Journal of Multivariate Analysis 138 (2015) 53–73

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Structured factor copula models: Theory, inference and computation

Pavel Krupskii *, Harry Joe

Department of Statistics, University of British Columbia, Vancouver BC, Canada V6T 1Z4

ARTICLE INFO

Article history:
Received 15 March 2014
Available online 13 November 2014

AMS 2000 subject classifications:
62H25
60H99

Keywords:
Bi-factor model
Conditional independence
Dependence measure
Factor analysis
Tail asymmetry
Tail dependence
Truncated vine

ABSTRACT

In factor copula models for multivariate data, dependence is explained via one or several common factors. These models are flexible in handling tail dependence and asymmetry with parsimonious dependence structures. We propose two structured factor copula models for the case where variables can be split into non-overlapping groups such that there is homogeneous dependence within each group. A typical example of such variables occurs for stock returns from different sectors. The structured models inherit most of dependence properties derived for common factor copula models. With appropriate numerical methods, efficient estimation of dependence parameters is possible for data sets with over 100 variables. We apply the structured factor copula models to analyze a financial data set, and compare with other copula models for tail inference. Using model-based interval estimates, we find that some commonly used risk measures may not be well discriminated by copula models, but tail-weighted dependence measures can discriminate copula models with different dependence and tail properties.

1. Introduction

Modeling high-dimensional data is a challenging task requiring flexible and tractable models. Models based on multivariate normality or Gaussianity are widely used in different applications due to their simplicity and tractability. In these models, special correlation structures are used to reduce the number of dependence parameters to a linear function of the dimension. A typical example is a Gaussian factor model where one or several common factors define the dependence structure for all of the variables. Factor copula models proposed in Krupskii and Joe [19] are extensions of the Gaussian factor model allowing greater flexibility when modeling non-Gaussian dependence. In particular, strong tail dependence and tail asymmetry can be accommodated. In data sets with a large number of variables, data can come from different sources or be clustered in different groups, for example, stock returns from different sectors or grouped item response data in psychometrics; thus dependence within each group and among different groups can be qualitatively different, and structured factor models can make use of the group information.

In psychometrics, sometimes a bi-factor correlation structure is used when variables or items can be split into non-overlapping groups; see for example Gibbons and Hedeker [10] and Holzinger and Swineford [12]. In a Gaussian bi-factor model, there is one common Gaussian factor which defines dependence between different groups, and one or several independent group-specific Gaussian factors which define dependence within each group. An alternative way to model dependence for grouped data is a nested model where the dependence in groups is modeled via dependent group-specific factors and the observed variables are assumed to be conditionally independent given these group-specific factors. The

* Corresponding author.
E-mail address: KrupskiiPV@yandex.ru (P. Krupskii).

http://dx.doi.org/10.1016/j.jmva.2014.11.002
0047-259X/

nested model is similar to Gaussian models with multilevel covariance structure; see Muthen [25]. Despite the simplicity, these two models have the same drawbacks as a common Gaussian factor model—they do not account for tail asymmetry and tail dependence.

In this paper, we propose copula extensions for bi-factor and nested Gaussian models. The extensions are called *structured factor copula models*. The proposed models contain 1- and 2-factor copula models introduced in Krupskii and Joe [19] as special cases, while allowing flexible dependence structure both for within group and between group dependence. As a result, the models can be suitable for modeling high-dimensional data sets consisting of several groups of variables with homogeneous dependence in each group.

The proposed multivariate copula models are built from a sequence of bivariate copulas in a similar way to vine copulas. Let $F_{\mathbf{X}}$ be the multivariate cumulative distribution function (cdf) of a random d-dimensional vector $\mathbf{X} = (X_1, \ldots, X_d)$, and let F_{X_j} be the cdf of X_j for $j = 1, \ldots, d$. The copula $C_{\mathbf{X}}$, corresponding to $F_{\mathbf{X}}$, is a multivariate uniform cdf such that $F_{\mathbf{X}}(x_1, \ldots, x_d) = C_{\mathbf{X}}(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))$. By Sklar [30], $C_{\mathbf{X}}$ is unique if $F_{\mathbf{X}}$ is continuous. Copula functions allow for different types of dependence structure and are popular for modeling non-Gaussian dependence, including stock returns, insurance and hydrology data; see for example see Patton [29], McNeil et al. [23], Salvadori et al. [24] and others.

