copula model for ordinal responses

In this section, we construct the factor copula model based on conditional independence of ordinal variables $Y_1, \ldots, Y_d$ given latent variables. The general theory is presented for one and two factors, and the extension to $p \geq 3$ factors follows the same approach. Before that, the first subsection has some background on copula models.

2.1. Overview and relevant background for copulas

A copula is a multivariate cumulative distribution function (cdf) with uniform $U(0, 1)$ margins. If F is a d-variate cdf with univariate margins $F_1, \ldots, F_d$, then Sklar's (1959) theorem implies that there is a copula C such that

$$F(y_1, \ldots, y_d) = C\Big(F_1(y_1), \ldots, F_d(y_d)\Big).$$

The copula is unique if $F_1, \ldots, F_d$ are continuous, but not if some of the F_j have discrete components. If F is continuous and $(Y_1, \ldots, Y_d) \sim F$, then the unique copula is the distribution of $(U_1, \ldots, U_d) = (F_1(Y_1), \ldots, F_d(Y_d))$ leading to

$$C(u_1, \ldots, u_d) = F\Big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\Big), \quad 0 \leq u_j \leq 1, j = 1, \ldots, d,$$

where F_j^{-1} are inverse cdfs. In particular, if $\Phi_d(\cdot; \Sigma)$ is the MVN cdf with correlation matrix Σ and N(0,1) margins, and Φ is the univariate standard normal cdf, then the MVN copula is

$$C(u_1, \ldots, u_d) = \Phi_d\Big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d); \Sigma\Big). \tag{1}$$

Copulas have become useful for flexible modeling of multivariate data when the variables Y_j are non-normal, examples of which are ordinal and extreme value. In particular, there are copula families that can lead to more dependence in the joint tail than with the MVN copula, and these are important for extreme value and risk analysis. The power of copulas for dependence modeling is due to the dependence structure being considered separate from the univariate margins; see, for example, Section 1.6 of Joe (1997). If $C(\cdot; \theta)$ is a parametric family of copulas and $F_j(\cdot; \eta_j)$ is a parametric model for the jth univariate margin, then

$$C\Big(F_1(y_1; \eta_1), \ldots, F_d(y_d; \eta_d); \theta\Big)$$

is a multivariate parametric model with univariate margins $F_1, \ldots, F_d$. For copula models, the variables can be continuous or discrete.

For a parametric copula model to be implemented for likelihood inference, a simple form for $C(\cdot; \theta)$ is desired. There are many simple bivariate copula families, but generally their multivariate extensions have limited dependence structures in comparison with that available from MVN copulas. In recent years, a popular and useful approach is the vine pair-copula construction (Aas et al., 2009; Kurowicka and Joe, 2011) which is based on $d(d-1)/2$ bivariate copulas, of which most are used to summarize conditional dependence; a special case occurs if all of these bivariate copulas are normal, and then the parametrization of an MVN copula is a set of correlations and partial correlations that are algebraically independent in $(-1, 1)^{d(d-1)/2}$.

Inspired by the vine approach, we present general parametric factor copula models, where for the first factor there are bivariate copulas that couple each observed variable to the first latent variable, and for the second factor there are copulas that link each observed variable to the second latent variable conditioned on the first factor (leading to conditional dependence parameters), etc. These factor models can be explained as truncated canonical vines rooted at the latent variables; but, more simply, we derive the models as conditional independence models to show that they are very general. The 1-factor copula model is the most general possible and the p-factor ($p \geq 2$) is the most general with the simplifying assumption that copulas for conditional distributions do not depend on the values of the conditioning variables.

The most general factor copula model for continuous variables is in Krupskii and Joe (2013), where dependence and tail properties are obtained. Our model for item response is a discrete counterpart, with implementation details that are quite different. Other continuous-variable models in the finance literature that are called factor models (e.g., Section 9.7.2 of McNeil et al. (2005), Hull and White (2004)) have an additive latent structure and are not as general.

2.2. 1-factor and 2-factor copula models

Let the cutpoints in the uniform $U(0,1)$ scale for the jth item/variable be $a_{j,k}$, $k = 1, \ldots, K-1$, with $a_{j,0} = 0$ and $a_{j,K} = 1$. These correspond to $a_{j,k} = \Phi(\alpha_{j,k})$ where $\alpha_{j,k}$ are cutpoints in the normal $N(0,1)$ scale.

For the 1-factor model, let X_1 be a latent variable, which we assume to be standard uniform (without loss of generality). From Sklar (1959), there is a bivariate copula $C_{X_1 j}$ such that $\Pr(X_1 \leq x, Y_j \leq y) = C_{X_1 j}(x, F_j(y))$ for $0 \leq x \leq 1$ where F_j is the cdf of Y_j; note that F_j is a step function with jumps at $0, \ldots, K-1$ and $F_j(y) = a_{j,y+1}$. Then it follows that

$$F_{j|X_1}(y|x) := \Pr(Y_j \leq y | X_1 = x) = \frac{\partial C_{X_1 j}(x, F_j(y))}{\partial x}. \tag{2}$$

Letting $C_{j|X_1}(a|x) = \partial C_{X_1 j}(x,a)/\partial x$ for shorthand notation, the pmf for the 1-factor model is

$$\pi_d(\mathbf{y}) = \int_0^1 \prod_{j=1}^d \Pr(Y_j = y_j | X_1 = x)\, dx = \int_0^1 \prod_{j=1}^d \left[C_{j|X_1}(F_j(y_j)|x) - C_{j|X_1}(F_j(y_j - 1)|x) \right] dx$$

$$= \int_0^1 \prod_{j=1}^d \left[C_{j|X_1}(a_{j,y_j+1}|x) - C_{j|X_1}(a_{j,y_j}|x) \right] dx = \int_0^1 \prod_{j=1}^d f_{j|X_1}(y_j|x)\, dx, \tag{3}$$

where $f_{j|X_1}(y|x) = C_{j|X_1}(a_{j,y+1}|x) - C_{j|X_1}(a_{j,y}|x)$ is the probability of $Y_j = y$ conditional on $X_1 = x$.

For the 2-factor model, consider two latent variables X_1, X_2 that are, without loss of generality, independent uniform $U(0,1)$ random variables. Let $C_{X_1 j}$ be defined as in the 1-factor model, and let $C_{X_2 j}$ be a bivariate copula such that

$$\Pr(X_2 \leq x_2, Y_j \leq y | X_1 = x_1) = C_{X_2 j}(x_2, F_{j|X_1}(y|x_1)), \quad 0 \leq x_1, x_2 \leq 1, \ y = 0, \ldots, K-1,$$

where $F_{j|X_1}$ is given in (2). Here we are making the simplifying assumption that the conditional copula for the univariate distributions $F_{X_2|X_1} = F_{X_2}$ and $F_{j|X_1}$ does not depend on x_1; this is a model assumption as by Sklar's theorem there exist such bivariate copulas that in general depend on $x_1 \in [0,1]$. Then for $0 \leq x_1, x_2 \leq 1$,

$$\Pr(Y_j \leq y | X_1 = x_1, X_2 = x_2) = \frac{\partial}{\partial x_2} \Pr(X_2 \leq x_2, Y_j \leq y | X_1 = x_1)$$

$$= \frac{\partial}{\partial x_2} C_{X_2 j}(x_2, F_{j|X_1}(y|x_1)) = C_{j|X_2}(F_{j|X_1}(y|x_1)|x_2),$$

where $C_{j|X_2}(a|x) = \partial C_{X_2 j}(x,a)/\partial x$ for shorthand notation.

The pmf for the 2-factor model is

$$\pi_d(\mathbf{y}) = \int_0^1 \int_0^1 \prod_{j=1}^d \Pr(Y_j = y_j | X_1 = x_1, X_2 = x_2)\, dx_1 dx_2$$

$$= \int_0^1 \int_0^1 \prod_{j=1}^d \left[C_{j|X_2}(F_{j|X_1}(y_j|x_1)|x_2) - C_{j|X_2}(F_{j|X_1}(y_j - 1|x_1)|x_2) \right] dx_1 dx_2$$

$$= \int_0^1 \int_0^1 \prod_{j=1}^d f_{X_2 j|X_1}(x_2, y_j | x_1)\, dx_1 dx_2, \tag{4}$$

where $f_{X_2j|X_1}(x_2, y|x_1) = C_{j|X_2}(F_{j|X_1}(y|x_1)|x_2) - C_{j|X_2}(F_{j|X_1}(y-1|x_1)|x_2)$. The idea in the derivation of this 2-factor model can be extended to $p \geq 3$ factors or latent variables; the pmf then involves p-dimensional integrals.

For parametric 1-factor and 2-factor models, we let C_{X_1j} and C_{X_2j} be parametric bivariate copulas, say with parameters θ_j and δ_j, respectively. For the set of all parameters, let $\boldsymbol{\theta} = \{a_{jk}, \theta_j : j = 1, \ldots, d; k = 1, \ldots, K-1\}$ for the 1-factor model and $\boldsymbol{\theta} = \{a_{jk}, \theta_j, \delta_j : j = 1, \ldots, d; k = 1, \ldots, K-1\}$ for the 2-factor model. By assuming X_1, X_2 to be independent random variables, we model the joint distribution of (Y_j, X_1, X_2) in terms of two bivariate copulas rather than one trivariate copula. There is much known about properties of parametric bivariate copula families in terms of dependence and tail behavior. Note that the copula C_{X_1j} links the jth response to the first latent variable X_1, and the copula C_{X_2j} links the jth response to the second latent variable X_2 conditional on X_1. Figure 1 depicts a graphical representation of the model. Our general statistical model allows for selection of C_{X_1j} and C_{X_2j} independently among a variety of parametric copula families, i.e., there are no constraints in the choices of parametric copulas $\{C_{X_1j}, C_{X_2j} : j = 1, \ldots, d\}$.
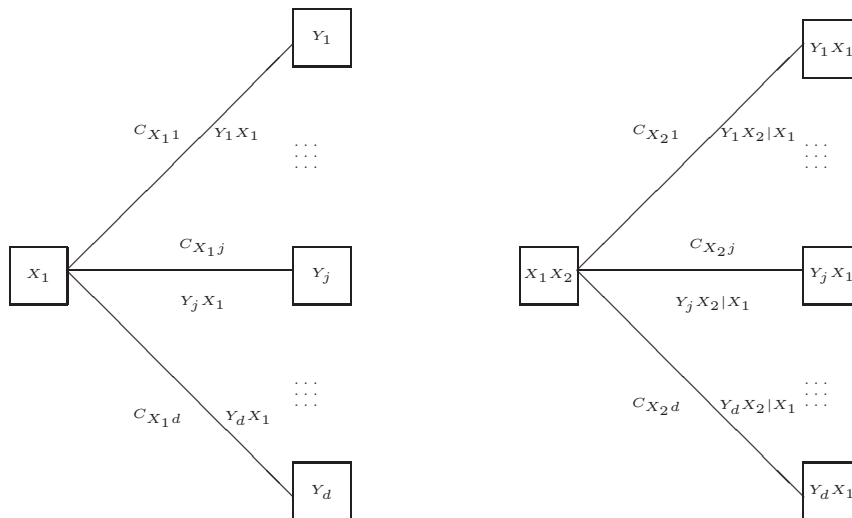
FIGURE 1.
Graphical representation of the factor copula model.

2.3. Normal ogive model as a special case

In this subsection, we show what happens when all the bivariate copulas are normal copulas. The resulting model is the same as the discretized MVN model with a p-factor correlation matrix, also known as the p-dimensional normal ogive model (Jöreskog and Moustaki, 2001). The relationship is only shown for one and two factors but holds for $p \geq 3$ factors.

The bivariate normal copula cdf from (1) is

$$C(u_1, u_2; \theta) = \Phi_2\Big(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta\Big), \quad -1 \leq \theta \leq 1,$$

where Φ_2 is the bivariate standard normal cdf with correlation parameter θ, and Φ is the

cdf of the univariate standard normal. Then, for $j = 1, \ldots, d$,

$$C_{X_1 j}(x, a_{j,y+1}) = \Phi_2\Big(\Phi^{-1}(x), \Phi^{-1}(a_{j,y+1}); \theta_j\Big) \quad \text{and} \quad C_{j|X_1}(a_{j,y+1}|x) = \Phi\left(\frac{\alpha_{j,y+1} - \theta_j \Phi^{-1}(x)}{\sqrt{1 - \theta_j^2}}\right).$$

Hence, the pmf for the 1-factor model in (3) becomes

$$\int_0^1 \prod_{j=1}^d \left\{ \Phi\left(\frac{\alpha_{j,y_j+1} - \theta_j \Phi^{-1}(x)}{\sqrt{1 - \theta_j^2}}\right) - \Phi\left(\frac{\alpha_{j,y_j} - \theta_j \Phi^{-1}(x)}{\sqrt{1 - \theta_j^2}}\right) \right\} dx$$

$$= \int_{-\infty}^{\infty} \prod_{j=1}^d \left\{ \Phi\left(\frac{\alpha_{j,y_j+1} - \theta_j z}{\sqrt{1 - \theta_j^2}}\right) - \Phi\left(\frac{\alpha_{j,y_j} - \theta_j z}{\sqrt{1 - \theta_j^2}}\right) \right\} \phi(z)\, dz.$$

This model is the same as the variant of Samejima's (1969) graded response IRT model, known as normal ogive model (McDonald, 1997) with a 1-factor correlation matrix $R = (r_{jk})$ with $r_{jk} = \theta_j \theta_k$ for $j \neq k$. The unidimensional (1-factor) normal ogive model is

$$Z_j = \theta_j Z + \sqrt{1 - \theta_j^2}\, \epsilon_j, \quad j = 1, \ldots, d,$$

where $Z, \epsilon_1, \ldots, \epsilon_d$ are independent $N(0,1)$ random variables, $-1 \leq \theta_j \leq 1$ for all j, and the connection between the latent variable Z_j and the ordinal variable Y_j is $Y_j = y$ if $\alpha_{j,y} < Z_j \leq \alpha_{j,y+1}$, with $\alpha_{j,K} = \infty$ and $\alpha_{j,0} = -\infty$.

For the 2-factor model,

$$C_{j|X_2}(F_{j|X_1}(y|x_1)|x_2) = \Phi\left(\left[\frac{\alpha_{j,y+1} - \theta_j \Phi^{-1}(x_1)}{\sqrt{1 - \theta_j^2}} - \delta_j \Phi^{-1}(x_2)\right] \bigg/ \sqrt{1 - \delta_j^2}\right)$$

$$= \Phi\left(\frac{\alpha_{j,y+1} - \theta_j \Phi^{-1}(x_1) - \delta_j \sqrt{1 - \theta_j^2}\, \Phi^{-1}(x_2)}{\sqrt{(1 - \theta_j^2)(1 - \delta_j^2)}}\right),$$

for $j = 1, \ldots, d$. Hence, the pmf for the 2-factor model in (4) becomes

$$\int \int \prod_{j=1}^d \left\{ \Phi\left(\frac{\alpha_{j,y_j+1} - \theta_j z_1 - \delta_j \sqrt{1 - \theta_j^2}\, z_2}{\sqrt{(1 - \theta_j^2)(1 - \delta_j^2)}}\right) - \Phi\left(\frac{\alpha_{j,y_j} - \theta_j z_1 - \delta_j \sqrt{1 - \theta_j^2}\, z_2}{\sqrt{(1 - \theta_j^2)(1 - \delta_j^2)}}\right) \right\} \phi(z_1)\phi(z_2)\, dz_1 dz_2.$$

This model is the same as the bidimensional (2-factor) normal ogive model with a 2-factor correlation matrix $R = (r_{jk})$ with $r_{jk} = \theta_j \theta_k + \delta_j \delta_k [(1 - \theta_j^2)(1 - \theta_k^2)]^{1/2}$ for $j \neq k$. The bidimensional (2-factor) normal ogive model (e.g., Joreskog and Moustaki, 2001) is

$$Z_j = \gamma_{j_1} Z_{01} + \gamma_{j2} Z_{02} + \sqrt{1 - \gamma_{j1}^2 - \gamma_{j2}^2}\, \epsilon_j, \quad j = 1, \ldots, d,$$

where,

$$\gamma_{j1} = \theta_j \text{ and } \gamma_{j2} = \delta_j (1 - \theta_j^2)^{1/2}, \tag{5}$$

$Z_{01}, Z_{02}, \epsilon_1, \ldots, \epsilon_d$ are independent $N(0,1)$ random variables and $0 \leq \gamma_{j1}^2 + \gamma_{j2}^2 \leq 1$ for all j. The parameter θ_j of $C_{X_1 j}$ is the correlation of Z_j with Z_{01}, and the parameter δ_j of $C_{X_2 j}$ is the partial correlation between Z_j and Z_{02} given Z_{01}. The off-diagonal entries of the

correlation matrix have the form $\gamma_{j1}\gamma_{k1} + \gamma_{j2}\gamma_{k2}$, $j \neq k$. This extends to a p-factor model ($p < d$) with correlation matrix $\Gamma\Gamma^\top + \Psi$, where Γ is a $d \times p$ matrix of loadings and Ψ is a diagonal matrix of "uniqueness" variances. The loading matrix Γ is not unique, because of invariance to rotations, $\Gamma\Gamma^\top = \Gamma P P^\top \Gamma^\top$, where P is an orthogonal $p \times p$ matrix. For the bidimensional (2-factor) normal ogive model the $2d$ parameters can be converted to $2d - 1$ parameters yielding the same latent correlation matrix, by applying a Givens (orthogonal) rotation $P = \begin{pmatrix} \omega & \sqrt{1-\omega^2} \\ -\sqrt{1-\omega^2} & \omega \end{pmatrix}$ with appropriate ω so that $(\gamma_{j1}, \gamma_{j2})P = (\gamma_{j1}^*, 0)$ so that $\gamma_{j2}^* = 0$ and $\delta_j^* = 0$ after the rotation.

2.4. Other choices of bivariate copulas with latent variables

We next proceed with a summary of choices of some other bivariate copula families that can be used when considering latent maxima, minima or mixtures. In the descriptions below, a bivariate copula C is *reflection symmetric* if its density $c(u_1, u_2) = \partial^2 C(u_1, u_2)/\partial u_1 \partial u_2$ satisfies $c(u_1, u_2) = c(1 - u_1, 1 - u_2)$ for all $0 \leq u_1, u_2 \leq 1$. Otherwise, it is reflection asymmetric often with more probability in the joint upper tail or joint lower tail. *Upper tail dependence* means that $c(1 - u, 1 - u) = O(u^{-1})$ as $u \to 0$ and *lower tail dependence* means that $c(u, u) = O(u^{-1})$ as $u \to 0$. Copulas that arise from multivariate extreme value theory have upper tail dependence. Under some regularity conditions, no upper tail dependence means that $c(1 - u, 1 - u) = O(u^{\kappa-2})$ as $u \to 0$ for some $\kappa > 1$. If $(U_1, U_2) \sim C$ for a bivariate copula C, then $(1 - U_1, 1 - U_2) \sim C_R$, where $C_R(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$ is the survival copula of C; this "reflection" of each uniform $U(0, 1)$ random variable about $1/2$ changes the direction of tail asymmetry.

Choices that we consider are the following:

(a) Reflection asymmetric copula family with upper tail dependence, such as the Gumbel extreme value copula with cdf

$$C(u_1, u_2; \theta) = \exp\left[-\left\{(-\log u_1)^\theta + (-\log u_2)^\theta\right\}^{1/\theta}\right], \quad \theta \geq 1.$$

The resulting model in this case has latent (ordinal) variables that can be considered as (discretized) maxima, and there is more probability in the joint upper tail.

(b) Reflection asymmetric copula family with lower tail dependence, such as the survival Gumbel (s.Gumbel) copula with cdf

$$C(u_1, u_2; \theta) = u_1 + u_2 - 1 + \exp\left[-\left\{\left(-\log(1 - u_1)\right)^\theta + \left(-\log(1 - u_2)\right)^\theta\right\}^{1/\theta}\right], \quad \theta \geq 1.$$

The resulting model in this case has latent (ordinal) variables that can be considered as (discretized) minima, and there is more probability in the joint lower tail.

(c) Copulas with reflection symmetric upper and lower tail dependence, such as the bivariate Student t_ν copula with cdf

$$C(u_1, u_2; \theta) = T_2\left(T^{-1}(u_1; \nu), T^{-1}(u_2; \nu); \theta, \nu\right), \quad -1 \leq \theta \leq 1,$$

where $T(; \nu)$ is the univariate Student t cdf with (non-integer) ν degrees of freedom, and T_2 is the cdf of a bivariate Student t distribution with ν degrees of freedom and correlation parameter θ. The resulting model in this case has latent (ordinal) variables

that can be considered as mixtures of (discretized) means, since the bivariate Student t distribution arises as a scale mixture of bivariate normals. A small value of ν, such as $1 \leq \nu \leq 5$, leads to a model with more probabilities in the joint upper and joint lower tails compared with the normal copula.

For the normal ogive model, the model parameters control the shape of the item characteristic curves (ICCs). For the factor copula models, ICCs can be computed (curves for one factor, surfaces for two factors), but their shape also depends on the choice of linking copulas. In Figure 2 we plot the ICCs for the item $Y_j \in \{0, 1, 2\}$,

$$\Pr(Y_j = y | X_1 = x) = C_{j|X_1}(F_j(y)|x; \theta_j) - C_{j|X_1}(F_j(y-1)|x; \theta_j), \qquad y = 0, 1, 2,$$

from the 1-factor copula model with normal, Gumbel, s.Gumbel, and t_2 copulas with the same model parameters $a_{j,1} = 1/3, a_{j,2} = 2/3$ and θ_j (in the Kendall's τ scale). We convert from $\tau_j = 0.5$ to the normal/t_2 and Gumbel/s.Gumbel parameter θ_j via the relations

$$\theta_j = \sin(\pi \tau_j / 2) \tag{6}$$

and

$$\theta_j = (1 - \tau_j)^{-1} \tag{7}$$

in Hult and Lindskog (2002) and Genest and MacKay (1986), respectively. As the form of the copulas changes, typically the middle part of the ICC is similar, but can differ more for extreme values of the latent variable because of different tail behavior of the bivariate copulas.