The proposed structured copula models are special cases of truncated-vine copula models with latent variables. In a vine model, bivariate linking copulas are applied to conditional cdfs to sequentially construct a multivariate distribution. The resulting vine model or pair-copula construction allows great flexibility in modeling different types of dependence structure by choosing appropriate linking copulas; see Kurowicka and Joe [20] and Brechmann and Czado [5] for more details. We show that depending on the choice of bivariate copulas in the structured copula models, different types of strength of dependence in the tails can be accommodated, similar to common factor copula models.

For a large number of variables which divide naturally into non-overlapping groups, it could be convenient to first separately model each group of variables followed by a method to combine the smaller models into a bigger model. Our structured copula models are one way to do this. The grouped t-copula of Demarta and McNeil [7] can handle groups but can only accommodate reflection symmetry. Another approach is hierarchical Kendall copulas in Brechmann [4]; it makes use of conditional independence given some group aggregation variables. Kendall functions only have simple form for exchangeable Archimedean copulas, so that hierarchical Kendall copulas are only convenient for exchangeable dependence within groups. Also nested Archimedean copulas (Section 4.1 of Joe [15]) are too parsimonious and have the property of exchangeable dependence within groups.

The details in this paper are given for continuous response variables, but the structured copula models can also be developed for discrete ordinal variables or mixed discrete/continuous variables. Factor copula models for item response are studied in Nikoloulopoulos and Joe [27], and if the items can be classified into non-overlapping groups, then the bi-factor or nested factor copula models are candidates when there is tail asymmetry or tail dependence.

The rest of the paper is organized as follows. In Section 2 we define bi-factor and nested copula models including a special case of Gaussian copulas, and compare the properties of these models with those of 1- and 2-factor copula models in Section 3. Section 4 has details on numerical maximum likelihood with a modified Newton–Raphson algorithm. Section 5 has a resampling method to obtain model-based interval estimates of the portfolio risk measures of Value-at-Risk and conditional tail expectation. In Section 6, we apply different copula–GARCH models to a financial data set and compare estimates of the Value-at-Risk, conditional tail expectations as well as some other tail-based quantities. The results show that structured factor copula models can parsimoniously estimate the dependence structure of the data. Value-at-Risk and other risk measures, which are widely used in financial applications, cannot efficiently differentiate models with different tail properties, and tail-weighted dependence measures are a better match to the fit of copula models based on the Akaike information criterion. Section 7 concludes with a discussion of future research.

2. Structured factor copula models

Common factor models assume that d observed variables are conditionally independent given $1 \le p \ll d$ latent variables that affect each observed variables; for identifiability, the latent variables are assumed to be independent. Structured factor models assume that there is structure to the observed variables and each latent variable is linked to a subset of the observed variables. For Gaussian structured factor models, this corresponds to many structured zeros in the matrix of loadings; in this case, with fewer parameters in the loading matrix compared with the common factor model, and the p latent variables could be dependent, as in the oblique factor model of Harris and Kaiser [11] and McDonald [22]. With a large d, structured Gaussian factor models are also parsimonious models to parameterize the correlation matrix in $O(d)$ parameters (instead of $d(d-1)/2$ parameters). The main goal of this section is to present the copula version of two Gaussian structured factor models; for the extension, the parameters of the Gaussian structured factor models are converted to a set of correlations and partial correlations that are algebraically independent and that have a truncated vine structure, and then the correlations and partial correlations are replaced by bivariate copulas. Similar copula extensions exist for other structured factor models.

A specific case of structured factor models occurs when variables can be divided into non-overlapping groups. Assume that we have G groups of variables and there are d_g variables in the gth group, $g = 1, \ldots, G$. Let $U_{ij} \sim U(0, 1), i = 1, \ldots, d_g$, and suppose variables $U_{1g}, \ldots, U_{d_g g}$ belong to the gth group. Denote the joint cdf of $\mathbf{U} = (U_{11}, \ldots, U_{d_1 1}, \ldots, U_{1G}, \ldots, U_{d_G G})$ by $C_{\boldsymbol{U}}$. Let $d = \sum_{g=1}^{G} d_g$ be the total number of variables.