For this paper, the above copula families are sufficient for the applications in Section 5. Many other parametric copula families are given in Joe (1997). If one is mainly concerned with using many options for factor copula models, there are many other choices that can be considered. The above list represents a few choices with ease of interpretation for item response data.

3. Assessing goodness-of-fit

Different parametric copula models can be compared via the log-likelihood at the maximum likelihood estimate or via the Akaike information criterion. In addition, for an overall goodness-of-fit of the discrete factor models, we will use the limited information method proposed by Maydeu-Olivares and Joe (2006), and we summarize the method in this section.

Consider the set of univariate and bivariate residuals that do not include category 0. This is a residual vector of dimension $s = d(K-1) + \binom{d}{2}(K-1)^2$. For our parametric model with parameter vector $\boldsymbol{\theta}$ of dimension q, let $\boldsymbol{\pi}_2(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{\pi}}_1(\boldsymbol{\theta})^\top, \dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})^\top\right)^\top$ be the column vector of the model-based marginal probabilities with $\dot{\boldsymbol{\pi}}_1(\boldsymbol{\theta})$ the $d(K-1)$ vector of univariate marginal probabilities, and $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$ the $\binom{d}{2}(K-1)^2$ vector of bivariate marginal probabilities. Also, let $\mathbf{p}_2 = (\dot{\mathbf{p}}_1^\top, \dot{\mathbf{p}}_2^\top)^\top$ be the vector of the observed sample proportions, with $\dot{\mathbf{p}}_1$ the $d(K-1)$ vector of univariate marginal proportions, and $\dot{\mathbf{p}}_2$ the $\binom{d}{2}(K-1)^2$ vector of the bivariate marginal proportions.

With a sample size N, the limited information statistic M_2 is given by

$$M_2 = M_2(\hat{\boldsymbol{\theta}}) = N\left(\mathbf{p}_2 - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}})\right)^\top \mathbf{C}_2(\hat{\boldsymbol{\theta}})\left(\mathbf{p}_2 - \boldsymbol{\pi}_2(\hat{\boldsymbol{\theta}})\right), \tag{8}$$
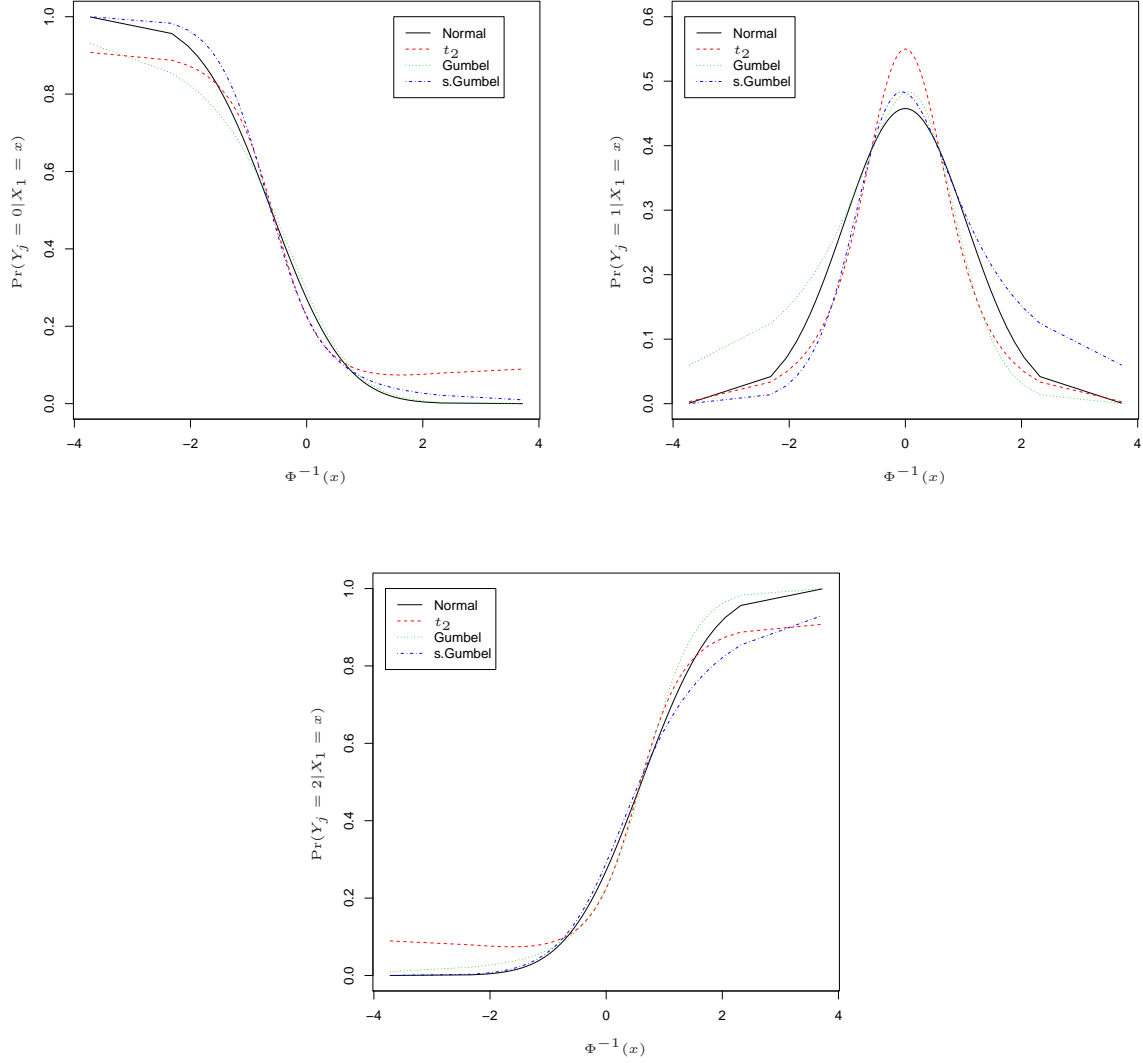
FIGURE 2.
Shape of the ICCs for item $Y_j \in \{0, 1, 2\}$ from the 1-factor copula model with normal, Gumbel, s.Gumbel, and t_2 copulas with the same model parameters $a_{j,1} = 1/3, a_{j,2} = 2/3$ and θ_j (in the Kendall's τ scale).

with

$$\mathbf{C}_2(\boldsymbol{\theta}) = \boldsymbol{\Xi}_2^{-1} - \boldsymbol{\Xi}_2^{-1}\boldsymbol{\Delta}_2(\boldsymbol{\Delta}_2^\top\boldsymbol{\Xi}_2^{-1}\boldsymbol{\Delta}_2)^{-1}\boldsymbol{\Delta}_2^\top\boldsymbol{\Xi}_2^{-1} = \boldsymbol{\Delta}_2^{(c)}\big([\boldsymbol{\Delta}_2^{(c)}]^\top\boldsymbol{\Xi}_2\boldsymbol{\Delta}_2^{(c)}\big)^{-1}[\boldsymbol{\Delta}_2^{(c)}]^\top, \quad (9)$$

where $\boldsymbol{\Delta}_2 = \partial\boldsymbol{\pi}_2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top$ is an $s \times q$ matrix with the derivatives of all the univariate and bivariate marginal probabilities with respect to the model parameters, $\boldsymbol{\Delta}_2^{(c)}$ is an $s \times (s-q)$ orthogonal complement to $\boldsymbol{\Delta}_2$, such that $[\boldsymbol{\Delta}_2^{(c)}]^\top\boldsymbol{\Delta}_2 = \mathbf{0}$, and $\boldsymbol{\Xi}_2 = N\mathrm{Cov}(\mathbf{p}_2)$ is the $s \times s$ covariance matrix of all the univariate and bivariate marginal sample proportions, excluding category 0. Due to equality in (9), $\mathbf{C}_2$ is invariant to the choice of orthogonal complement. The limited information statistic M_2 has a null asymptotic distribution that is χ^2 with $s - q$ degrees of freedom when the estimate $\hat{\boldsymbol{\theta}}$ is $\sqrt{N}$-consistent.

The covariance matrix can be derived by noticing that each Y_j has a Multinomial(1, $\pi_{j0}, \ldots, \pi_{j,K-1}$) distribution, where $\pi_{jy} = \Pr(Y_j = y)$. Hence, the joint distribution of

$Y_1, \ldots, Y_d$ is multivariate multinomial with univariate margins that are multinomial and thus

$$\boldsymbol{\Xi}_2 = \mathrm{diag}(\boldsymbol{\pi}_2(\boldsymbol{\theta})) - \boldsymbol{\pi}_2(\boldsymbol{\theta})\boldsymbol{\pi}_2(\boldsymbol{\theta})^\top.$$

Note that the computation of $\boldsymbol{\Xi}_2$ involves the computation of up to 4-dimensional probabilities. The matrix $\boldsymbol{\Delta}_2$ consists of the derivatives of the univariate and bivariate estimated probabilities with respect to the model parameters $\boldsymbol{\theta}$. In Appendix A, Table A.1 lists the derivatives of the univariate probability, $\pi_{jy} = \mathrm{Pr}(Y_j = y) = a_{j,y+1} - a_{j,y}$, with respect to the cutpoint a_{jk}, and of the bivariate probability, $\pi_{j_1 j_2, y_1 y_2} = \mathrm{Pr}(Y_{j_1} = y_1, Y_{j_2} = y_2)$, with respect to the cutpoint a_{jk} and the copula parameter θ_j for the 1-factor model; Table A.2 lists the derivatives of the bivariate probability $\pi_{j_1 j_2, y_1 y_2}$ with respect to the cutpoint a_{jk}, and the copula parameter θ_j for the first latent variable, and the copula parameter δ_j for the second latent variable for the 2-factor model for $j, j_1, j_2 = 1, \ldots, d,\quad y, y_1, y_2 = 1, \ldots, K - 1$, and $k = 1, \ldots, K - 1$. Computational details and pseudo-code for the evaluation of $\boldsymbol{\Xi}_2$ and $\boldsymbol{\Delta}_2$ are given in the Appendices B and C, respectively.

4. Behavior of log-likelihood and computational details

In this section, we discuss estimation techniques and computational details. We give details on the behavior of the log-likelihood for the 2-factor copula model and discuss cases when the model with $2d$ copulas is identifiable or has near non-identifiability.

For a sample of size N with data $\mathbf{y}_1, \ldots, \mathbf{y}_N$, the joint log-likelihood of the factor copula model is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \boldsymbol{\pi}_d(\mathbf{y}_i; \boldsymbol{\theta}). \tag{10}$$

Maximization of (10) is numerically possible but is time-consuming for large d because of many univariate and dependence parameters. When the dependence is not too strong, which is a realistic scenario for item response data, a two-step approach such as in Olsson (1979), called Inference Function of Margins (IFM) method in Joe (1997, 2005), can efficiently (in the sense of computing time and asymptotic variance) estimate the model parameters. In the first step, the univariate cutpoints are estimated as

$$\hat{a}_{j0} = p_{j0},\ \hat{a}_{j1} = p_{j0} + p_{j1},\ \ldots,\ \hat{a}_{j,K-1} = p_{j0} + p_{j1} + \ldots + p_{j,K-1}, \tag{11}$$

where p_{jy}, $y = 0, \ldots, K - 1$, for $j = 1, \ldots, d$ are the univariate sample proportions; and in the second step the joint log-likelihood (10) is maximized over the copula parameter vector with the cutpoints fixed at the estimated values from the first step. For the second step, a quasi-Newton or modified Newton-Raphson method can be used, even for large d. The likelihood estimation is especially very fast and stable for the 1-factor model. For the 2-factor model, one should try using different starting points in the maximization algorithm, when computing maximum likelihood estimates, to ensure that a global optimum is obtained. Details for the computational implementation and pseudo-code for the likelihood evaluation in (10) are given in the Appendices B and C, respectively.

4.1. Behavior of the log-likelihood for the 2-factor copula model

For the special case of the normal ogive model with 2 factors, one of the copulas in the set $\{C_{X_2 j}\}$ can be set as an independence copula without loss of generality, because the model with $2d$ parameters is not identifiable. What happens if other copulas such as Gumbel and t_ν are used for bivariate linking copulas? Is the model with $2d$ bivariate linking copulas still not identifiable? The purpose of this section is to provide answers to such questions.

Given that near non-identifiability occurs when there are large standard errors (SEs) in parameter estimates, we will compare the asymptotic covariance matrix of the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ between a model with $2d$ bivariate linking copulas and a model with $2d - 1$ bivariate linking copulas where one of the $C_{X_2 j}$'s is the independence copula. By using the asymptotic covariance matrices, we do not need time-consuming Monte Carlo simulations for comparisons; and we can quickly vary factors such as parametric family of copulas $C_{X_1 j}, C_{X_2 j}$, dimension d, $\boldsymbol{\theta}$ values, number of categories K and see the effects.

For known $\boldsymbol{\theta}$, the Fisher information matrix for the parameter $\boldsymbol{\theta}$ is given by

$$\mathcal{I} = \sum_{\mathbf{y}} \frac{\partial \pi_d(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \pi_d(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \bigg/ \pi_d(\mathbf{y}; \boldsymbol{\theta}),$$

where the inner sum is taken over all the possible vectors $\mathbf{y}$. The partial derivatives are similar in form to those in Table A.2. Under regularity conditions, the inverse of $\mathcal{I}$ is the asymptotic covariance matrix of the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ (Lehmann, 1998, p. 545). To get the magnitudes to be interpretable, we scale the asymptotic SEs (the square root of the diagonal of the inverse of $\mathcal{I}$) by $1/\sqrt{n}$, where n is the nominal sample size. Note that asymptotic SEs are not so meaningful since one does not know how large the sample size needs to be before the approximations provided by the asymptotic results are adequate in practice. Results reported by de Menezes (1999) show that the asymptotic theory may provide a very poor approximation, even if the sample size is of the order of 1000. Nevertheless, the theoretical comparison of $2d$ versus $2d - 1$ parameters is valid.

To make 2-factor copula models with $2d$ and $2d - 1$ parameters comparable and also with t_ν versus Gumbel, for a given number of ordinal categories (assumed uniformly distributed), we selected random choices of copula parameters as described below:

$2d$ t_ν **copulas:** the random choices for the parameters θ_j were simulated from $N(0.6, 0.1^2)$ and for the parameters δ_j were simulated from $N(0.3, 0.1^2)$ for $j = 1, ..., d$.

$(2d - 1)$ t_ν **copulas:** The $2d$ parameters at the preceding step were converted to $2d - 1$ parameters with the same dependence as indicated below:

  1. The factor copula parameters were converted to loadings via (5).
  2. A Givens rotation was applied in order to get the rotated loadings $\gamma_{j1}^*, \gamma_{j2}^*$, with the last loading for the second factor set to 0.
  3. The rotated loadings are converted back to new factor copula parameters θ_j^*, δ_j^* via, $\theta_j^* = \gamma_{j1}^*$ and $\delta_j^* = \gamma_{j2}^*/(1 - \theta_j^{*2})^{1/2}$.

$2d$ **Gumbel copulas:** Each parameter of $2d$ t_ν copulas was converted to Kendall's tau via the functional inverse of (6), and then from Kendall's tau to Gumbel parameter

with the relation in (7). Note that in this way both models have the same copula
parameters in the Kendall's tau scale.

$(2d - 1)$ **Gumbel copulas:** The factor copula parameters θ_j^*, δ_j^* in $2d-1$ t_ν copulas were
converted to Kendall τ's and then to Gumbel parameters. Negative τ values were
converted to positive, since Gumbel copulas can model only positive dependence.
Note that in this way the overall dependence is still about the same.

The random factor copula parameters as described above are exchangeable so that
the average of the asymptotic SEs leads to the same value for each parameter of Factor
1 and Factor 2. Table 1 summarizes the average asymptotic SEs for each factor over 50
repetitions; the asymptotic SEs are scaled using $n = 10^4$.

TABLE 1.

Asymptotic SEs for models with $2d$ and $2d - 1$ bivariate linking copulas scaled for a sample of
size $n = 10^4$; d=number of items, K=number of categories (assumes equally weighted); ASE1
and ASE2 are averages of the asymptotic SEs for the first factor and second factor, respectively.

	$d = 7, K = 3$		$d = 7, K = 6$		$d = 10, K = 3$	
Model	ASE1	ASE2	ASE1	ASE2	ASE1	ASE2
$2d$ Gumbel	0.069	0.071	0.051	0.052	0.048	0.050
$2d - 1$ Gumbel	0.036	0.066	0.028	0.048	0.029	0.046
$2d$ t_1	0.020	0.063	0.015	0.047	0.015	0.046
$2d - 1$ t_1	0.012	0.108	0.010	0.086	0.010	0.075
$2d$ t_2	0.029	0.084	0.023	0.064	0.020	0.056
$2d - 1$ t_2	0.014	0.106	0.011	0.086	0.010	0.071
$2d$ t_3	0.042	0.122	0.031	0.089	0.027	0.079
$2d - 1$ t_3	0.014	0.109	0.012	0.085	0.010	0.071
$2d$ t_4	0.063	0.184	0.045	0.129	0.040	0.119
$2d - 1$ t_4	0.015	0.111	0.012	0.086	0.010	0.071
$2d$ t_5	0.093	0.273	0.063	0.183	0.060	0.176
$2d - 1$ t_5	0.016	0.114	0.012	0.086	0.010	0.071
$2d$ t_7	0.179	0.531	0.112	0.329	0.115	0.343
$2d - 1$ t_7	0.016	0.116	0.012	0.088	0.010	0.072
$2d$ t_{10}	0.366	1.089	0.196	0.580	0.232	0.696
$2d - 1$ t_{10}	0.016	0.118	0.013	0.091	0.010	0.072
$2d$ t_{20}	0.853	2.555	0.285	0.847	0.530	1.595
$2d - 1$ t_{20}	0.017	0.121	0.013	0.095	0.010	0.072

From the results, a model with $2d$ Gumbel or t_ν copulas with $\nu \leq 3$ is clearly identifi-
able. For $2d$ t_ν copulas there is much variability for $\nu \geq 5$, and the usage of one indepen-
dence copula for the second factor leads to smaller asymptotic SEs for the first factor. Note
that for this case we have also checked, using the Jacobian of the transformation, that the
variability of factor correlations is small and similar for the various choices of the degrees
of freedom. For the factor model with $2d$ t_ν copulas with larger values of ν, the correlation
matrix based on the parameters is stable, even though the copula parameters (especially
for the second factor) are not so interpretable. Another pattern is that the asymptotic SEs
decrease as the dimension and/or the number of ordinal categories increase. The latter
means that a part of near non-identifiability is due to discretization.

These summaries agree with what we see in fit to data sets with sample size 500–1000.
For t_ν with ν small or for Gumbel, there appears to be a single global/local maximum
to the log-likelihood in that the same convergence is obtained in numerical optimization
from different starting points. What this means is that if the factor copula model with

t_ν ($\nu \le 3$) or Gumbel linking copulas is a good fit, then the "loading" parameters are identifiable and can be interpreted without a need for something like a varimax transform (Kaiser, 1958) of the loading matrix.

For t_ν with larger values of ν, when there are large standard errors (SEs) in parameter estimates and/or the latter are greater than 0.9 in the Kendall's tau scale, that might indicate that the model with $2d$ copulas is nearly non-identifiable and there is a direction for which the log-likelihood is quite flat; there are local maxima to the log-likelihood and different convergence values are obtained from different starting points. For such cases, we can set one of the $C_{X_2 j}$'s to be an independence copula, i.e., use $2d - 1$ copulas, in order to avoid the local maxima issues and make the parameter estimates interpretable.

5. Applications

In this section we illustrate the factor copula models for item response data sets. For a preliminary analysis, the bivariate copulas are applied to each of the pair of items to compare observed versus model-based bivariate counts. Then, we fit the discrete factor models with normal (this is the special case of the normal ogive model), Gumbel, s.Gumbel, and t_ν bivariate linking copulas. For Student t_ν, choices of ν were $1, 2, \ldots, 10$. For ease of interpretation, we do not mix Gumbel, t_ν and normal for a single factor; hence, for the 2-factor models we allow two different copula families, one for Factor 1 and one for Factor 2. Note that Gumbel copulas for both factors means that conditional dependence is positive only; Gumbel for Factor 1 and t_ν for Factor 2 allow for conditional negative dependence of items and latent variable 2 given latent variable 1. The conditional standard errors (SEs) of the estimated parameters are obtained by the inversion of the Hessian matrix at the second step of the IFM method. These SEs are adequate to assess the flatness of the log-likelihood. Proper SEs that account for the estimation of cutpoints can be obtained by jackknifing the two-stage estimation procedure. To make it easier to compare strengths of dependence, we convert the estimated parameters to Kendall's τ's in $(0, 1)$ via the functional inverse of (6) for elliptical copulas and via the functional inverse of (7) for Gumbel copulas; SEs are also converted via the delta method. Note that Kendall's tau only accounts for the dependence dominated by the middle of the data, and it is expected to be similar amongst different families of copulas. However, the tail dependence varies, as explained in Subsection 2.4, and is a property to consider when choosing amongst different families of copulas.

For the 1-factor model with t_ν we summarize the choice of integer ν with the largest log-likelihood, while for the 2-factor models we summarize the first three models in terms of largest likelihood along with the bidimensional normal ogive model as a baseline. For this (normal ogive) model, to obtain a unique solution we must impose sufficient constraints. One parameter for the second factor can be set to zero and the likelihood can be maximized with respect to other $2d - 1$ parameters. We report the varimax transform of the loadings (a reparametrization of $2d$ parameters), converted to factor copula parameters (and Kendall's taus), as these τ's might be more comparable with the maximum likelihood estimates from copula-based factor models. Note also that, in terms of identifiability of signs of parameters, the factor copula model based on t_ν is like that based on normal. If $\theta_j \to -\theta_j$, $j = 1, \ldots, d$ or if $\delta_j \to -\delta_j$, $j = 1, \ldots, d$, then the model is the same, because only the orientation of the

latent variable has been reversed. For simplicity, we report these correlation parameters as being positive for stronger dependence. In addition, we summarize the M_2 statistics from (8). Note that the M_2 statistics in the case with $2d - 1$ copulas (one set to independence for the second factor) are computed with $\mathbf{\Delta}_2$ having one less column.

5.1. LOT item response data

The first example deals with the Life Orientation Test (LOT) scale (Scheier and Carver, 1985) which consists of twelve 5-point rating items designed to measure generalized outcome expectancies. The sample of $N = 389$ is based on a survey of university male and female undergraduate students from Carnegie-Mellon University and the University of Miami. Students were asked to respond to each item using one of five categories: "0=strongly disagree"; "1=disagree"; "2=neutral"; "3=agree"; "4=strongly agree". Chang et al. (1994) analyzed the following items, Y_1: In uncertain times, I usually expect the best; Y_2: I always look on the bright side of things; Y_3: I'm always optimistic about my future; Y_4: I'm a believer in the idea that "every cloud has a silver lining"; Y_5: If something can go wrong for me, it will; Y_6: I hardly ever expect things to go my way; Y_7: Things never work out the way I want them to; Y_8: I rarely count on good things happening to me.

We will analyze these item response data with $N = 389$, with the scoring of the last 4 items reversed, so that all items are positively correlated; the positive correlation is needed for Gumbel copulas, which only have positive dependence. Note that the four positively worded items measure optimism, while the four negatively worded items measure pessimism.

For these items, a respondent might be thinking about the average "sensation" of many past relevant events, leading to latent means. That is, based on the item descriptions, this seems more natural than a discretized maxima or minima. Since the sample is a mixture (male and female students, and two locations) we can expect a priori that a factor copula model with t_ν copulas might be plausible, as in this case the items can be considered as mixtures of discretized means.

The summary of findings for the preliminary analysis show that there is more probability in the joint upper and lower tails than with a discretized MVN, suggesting that a factor model with bivariate Student t_ν linking copulas might provide a better fit. Some representative results, for margin $(5, 6)$, are presented in Table 2. Table 3 gives the estimated parameters, their standard errors (SE) in Kendall's tau scale, and joint log-likelihoods ℓ for the 1-factor and 2-factor models; and Table 4 presents the M_2 statistics and corresponding p-values along with the maximum deviations of observed and expected counts for each bivariate margin, that is, $D_{j_1 j_2} = N \max_{y_1, y_2} |p_{j_1, j_2, y_1, y_2} - \pi_{j_1, j_2, y_1, y_2}(\hat{\boldsymbol{\theta}})|$, to show where the fit is improving relative to the normal ogive model.

For these results, in 1-factor models, the weakest dependence can be found in the copula $C_{X_1 4}(\cdot; \theta_4)$ linking the latent variable and the fourth item. The strongest dependence can be found in the copula $C_{X_1 6}(\cdot; \theta_6)$ linking the latent variable and the sixth item. The best fit for the 1-factor model is based on t_2 linking copulas, where there is a big improvement over the normal ogive model. However, it is revealed that one latent variable is not adequate to model the dependencies among the items (Table 4, 1-factor model).

In the 2-factor models dependence parameter estimates, we report the best three

TABLE 2.
Diagnostics for Items 5 and 6 for the LOT response data, based on the fit of the bivariate normal, Gumbel, s.Gumbel, and t_5 copulas, at each of the pair of items, comparing observed versus model-based bivariate counts with an emphasis on the tails.

Y_5	Y_6	observed	normal	Gumbel	s.Gumbel	t_5
0	0	7	5	4	9	7
0	1	7	8	8	9	9
0	2	4	7	7	4	5
0	3	5	4	5	2	3
0	4	1	0	0	0	1
1	0	2	6	6	4	5
1	1	28	20	20	23	22
1	2	29	28	29	30	30
1	3	20	31	31	28	28
1	4	10	4	3	5	4
2	0	3	2	3	1	1
2	1	9	13	14	10	11
2	2	37	28	29	29	30
2	3	52	52	54	54	56
2	4	7	12	8	14	10
3	0	0	1	1	0	1
3	1	2	7	7	5	6
3	2	15	21	19	21	19
3	3	89	73	75	75	77
3	4	33	37	36	37	37
4	0	2	0	0	0	0
4	1	2	0	0	1	1
4	2	0	1	1	2	1
4	3	6	11	6	13	8
4	4	19	16	22	14	19

models based on the joint log-likelihood which are all built with bivariate t_ν copulas with $\nu = 2$ for the first factor and $\nu = 7$ or 8 or 9 for the second factor. The improvement over the bidimensional normal ogive model is substantial for these models (Table 4, 2-factor model). However, it is not so clear from the goodness-of-fit p-values, that the response patterns are satisfactorily explained by two latent variables. The discrepancy of 14/389 maximum occurs with a larger count. For the last 4 reversed items, there are more counts in vectors of the form $\{Y_5 = \cdot, Y_6 = 3, Y_7 = 3, Y_8 = 3\}$ and $\{Y_5 = \cdot, Y_6 = \cdot, Y_7 = 2, Y_8 = 2\}$ than expected with the models and less in $\{Y_5 = \cdot, Y_6 = \cdot, Y_7 = 3, Y_8 = 2\}$. The overall model fit is not bad except for this. Note that either a discrete 3-factor model or a 3-factor model with a special structure known as bifactor model (see, e.g., Gibbons and Hedeker, 1992) does not improve the fit.

In this example, it is highlighted that a factor model with t_ν copulas is plausible for a population that is a mixture of subpopulations, while a normal model might be adequate for smaller homogeneous subgroups. For the 2-factor model based on t_ν copulas with small ν, note that, without the need for a varimax rotation, the unique loading matrix shows that one factor is loaded mainly on Items 1–4 and the other factor is loaded mainly on Items 5–8. This was expected from the 'optimistically' and 'pessimistically' worded items.

5.2. Science item response data

This data set comes from the Consumer Protection and Perceptions of Science and Technology section of the 1992 Euro-Barometer Survey based on a sample of $N = 392$ from

Table 3.
Estimated parameters (Est.), their standard errors (SE) on Kendall's tau scale, and joint log-likelihoods for the discrete factor models built by normal, Gumbel, s.Gumbel, and t_ν bivariate linking copulas for the LOT item response data.

1-factor	normal		Gumbel		s.Gumbel		t_2	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
θ_1	0.27	0.03	0.24	0.04	0.28	0.03	0.24	0.04
θ_2	0.34	0.03	0.31	0.04	0.35	0.03	0.34	0.04
θ_3	0.33	0.03	0.33	0.04	0.32	0.03	0.33	0.04
θ_4	0.21	0.04	0.20	0.04	0.20	0.04	0.19	0.04
θ_5	0.43	0.03	0.46	0.03	0.45	0.03	0.47	0.04
θ_6	0.66	0.03	0.72	0.03	0.66	0.03	0.74	0.03
θ_7	0.63	0.03	0.67	0.03	0.61	0.03	0.65	0.03
θ_8	0.50	0.03	0.54	0.03	0.51	0.03	0.58	0.03
ℓ	-3902.1		-3895.1		-3863.5		-3799.4	

2-factor	normal	t_2/t_7		t_2/t_8		t_2/t_9	
	Est.	Est.	SE	Est.	SE	Est.	SE
θ_1	0.09	0.13	0.05	0.13	0.05	0.13	0.05
θ_2	0.11	0.18	0.06	0.17	0.06	0.17	0.06
θ_3	0.15	0.20	0.05	0.20	0.05	0.20	0.05
θ_4	0.10	0.13	0.05	0.13	0.04	0.13	0.04
θ_5	0.43	0.49	0.04	0.49	0.04	0.49	0.04
θ_6	0.61	0.74	0.04	0.74	0.04	0.74	0.04
θ_7	0.66	0.70	0.03	0.70	0.03	0.69	0.03
θ_8	0.44	0.54	0.04	0.54	0.04	0.54	0.04
δ_1	0.46	0.44	0.04	0.44	0.04	0.44	0.04
δ_2	0.65	0.64	0.05	0.64	0.05	0.64	0.05
δ_3	0.47	0.46	0.04	0.46	0.04	0.46	0.04
δ_4	0.26	0.25	0.04	0.25	0.04	0.25	0.04
δ_5	0.12	0.07	0.07	0.07	0.07	0.07	0.06
δ_6	0.33	0.25	0.12	0.25	0.11	0.26	0.11
δ_7	0.23	0.14	0.11	0.15	0.10	0.15	0.10
δ_8	0.24	0.19	0.07	0.20	0.07	0.20	0.07
ℓ	-3800.7	-3694.4		-3694.4		-3694.4	

Great Britain; see Rizopoulos (2011) and the references therein. The questions (items) asked are, Y_1: Science and technology are making our lives healthier, easier and more comfortable; Y_2: Scientific and technological research cannot play an important role in protecting the environment and repairing it; Y_3: The application of science and new technology will make work more interesting; Y_4: Thanks to science and technology, there will be more opportunities for the future generations; Y_5: New technology does not depend on basic scientific research; Y_6: Scientific and technological research do not play an important role in industrial development; Y_7: The benefits of science are greater than any harmful effect it may have. All of the items were measured on a four-group scale with response categories "0=strongly disagree", "1=disagree to some extent", "2=agree to some extent" and "3=strongly agree".

For some items such as the first three, it is plausible that a respondent might be thinking about the maximum benefit (or a high quantile) of many past events. For example, for Item 1, a participant might reflect on past relevant events where science and technology improved her/his life; then by considering the best case, i.e., the event where the positive effect of science and technology in her/his life was substantial, s/he chooses an appropriate ordinal response. For some of the other items, one might consider a median or less extreme benefit of past relevant events. To sum up, the items appear to be a mixed selection between

TABLE 4.

M_2 statistics and corresponding p-values along with some summaries for bivariate margins $(D_{j_1 j_2})$ for the discrete factor models built by normal, Gumbel, s.Gumbel, and t_ν bivariate linking copulas for the LOT item response data. Sample size $N = 389$.

Model	1-factor				2-factor			
	normal	Gumbel	s. Gumbel	t_2	normal	t_2/t_7	t_2/t_8	t_2/t_9
M_2	1178.8	1074.0	1115.9	809.9	959.2	573.6	573.3	573.1
df	440	440	440	440	433	432	432	432
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
D_{12}	20	21	18	18	9	6	6	6
D_{13}	15	17	14	14	10	8	8	8
D_{14}	8	9	8	7	5	5	5	5
D_{15}	10	9	11	11	10	12	11	11
D_{16}	9	10	8	8	10	9	9	9
D_{17}	5	7	5	8	5	7	7	7
D_{18}	12	11	11	9	12	9	9	9
D_{23}	21	22	19	17	10	6	7	7
D_{24}	13	14	12	11	8	9	8	8
D_{25}	13	13	11	9	15	11	11	11
D_{26}	7	10	7	9	8	7	7	7
D_{27}	10	12	8	11	13	12	12	12
D_{28}	9	9	9	8	9	8	8	8
D_{34}	7	6	8	5	5	4	4	4
D_{35}	5	5	4	5	7	5	5	5
D_{36}	15	12	13	6	17	7	7	7
D_{37}	4	9	5	8	5	7	7	7
D_{38}	10	10	10	10	10	10	10	10
D_{45}	7	8	7	8	9	8	8	8
D_{46}	5	6	5	7	6	7	7	7
D_{47}	10	9	10	10	11	11	11	11
D_{48}	6	6	7	4	6	4	4	4
D_{56}	17	14	14	11	16	9	9	9
D_{57}	13	11	11	9	11	7	7	7
D_{58}	13	12	10	8	13	7	7	7
D_{67}	24	18	24	14	19	11	11	11
D_{68}	25	17	22	11	25	11	11	11
D_{78}	22	15	20	15	21	14	14	14

averages and maxima so that a factor model with more probability in the joint upper tail might be an improvement over a factor model based on a discretized MVN.

The summary of findings for the preliminary analysis show that there is more probability in the upper tail compared with a discretized MVN, suggesting that a factor model with bivariate Gumbel linking copulas might provide a better fit. Some representative results, for margin $(3, 7)$, are presented in Table 5. Note that the bivariate Gumbel copula failed to model the upper tail in some bivariate margins. This result might indicate that a 2-factor model that combines Gumbel (more probability in the upper tail) with t_ν copulas, to add some additional dependence in both tails, might provide the better fit. Table 6 gives the estimated parameters, their standard errors (SE) in Kendall's tau scale, joint log-likelihoods ℓ, M_2 statistics and corresponding p-values for the 1-factor and 2-factor models.

For these results, in 1-factor models the weakest dependence can be found in the copula $C_{X_1 5}(\cdot; \theta_5)$ linking the latent variable and the fifth item. The strongest dependence can be found in the copula $C_{X_1 4}(\cdot; \theta_4)$ linking the latent variable and the fourth item. The best fit for the 1-factor model is based on t_2 linking copulas, where there is a big

TABLE 5.

Diagnostics for Items 3 and 7 for the Science response data, based on the fit of the bivariate normal, Gumbel, s.Gumbel, and t_5 copulas, at each of the pair of items, comparing observed versus model-based bivariate counts with an emphasis on the tails.

Y_3	Y_7	observed	normal	Gumbel	s.Gumbel	t_5
0	0	4	3	3	5	5
0	1	8	11	11	11	11
0	2	12	15	15	13	12
0	3	9	3	4	4	5
1	0	6	7	7	6	6
1	1	34	29	30	29	31
1	2	47	47	48	47	47
1	3	11	15	13	16	14
2	0	8	9	10	8	7
2	1	52	50	50	49	49
2	2	111	104	106	106	111
2	3	35	43	40	44	40
3	0	3	1	1	2	3
3	1	6	10	9	11	10
3	2	23	28	24	28	24
3	3	23	17	21	14	19

improvement over the normal ogive model. However, it is revealed that one latent variable is not adequate to model the dependencies among the items (Table 6, 1-factor model).

In 2-factor models, the best models in terms of the log-likelihood principle are provided when we use Gumbel for one factor and t_ν ($2 \leq \nu \leq 3$) copulas for the other factor. The improvement over the bidimensional normal ogive model is substantial according to the M_2 statistics presented in Table 6. Overall the (residual) dependence after a latent variable for maxima is quite strong, as the overall dependence of the factor with t_ν copulas is strong and varying in sign. The latent variable for maxima is positively associated with all items, while the other latent variable is positively associated with Items 1,3,4,7 and negatively associated with the remaining items. The combination of Gumbel and t_ν copulas can handle negative (conditional) association, while this is not the case for a model with Gumbel copulas for both factors. A 2-factor model with Gumbel linking copulas for both factors was not a good fit because some parameters for Factor 2 tended to the independence copula, which suggests some negative conditional dependence. Note here that the varimax solution of the normal ogive model is closer to the 2-factor copula model with Gumbel for Factor 2.

5.3. Environment item response data

This data set comes from the Environment section of the 1990 British Social Attitudes Survey; see Rizopoulos (2011) and the references therein. All of the items were measured on a three-group scale with response categories "0=very concerned", "1=slightly concerned" and "2=not very concerned". A sample of 291 responded to the questions, Y_1: Lead from petrol; Y_2: River and sea pollution; Y_3: Transport and storage of radioactive waste; Y_4: Air pollution; Y_5: Transport and disposal of poisonous chemicals; Y_6: Nuclear Risks from nuclear power station.

For all the items, it seems plausible that latent minima are involved. The items are on a "reversed" ordinal scale; for an item, a respondent might be thinking about the

TABLE 6.

Estimated parameters, their standard errors (SE) in Kendall's tau scale, and joint log-likelihoods ℓ, along with M_2 statistics and corresponding p-values, for the discrete factor models built by normal, Gumbel, s.Gumbel, and t_ν bivariate linking copulas for the Science item response data.

1-factor	normal		Gumbel		s.Gumbel		t_2	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
θ_1	0.32	0.05	0.32	0.05	0.38	0.07	0.34	0.06
θ_2	-0.01	0.05	0.07	0.04	0.01	0.03	0.07	0.06
θ_3	0.37	0.04	0.37	0.04	0.30	0.06	0.34	0.06
θ_4	0.59	0.07	0.60	0.08	0.43	0.07	0.52	0.07
θ_5	-0.01	0.05	0.05	0.04	0.00	0.05	0.06	0.06
θ_6	0.09	0.05	0.16	0.05	0.09	0.05	0.18	0.06
θ_7	0.33	0.05	0.34	0.05	0.36	0.06	0.38	0.06
ℓ	-3002.0		-2992.7		-3011.1		-2957.0	
M_2	540.4		500.2		580.7		401.7	
df	182		182		182		182	
p-value	< 0.001		< 0.001		< 0.001		< 0.001	

2-factor	normal	Gumbel/t_2		t_2/Gumbel		t_3/Gumbel	
	Est.	Est.	SE	Est.	SE	Est.	SE
θ_1	0.32	0.27	0.05	0.22	0.07	0.24	0.06
θ_2	-0.03	0.36	0.05	-0.18	0.07	-0.17	0.07
θ_3	0.38	0.15	0.05	0.36	0.06	0.37	0.05
θ_4	0.58	0.28	0.06	0.47	0.07	0.50	0.07
θ_5	-0.03	0.36	0.06	-0.24	0.08	-0.22	0.08
θ_6	0.09	0.44	0.05	-0.08	0.08	-0.06	0.07
θ_7	0.34	0.21	0.05	0.32	0.06	0.32	0.06
δ_1	0.13	0.20	0.07	0.24	0.06	0.24	0.06
δ_2	0.46	-0.31	0.07	0.42	0.05	0.43	0.05
δ_3	-0.09	0.36	0.06	0.13	0.07	0.11	0.07
δ_4	-0.01	0.49	0.07	0.30	0.09	0.30	0.09
δ_5	0.49	-0.37	0.07	0.47	0.06	0.48	0.06
δ_6	0.44	-0.21	0.09	0.50	0.05	0.50	0.05
δ_7	0.02	0.30	0.06	0.19	0.06	0.17	0.06
ℓ	-2921.9	-2864.7		-2866.3		-2866.7	
M_2	296.8	169.9		169.9		175.0	
df	176	175		175		175	
p-value	< 0.001	0.59		0.59		0.49	

maximum concern of many past relevant events and then discretize the negative of the latent maximum. Since each form of the pollution of the environment is very important, the most harmful case of each form of the pollution in the environment might be reflected on and then converted to an ordinal scale with the low category representing the most concern. Hence, we expect that a factor copula model with survival Gumbel copulas might be plausible, as the ordinal responses in this case are discretized minima.

The summary of findings for the preliminary analysis show that there is more joint probability in the joint lower tail, suggesting that a factor model with bivariate s.Gumbel linking copulas might provide a better fit. Some representative results, for margin $(1,5)$, are presented in Table 7. Table 8 gives the estimated parameters, their standard errors (SE) in Kendall's tau scale, joint log-likelihoods ℓ, M_2 statistics and corresponding p-values for the 1-factor and 2-factor models.

The best fit for the 1-factor model is based on t_8 linking copulas, where there is a slight improvement over the normal ogive model. However, it is revealed that one latent variable is not adequate to model the dependencies among the items (Table 8, 1-factor model).

TABLE 7.

Diagnostics for Items 1 and 5 of the Environment response data, based on the fit of the bivariate normal, Gumbel, s.Gumbel, and t_5 copulas at each of the pair of items, comparing observed versus model-based bivariate counts with an emphasis on the tails.

Y_1	Y_5	observed	normal	Gumbel	s.Gumbel	t_5
0	0	153	151	149	152	151
0	1	22	24	25	22	24
0	2	4	4	4	5	5
1	0	57	60	63	59	61
1	1	27	26	25	28	27
1	2	11	9	7	9	7
2	0	8	7	6	7	7
2	1	7	6	5	6	5
2	2	2	4	6	3	5

TABLE 8.

Estimated parameters (Est.), their standard errors (SE) in Kendall's tau scale, and joint log-likelihoods ℓ, along with M_2 statistics and corresponding p-values, for the discrete factor models built by normal, Gumbel, s.Gumbel, and t_ν for the Environment item response data.

1-factor	normal		Gumbel		s.Gumbel		t_8	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
θ_1	0.42	0.04	0.36	0.04	0.47	0.05	0.42	0.05
θ_2	0.58	0.05	0.49	0.05	0.65	0.05	0.58	0.05
θ_3	0.67	0.04	0.60	0.05	0.74	0.04	0.67	0.05
θ_4	0.67	0.04	0.62	0.05	0.72	0.04	0.68	0.05
θ_5	0.66	0.04	0.58	0.05	0.73	0.04	0.66	0.04
θ_6	0.52	0.04	0.47	0.04	0.55	0.04	0.52	0.04
ℓ	-1093.3		-1098.4		-1092.8		-1092.7	
M_2	120.6		140.5		115.4		119.8	
df	54		54		54		54	
p-value	< 0.001		< 0.001		< 0.001		< 0.001	
2-factor	normal		s.Gum./s.Gum		s.Gumbel/t_4		s.Gumbel/t_3	
	Est.		Est.	SE	Est.	SE	Est.	SE
θ_1	0.45		0.54	0.06	0.51	0.08	0.51	0.08
θ_2	0.57		0.70	0.06	0.66	0.11	0.65	0.12
θ_3	0.24		0.52	0.17	0.25	0.15	0.22	0.16
θ_4	0.58		0.82	0.16	0.59	0.12	0.57	0.13
θ_5	0.26		0.52	0.15	0.30	0.14	0.28	0.15
θ_6	0.18		0.38	0.15	0.21	0.12	0.19	0.13
δ_1	0.21		0.00	0.24	0.21	0.14	0.22	0.14
δ_2	0.41		0.18	0.29	0.48	0.17	0.51	0.16
δ_3	0.73		0.72	0.13	0.77	0.10	0.78	0.10
δ_4	0.58		0.33	0.50	0.60	0.12	0.61	0.11
δ_5	0.64		0.64	0.12	0.63	0.07	0.64	0.07
δ_6	0.55		0.52	0.11	0.53	0.07	0.52	0.07
ℓ	-1072.3		-1068.0		-1069.3		-1069.4	
M_2	66.6		60.8		59.5		59.1	
df	49		48		48		48	
p-value	0.048		0.102		0.123		0.132	

The best fitted 2-factor models result when we use s.Gumbel for Factor 1 and s.Gumbel or t_3 or t_4 copulas for Factor 2 (s.Gumbel or s.Gumbel/t_3 or s.Gumbel/t_4). Note that, for the conditional independence model with s.Gumbel copulas, some SEs are large and the first parameter for the second factor is close to subindependence. Thus, we have also used $2d - 1$ s.Gumbel copulas with the $C_{X_2 1}$ copula set to independence; but the results were similar and, hence, are not reported. The improvement over the bidimensional normal

ogive model is much smaller in this data set, but for the M_2 statistics there is enough improvement to change a P-value < 0.05 to one > 0.05 (Table 8, 2-factor model).

Interpreting the estimated parameters, the first latent variable is moderately positively associated with all items and the second latent variable is moderately to highly positively associated with almost all items. Similar interpretation comes from the varimax solution of the normal ogive model.

6. Discussion

We have proposed factor or conditional independence models where we replace bivariate normal distributions, between observed and latent variables, with bivariate copulas. It is the most general factor model, with univariate parameters separated from dependence parameters. Factor copula models use bivariate copulas linking observed variables and unobserved factors or latent variables, rather than logistic or normal distributions conditional on latent variables. Our factor copula construction includes the normal ogive model as a special case and can provide a substantial improvement over the latter (normal ogive model) based on log-likelihood and goodness-of-fit. This improvement relies on the fact that when we use appropriate bivariate copulas other than normal copulas in the construction, there is an interpretation of latent variables that can be maxima/minima or high/low quantiles instead of means.

Building on the basic model proposed in this paper, there are several extensions that can be implemented. The adoption of the special structure of the bifactor model (e.g., Gibbons and Hedeker, 1992) is straightforward. This has been implemented for the LOT dataset but the fit was not improved. However, this special 3-factor dependence structure may prove useful in other applications. The discrete factor model can also easily be extended to other types of discrete data and to inclusion of covariates. One can carry out the model building process in two steps: (a) the selection of the univariate margins as functions of some explanatory variables and, (b) the selection of the bivariate copulas to build the factor copula model according to the actual dependence among the discrete responses. The estimation via the IFM method is convenient for fitting the univariate marginal distributions separately from fitting the factor model dependence structure.

Another direction of future research is to extend our factor model to capture the residual dependence as in Braeken et al. (2007). Therein multivariate Archimedean copulas have been used for subgroups of items that are chosen from the context. Without a priori knowledge of the subgroups of items, a more general approach makes use of truncated vine copula models, conditional on latent variables, to model the residual dependence with $O(d)$ dependence parameters for d items. Truncated vine models are studied in Brechmann et al. (2012) for continuous response variables and in Panagiotelis et al. (2012) for discrete response variables.

Acknowledgments

Appendix A. Derivatives needed for the M_2 statistic

We summarize the integral form of the derivatives needed for the M_2 statistic in Tables A.1 and A.2.

TABLE A.1.

Derivatives of the univariate probability π_{jy} with respect to the cutpoint a_{jk}, and of the bivariate probability $\pi_{j_1j_2,y_1y_2}$ with respect to the cutpoint a_{jk} and the copula parameter θ_j for the 1-factor model for $j, j_1, j_2 = 1, \ldots, d$, $y, y_1, y_2 = 1, \ldots, K-1$, and $k = 1, \ldots, K-1$. Note that $c_{X_1j}(x,a) = \partial^2 C_{X_1j}(x,a)/\partial x \partial a$, $\dot{C}_{j|X_1}(\cdot;\theta_j) = \partial C_{j|X_1}(\cdot;\theta_j)/\partial\theta_j$, $\dot{f}_{j|X_1}(y|x) = \partial f_{j|X_1}(y|x)/\partial\theta_j = \dot{C}_{j|X_1}(a_{j,y+1}|x;\theta_j) - \dot{C}_{j|X_1}(a_{j,y}|x;\theta_j)$.

$\partial\pi_{j_1y}/\partial a_{jk}$	If
1	$a_{jk} = a_{j_1,y+1}$
-1	$a_{jk} = a_{j_1,y}$
0	elsewhere
$\partial\pi_{j_1j_2,y_1y_2}/\partial a_{jk}$	If
$\int_0^1 c_{X_1j_1}(x,a_{j_1,y_1+1};\theta_{j_1})f_{j_2|X_1}(y_2|x;\theta_{j_2})\,dx$	$a_{jk} = a_{j_1,y_1+1}$
$-\int_0^1 c_{X_1j_1}(x,a_{j_1,y_1};\theta_{j_1})f_{j_2|X_1}(y_2|x;\theta_{j_2})\,dx$	$a_{jk} = a_{j_1,y_1}$
$\int_0^1 c_{X_1j_2}(x,a_{j_2,y_2+1};\theta_{j_2})f_{j_1|X_1}(y_1|x;\theta_{j_1})\,dx$	$a_{jk} = a_{j_2,y_2+1}$
$-\int_0^1 c_{X_1j_2}(x,a_{j_2,y_2};\theta_{j_2})f_{j_1|X_1}(y_1|x;\theta_{j_1})\,dx$	$a_{jk} = a_{j_2,y_2}$
0	elsewhere
$\partial\pi_{j_1j_2,y_1y_2}/\partial\theta_j$	If
$\int_0^1 \dot{f}_{j_1|X_1}(y_1|x;\theta_{j_1})f_{j_2|X_1}(y_2|x;\theta_{j_2})\,dx$	$j = j_1$
$\int_0^1 \dot{f}_{j_2|X_1}(y_2|x;\theta_{j_2})f_{j_1|X_1}(y_1|x;\theta_{j_1})\,dx$	$j = j_2$
0	elsewhere

Appendix B. Numerical integration

For the log-likelihood and goodness-of-fit computations, Gauss-Legendre quadrature (Stroud and Secrest, 1966) is used to evaluate one-dimensional integrals for all quantities for the 1-factor model. For example,

$$\pi_d(\mathbf{y}) = \int_0^1 \prod_{j=1}^d \Pr(Y_j = y_j|X_1 = x)\,dx \approx \sum_{q=1}^{n_q} w_q \prod_{j=1}^d \Pr(Y_j = y_j|X_1 = x_q),$$

where $\{x_q : q = 1, \ldots, n_q\}$ are the quadrature points and $\{w_q : q = 1, \ldots, n_q\}$ are the quadrature weights. To compute two-dimensional integrals for the 2-factor model, the approximation uses Gauss-Legendre quadrature points in a double sum:

$$\pi_d(\mathbf{y}) = \int_0^1 \int_0^1 \prod_{j=1}^d \Pr(Y_j = y_j|X_1 = x_1, X_2 = x_2)\,dx_1 dx_2$$

$$\approx \sum_{q_1=1}^{n_q} \sum_{q_2=1}^{n_q} w_{q_1} w_{q_2} \prod_{j=1}^d \Pr(Y_j = y_j|X_1 = x_{q_1}, X_2 = x_{q_2}).$$

TABLE A.2.

Derivatives of the bivariate probability $\pi_{j_1j_2,y_1y_2}$ with respect to the cutpoint a_{jk}, the copula parameter θ_j for the first latent variable and the copula parameter δ_j for the second latent variable, for the 2-factor model for $j, j_1, j_2 = 1, \ldots, d,$ $y, y_1, y_2 = 1, \ldots, K-1$, and $k = 1, \ldots, K-1$. Note that $c_{X_1j}(x,a) = \partial^2 C_{X_1j}(x,a)/\partial x \partial a$, $c_{X_2j}(x,a) = \partial^2 C_{X_2j}(x,a)/\partial x \partial a$, $\dot{f}_{X_2j|X_1}(x_2,y|x_1) = \partial f_{X_2j|X_1}(x_2,y|x_1)/\partial \delta_j = \dot{C}_{j|X_2}\big[C_{j|X_1}(a_{j,y+1}|x_1)|x_2\big] - \dot{C}_{j|X_2}\big[C_{j|X_1}(a_{j,y}|x_1)|x_2\big]$, $\bar{f}_{X_2j|X_1}(x_2,y|x_1) = \partial f_{X_2j|X_1}/\partial \theta_j = c_{X_2j}\big[x_2, C_{j|X_1}(a_{j,y+1}|x_1)\big]\dot{C}_{j|X_1}(a_{j,y+1}|x_1) - c_{jX_2}\big[x_2, C_{j|X_1}(a_{j,y}|x_1)\big]\dot{C}_{j|X_1}(a_{j,y}|x_1)$, $\dot{C}_{j|X_1}(\cdot;\theta_j) = \partial C_{j|X_1}(\cdot;\theta_j)/\partial \theta_j$, $\dot{C}_{j|X_2}(\cdot;\delta_j) = \partial C_{j|X_2}(\cdot;\delta_j)/\partial \delta_j$; the dependence of C_{X_1j} and C_{X_2j} on θ_j and δ_j, respectively, has been suppressed to save space.

$\partial \pi_{j_1j_2,y_1y_2}/\partial a_{jk}$	If
$\int_0^1 \int_0^1 f_{X_2j_2|X_1}(x_2,y_2|x_1) c_{X_2j_1}\big[x_2, C_{j_1|X_1}(a_{j_1,y_1+1}|x_1)\big] c_{X_1j_1}(x_1, a_{j_1,y_1+1})\, dx_1 dx_2$	$a_{jk} = a_{j_1,y_1+1}$
$-\int_0^1 \int_0^1 f_{X_2j_2|X_1}(x_2,y_2|x_1) c_{X_2j_1}\big[x_2, C_{j_1|X_1}(a_{j_1,y_1}|x_1)\big] c_{X_1j_1}(x_1, a_{j_1,y_1})\, dx_1 dx_2$	$a_{jk} = a_{j_1,y_1}$
$\int_0^1 \int_0^1 f_{X_2j_1|X_1}(x_2,y_1|x_1) c_{X_2j_2}\big[x_2, C_{j_2|X_1}(a_{j_2,y_2+1}|x_1)\big] c_{X_1j_2}(x_1, a_{j_2,y_2+1})\, dx_1 dx_2$	$a_{jk} = a_{j_2,y_2+1}$
$-\int_0^1 \int_0^1 f_{X_2j_1|X_1}(x_2,y_1|x_1) c_{X_2j_2}\big[x_2, C_{j_2|X_1}(a_{j_2,y_2}|x_1)\big] c_{X_1j_2}(x_1, a_{j_2,y_2})\, dx_1 dx_2$	$a_{jk} = a_{j_2,y_2}$
0	elsewhere
$\partial \pi_{j_1j_2,y_1y_2}/\partial \theta_j$	If
$\int_0^1 \int_0^1 f_{X_2j_2|X_1}(x_2,y_2|x_1) \bar{f}_{X_2j_2|X_1}(x_2,y_1|x_1)\, dx_1 dx_2$	$j = j_1$
$\int_0^1 \int_0^1 f_{X_2j_1|X_1}(x_2,y_1|x_1) \bar{f}_{X_2j_1|X_1}(x_2,y_2|u_1)\, dx_1 dx_2$	$j = j_2$
0	elsewhere
$\partial \pi_{j_1j_2,y_1y_2}/\partial \delta_j$	If
$\int_0^1 \int_0^1 f_{X_2j_2|X_1}(x_2,y_2|x_1) \dot{f}_{X_2j_2|X_1}(x_2,y_1|x_1)\, dx_1 dx_2$	$j = j_1$
$\int_0^1 \int_0^1 f_{X_2j_1|X_1}(x_2,y_1|x_1) \dot{f}_{X_2j_1|X_1}(x_2,y_2|x_1)\, dx_1 dx_2$	$j = j_2$
0	elsewhere

With Gauss-Legendre quadrature, the same nodes and weights are used for different functions; this helps in yielding smooth numerical derivatives for numerical optimization via quasi-Newton or Newton-Raphson iterations. To decide on n_q, we have calculated several d-dimensional probabilities for different copulas, where the integrals were approximated with Gauss-Legendre quadrature and evaluated numerically with the R function `integrate` for the 1-factor model or the R function `adaptIntegrate` in the R package `cubature` for the 2-factor model. Our comparisons show that $n_q = 15$ is adequate with good precision to at least at three decimal places.

Appendix C. Pseudo-code

Many of the steps in the algorithms can be vectorized if using programming languages such as R or Fortran90. Assuming the number of quadrature points n_q is larger than the number of categories K per item, vectorization would be done over quadrature points.

1-factor log-likelihood given parameter value $\boldsymbol{\theta}$

- Set the cutpoints on the uniform scale: $a_{11}, \ldots, a_{1,K-1}, \ldots, a_{d1}, \ldots, a_{d,K-1}$. Also boundary cutpoints 0 and 1 for each item.
- For each j, compute/store $C_{j|X_1}(a_{jk}|x_q)$ for $k = 1, \ldots, K-1$ and quadrature points x_q for $q = 1, \ldots, n_q$. Store values of $0 = C_{j|X_1}(0|x_q)$ and $1 = C_{j|X_1}(1|x_q)$ by appending, resulting in an $n_q \times (K+1) \times d$ array.
- For each j, compute the probability $f_{j|X_1}(k-1|x_q) = C_{j|X_1}(a_{jk}|x_q) - C_{j|X_1}(a_{j,k-1}|x_q)$ for $k = 1, \ldots, K$ and each quadrature point x_q. Overall after this step, assume that these densities are stored in an $n_q \times K \times d$ array, say `fden`.
- Data loop: For $\mathbf{y}_i$ $(i = 1, \ldots, N$, sample size $N)$,

$$\texttt{fproduct} \leftarrow \texttt{fden}[, y_i[1]+1, 1] \times \cdots \times \texttt{fden}[, y_i[d]+1, d].$$

  Then to update the log-likelihood, use

$$\texttt{loglikcontribution} \leftarrow \log\left\{\sum(\texttt{fproduct} \times \texttt{weightvec})\right\},$$

  where $\texttt{weightvec} = (w_q; q = 1, \ldots, n_q)$ is the vector of quadrature weights.

2-factor log-likelihood given parameter value $\boldsymbol{\theta}$

- Similarly as above to get an $n_q \times (K-1) \times d$ array `fcond1` with $s_{q_1,k,j} = C_{j|X_1}(a_{jk}|x_{q_1}) = F_{j|X_1}(k|x_{q_1})$. Values on the boundary cutpoints are not needed.
- Compute/store $t_{q_1,q_2,k,j} = C_{j|X_2}(C_{j|X_1}(a_{jk}|x_{q_1})|x_{q_2}) = C_{j|X_2}(s_{q_1,k,j}|x_{q_2})$ in an $n_q \times n_q \times (K-1) \times d$ array. Store values of $0 = C_{j|X_2}(0|x_{q_2})$ and $1 = C_{j|X_2}(1|x_{q_2})$ by appending, resulting in an $n_q \times n_q \times (K+1) \times d$ array. Then compute $t_{q_1,q_2,k,j} - t_{q_1,q_2,k-1,j}$ in an appropriate vectorized operation to get an $n_q \times n_q \times K \times d$ array, say `fden2`.
- Data loop: For $\mathbf{y}_i$ $(i = 1, \ldots, N$, sample size $N)$,

$$\texttt{fproduct} \leftarrow \texttt{fden2}[,, y_i[1]+1, 1] \times \cdots \times \texttt{fden2}[,, y_i[d]+1, d]$$

  (as a $n_q \times n_q$ matrix). Then,

$$\texttt{loglikcontribution} \leftarrow \log\left\{\sum(\texttt{fproduct} \times \texttt{weightmatrix})\right\},$$

  where $\texttt{weightmatrix} = (w_{q_1} w_{q_2})_{1 \leq q_1, q_2 \leq n_q}$.

1-factor M_2 statistic with $\sqrt{N}$-consistent $\hat{\boldsymbol{\theta}}$

- With similar calculations as the ones for the likelihood evaluation:

  - Get a 3-dimensional $n_q \times (K+1) \times d$ array `fcond` with $s_{q,k,j} = C_{j|X_1}(a_{jk}|x_q)$.
  - Get a 3-dimensional $n_q \times (K+1) \times d$ array `fdotcond` with $\dot{s}_{q,k,j} = \dot{C}_{j|X_1}(a_{jk}|x_q)$.
  - Get a 3-dimensional $n_q \times (K-1) \times d$ array `cdens` with $c_{q,k,j} = c_{X_1j}(x_q, a_{jk})$. Values on the boundary cutpoints are not needed.

- Compute/store $s_{q,k,j} - s_{q,k-1,j}$ to get an $n_q \times n_q \times K \times d$ array, say `fden`.
- Compute/store $\dot{s}_{q,k,j} - \dot{s}_{q,k-1,j}$ to get an $n_q \times n_q \times K \times d$ array, say `fdotden`.
- The matrices $\boldsymbol{\Delta}_2, \boldsymbol{\Delta}_2^{(c)}, \boldsymbol{\Xi}_2$ can be computed in a straightforward way from the above arrays.

2-factor model M_2 statistic with $\sqrt{N}$-consistent $\hat{\boldsymbol{\theta}}$

- Similarly as above to get:

  - a 3-dimensional $n_q \times (K-1) \times d$ array `fcond1` with $s_{q_1,k,j} = C_{j|X_1}(a_{jk}|x_{q_1})$.
  - a 3-dimensional $n_q \times (K-1) \times d$ array `fdotcond1` with $\dot{s}_{q_1,k,j} = \dot{C}_{j|X_1}(a_{jk}|x_{q_1})$.
  - a 3-dimensional $n_q \times (K-1) \times d$ array `cdens1` with $c_{q_1,k,j} = c_{X_1 j}(x_{q_1}, a_{jk})$.

- For each j, for $k = 1, \ldots, K$ and each quadrature point x_{q_1} compute/store:

  - $t_{q_1,q_2,k,j} = C_{j|X_2}(C_{j|X_1}(a_{jk}|x_{q_1})|x_{q_2}) = C_{j|X_2}(s_{q_1,k,j}|x_{q_2})$ to get an $n_q \times n_q \times (K-1) \times d$ array. Store values of $0 = C_{j|X_2}(0|x_{q_2})$ and $1 = C_{j|X_2}(1|x_{q_2})$ by appending to get an $n_q \times n_q \times (K+1) \times d$ array, say `fcond2`.
  - $\dot{t}_{q_1,q_2,k,j} = \dot{C}_{j|X_2}(C_{j|X_1}(a_{jk}|x_{q_1})|x_{q_2}) = \dot{C}_{j|X_2}(s_{q_1,k,j}|x_{q_2})$ to get an $n_q \times n_q \times (K-1) \times d$ array. Store values of $0 = \dot{C}_{j|X_2}(0|x_{q_2})$ and $0 = \dot{C}_{j|X_2}(1|x_{q_2})$ by appending to get an $n_q \times n_q \times (K+1) \times d$ array, say `fdotcond2`
  - $z_{q_1,q_2,k,j} = c_{X_2 j}(x_{q_2}, C_{j|X_1}(a_{jk}|x_{q_1})) = c_{X_2 j}(x_{q_2}, s_{q_1,k,j})$ to get an $n_q \times n_q \times (K-1) \times d$ array, say `cdens2`.
  - $z_{q_1,q_2,k,j} \times c_{q_1,k,j}$ to get an $n_q \times n_q \times (K-1) \times d$ array, say `cdens12`.
  - $b_{q_1,q_2,k,j} = z_{q_1,q_2,k,j} \times \dot{s}_{q_1,k,j}$ to get an $n_q \times n_q \times (K-1) \times d$ array. Store values of $0 = c_{X_2 j}(x_{q_2}, 0)\dot{C}_{j|X_2}(0|x_{q_2})$ and $0 = c_{X_2 j}(x_{q_2}, 1)\dot{C}_{j|X_2}(1|x_{q_2})$ by appending to get an $n_q \times n_q \times (K+1) \times d$ array, say `cbar`.

- Then compute/store

  - $t_{q_1,q_2,k,j} - t_{q_1,q_2,k-1,j}$ to get an $n_q \times n_q \times K \times d$ array, say `fden2`.
  - $\dot{t}_{q_1,q_2,k,j} - \dot{t}_{q_1,q_2,k-1,j}$ to get an $n_q \times n_q \times K \times d$ array, say `fdotden2`.
  - $b_{q_1,q_2,k,j} - b_{q_1,q_2,k-1,j}$ to get an $n_q \times n_q \times K \times d$ array, say `fbarden`.

- The matrices $\boldsymbol{\Delta}_2, \boldsymbol{\Delta}_2^{(c)}, \boldsymbol{\Xi}_2$ can be computed from the above arrays.

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics, 44*, 182–198.

Bartholomew, D., Knott, M., & Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach, 3rd Edition.* New York: Wiley.

Braeken, J., Tuerlinckx, F., & De Boeck, P. (2007). Copula functions for residual dependency. *Psychometrika, 72*, 393–411.

Braeken, J. (2011). A boundary mixture approach to violations of conditional independence. *Psychometrika, 76*, 57–76.

Brechmann, E. C., Czado, C., & Aas, K. (2012). Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics, 40*, 68–85.

Chang, E., D'Zurilla, T., & Maydeu-Olivares, A. (1994). Assessing the dimensionality of optimism and pessimism using a multimeasure approach. *Cognitive Therapy and Research, 18*, 143–160.

de Menezes, L. M. (1999). On fitting latent class models for binary data: the estimation of standard errors. *British Journal of Mathematical and Statistical Psychology, 52*, 149–168.

Genest, C. & MacKay, J. (1986). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician, 40*, 280–283.

Gibbons, R. & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423–436.

Hull, J. & White, A. (2004). Valuation of a CDO and an nth to default CDS without Monte Carlo simulation. *Journal of Derivatives, 12*, 8–23.

Hult, H. & Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability, 34*, 587–608.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis, 94*, 401–419.

Jöreskog, K. G. & Moustaki, I. (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research, 36*, 347–387.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.

Krupskii, P. & Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis, 120*, 85–101.

Kurowicka, D. & Joe, H. (2011). *Dependence Modeling: Vine Copula Handbook*. Singapore: World Scientific.

Lehmann, E. L. (1998). *Elements of Large Sample Theory*. New York: Springer.

Maydeu-Olivares, A. (2006). Limited information estimation and testing of discretised multivariate normal structural models. *Psychometrika, 71*, 57–77.

Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*, 713–732.

McDonald, R. P. (1997). Normal ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257–269). New York: Springer.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: University Press.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

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

Panagiotelis, A., Czado, C., & Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association, 107*, 1063–1072.

Rizopoulos, D. (2011). *ltm: Latent Trait Models under IRT*. R package version 0.9-7.

Samejima, F. (1969). Calibration of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 17*.

Scheier, M. & Carver, C. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Cognitive Therapy and Research, 4*, 219–247.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris, 8*, 229–231.

Stroud, A. H. & Secrest, D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall.