Jinguo Gong, Yadong Li and Qiwei Yao

Estimation of extreme quantiles for functions of dependent random variables

Article (Accepted version) (Refereed)

Estimation of Extreme Quantiles for Functions of Dependent Random Variables

Jinguo Gong

Southwestern University of Finance and Economics, Chengdu, China

Yadong Li

Quantitative Analytics, Barclays, New York, USA

Liang Peng

Georgia State University, Atlanta, USA

and Qiwei Yao

London School of Economics and Political Science, UK, and Guanghua School of Management, Peking University, China

Summary. We propose a new method for estimating the extreme quantiles for a function of several dependent random variables. In contrast to the conventional approach based on extreme value theory, we do not impose the condition that the tail of the underlying distribution admits an approximate parametric form, and, furthermore, our estimation makes use of the full observed data. The proposed method is semiparametric as no parametric forms are assumed on the marginal distributions. But we select appropriate bivariate copulas to model the joint dependence structure by taking the advantage of the recent development in constructing large dimensional vine copulas. Consequently a sample quantile resulted from a large bootstrap sample drawn from the fitted joint distribution is taken as the estimator for the extreme quantile. This estimator is proved to be consistent under the regularity conditions on the closeness between a quantile set and its truncated set, and the empirical approximation for the truncated set. The simulation results lend further support to the reliable and robust performance of the proposed method. The method is further illustrated by an real world example in backtesting financial risk models.

Keywords: Bootstrap, D-vine copula, empirical distribution function, extreme quantile, sample quantiles, time series.

Address for correspondence: Qiwei Yao, Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK.

E-mail: q.yao@lse.ac.uk

1 Introduction

Let $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ be a sample from the population of a p-variate random vector $\mathbf{X} = (X_1, \cdots, X_p)$. Let $\xi = h(\mathbf{X})$ be a random variable defined as a function of $\mathbf{X}$, where the function $h(\cdot)$ is known. The goal of this paper is to estimate the $(1-\alpha)$-th quantile of ξ, i.e.

$$Q_\xi(\alpha) = \min\{\, x : P(\xi \le x) \ge 1 - \alpha \,\}, \tag{1.1}$$

where $\alpha > 0$ is a very small constant such that $n\alpha$ is small. When $\alpha < 1/n$, $Q_\xi(\alpha)$ is outside the range of observed data. This rules out the possibility to estimate $Q_\xi(\alpha)$ by the sample quantile of $\{\xi_1, \cdots, \xi_n\}$, where $\xi_i = h(\mathbf{X}_i)$. This study was motivated by a backtesting problem in financial risk management, for which we need to estimate the $(1-\alpha)$-th quantile of $\xi = h(X_1, \cdots, X_p)$ with $\alpha = 0.0005$ or 0.0001, p in the range from 10 to 200, and sample size n in the order of a few hundreds to thousands. See section 5 below.

The standard approach to estimate quantiles outside the range of the data is to assume that the distribution of ξ is in the domain of attraction of an extreme value distribution. Based on the characterization of this assumption (Proposition 3.3.2 of Embrechts, Klüppelberg and Mikosch, 1997), extreme quantiles can be estimated via the estimation for the parameters in the extreme value distribution and the normalized constants. However the estimation is inefficient as only a small proportion of the observations at a tail can be used. This causes further difficulties in practice as the estimation is often sensitive to the proportion of the data used, although there exist in the literature the data driven methods for choosing the sample fraction (Ferreira, de Haan and Peng, 2003) and the bias-reduced estimators (Gomes and Pestana, 2007, and Beirlant et al., 2004). See, e.g., Embrechts, Klüppelberg and Mikosch (1997), Coles (2001) and de Haan and Ferreira (2006) for a detailed account of this approach.

In addition to the methods based on univariate extreme value theory, one can also assume that $\mathbf{X}$ lies in the domain of attraction of a multivariate extreme value distribution; see de Haan and Ferreira (2006). This implies that the tail distribution of each component of $\mathbf{X}$ can be approximated by a parametric form determined by an extreme value distribution while the joint tail dependence has a nice homogeneous property. For estimating extreme quantiles for the functions of $\mathbf{X}$, one can model the joint tail dependence either parametrically (Coles and Tawn, 1994) or nonparametrically, and then extrapolate data based on the homogeneous property (de Haan and Sinha, 1999, and Drees and de Haan, 2013). Although using multivariate extreme value theory may be more efficient than using univariate extreme value theory (Bruun and Tawn, 1988), the sensitivity on the amount of data used in estimation remains as a serious drawback. Furthermore, when the dimension of $\mathbf{X}$ is not small, finding a parametric family for the joint tail dependence is extremely difficult and the nonparametric estimation for the joint tail dependence

is too poor to be practically usable.

In this paper, we propose a new semiparametric method for estimating $Q_\xi(\alpha)$. It consists of three steps: (i) we apply the empirical distribution transformation to each components of $\mathbf{X}$ to make all the marginal distributions approximate $U[0,1]$, (ii) we then select an appropriate copula to model the joint dependence structure, (iii) finally we draw a large bootstrap sample $\{\mathbf{X}_1^\star, \cdots, \mathbf{X}_m^\star\}$ from the fitted joint distribution derived from (i) and (ii), and estimate $Q_\xi(\alpha)$ by the $(1-\alpha)$-th sample quantile of $\{\xi_1^\star, \cdots, \xi_m^\star\}$, where $\xi_i^\star = h(\mathbf{X}_i^\star)$. Fitting a p-dimensional copula in (ii) is feasible due to the recent development of vine copula construction; see section 2 below. The bootstrap sample size m can be arbitrarily large. In practice we typically require, e.g. $m\alpha \geq 20$. This method does not impose a parametric form directly on the tail of the distribution of ξ or the marginal distributions of $\mathbf{X}$. It is free from choosing the fraction of the whole sample to be used in estimation, which is a notorious tuning parameter often causing difficulties in practice.

Our new proposal can only work when $p > 1$. It is based on an important observation that it is not necessary to go to extremes along any component of $\mathbf{X} = (X_1, \cdots, X_p)$ in order to observe the joint extreme event $\{h(X_1, \cdots, X_p) > Q_\xi(\alpha)\}$. Therefore we only need to capture the dependence among $X_1, \cdots, X_p$ within the observed range, which is practically feasible. The fact that $p > 1$ also makes it possible to generate a bootstrap sample of size m greater, or much greater, than n. Although this method can handle the cases when the components of $\mathbf{X}$ are dependent with each other, its intuition is at its clearest when all $X_1, \cdots, X_p$ are independent, as then a bootstrap sample for $\mathbf{X}$ can be easily obtained by sampling each component separately from its n observations. Note that the corresponding bootstrap sample space consists of n^p elements. It ensures sufficient diversity in the bootstrap sample even for m much greater than n. Hence $Q_\xi(\alpha)$ can be well estimated by the $(1-\alpha)$-th sample quantile from a bootstrap sample with m sufficiently large (Theorem 3.1 of Dekkers and de Haan, 1989).

However, as stated above, the fundamental reason for our approach to be a creditable one is that it is not necessary to go to extremes along any component of $\mathbf{X}$ in order to observe a joint extreme event. We report a simple simulation result below to illustrate this key point. Let all components X_j be i.i.d., and $\xi = \frac{1}{p}\sum_{1\leq j\leq p} X_j$. We approximate the probability $\alpha = P\{\xi > Q_\xi(\alpha)\}$ by

$$\widehat{\alpha}_n = P\{\ \xi > Q_\xi(\alpha),\ F_j^{-1}(1/n) \leq X_j \leq F_j^{-1}(1-1/n)\ \text{ for } 1 \leq j \leq p\ \},$$

where $F_j(\cdot)$ denotes the marginal distribution function of X_j. With available n observations, the distribution range for X_j covered by the data can be regarded as from $F_j^{-1}(1/n)$ to $F_j^{-1}(1-1/n)$. This range cannot be enlarged by resampling from the observed data. Thus $\widehat{\alpha}_n$ can be regarded as the probability of the event $\{\xi > Q_\xi(\alpha)\}$ truncated within the range covered by a sample of

size n. Our method will work when $\widehat{\alpha}_n$ is close to α, as we can only model the joint distribution well within the observed range.

The table below lists the values of $\widehat{\alpha}_n$ calculated by a simulation with 1,000,000 replications for $p = 20$, $n = 500$ or $1,000$ and the distribution of X_j being uniform on the unit interval, standard normal or Student's t with 4 degrees freedom. Note that t_4 is a very heavy-tailed distribution, as $E(X_j^4) = \infty$ if $X_j \sim t_4$.

Distribution of X_j	n	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$U(0,1)$	500	.04741	.00942	.00436	.00078	.00045
	1000	.04809	.00949	.00438	.00084	.00046
$N(0,1)$	500	.04360	.00829	.00401	.00075	.00038
	1000	.04645	.00896	.00439	.00083	.00043
t_4	500	.03629	.00540	.00204	.00013	.00004
	1000	.04183	.00609	.00251	.00020	.00005

This simulation indicates that it is possible to estimate $Q_\xi(\alpha)$ accurately for α as small as 0.0005 even with sample size $n = 500$ when $\mathbf{X}$ is uniformly distributed or normal. However for the heavy-tailed distributions such as t_4, the proposed method may incur large estimation errors, and therefore is not adequate. In fact our approach does not involve any direct extrapolations, it can estimate extreme but *not too extreme* quantiles. How extreme it can go depends on the underlying distribution, the sample size n, and the form of function $h(\cdot)$ which defines ξ. However when ξ is defined in terms of empirical marginal distribution functions, all marginal distributions are effectively $U(0,1)$. Then our method will provide accurate estimation even for very small α (see also sections 4 & 5 below). In fact many risk metrics used in backtesting fall into this category.

Wang, Li and He (2012) proposed a method for estimating high conditional quantiles by combining quantile regression with extreme value theory. It remains as an open question if the method proposed in this paper can be further developed for estimating conditional quantiles. Investigation in this direction is beyond the scope of this paper.

The rest of the paper is organized as follows. The methodology is presented in section 2. It also contains a brief introduction of D-vine copulas. The asymptotic properties are developed in section 3. We have shown that the proposed estimator is consistent under the regularity conditions on the closeness between a quantile set and its truncated set, and the empirical approximation for the truncated set. Simulation illustration is reported in section 4. Section 5 contains a brief introduction on a backtesting problem in financial risk management, which actually motivated this study. Using the example with a calibrated GJR-GARCH model for daily S&P 500 indices in 2005

– 2014, we illustrate how the proposed method provides an adequate solution for a challenging backtesting problem. Comparison with the conventional methods based on the extreme value theory is reported in both sections 4 & 5. A generic R-code for implementing the proposed method is available online at `stats.lse.ac.uk/q.yao/qyao.links/paper/EEQ-Copula.R`.

2 Methodology

2.1 Notation

Let $\mathbf{X} = (X_1, \cdots, X_p)$, $F(\cdot)$ be the cumulative distribution function (CDF) of $\mathbf{X}$, $F_j(\cdot)$ be the CDF of X_j, and $U_j = F_j(X_j)$. Then $U_j \sim U[0,1]$ for $1 \le j \le p$. Let $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})$, $i = 1, \cdots, n$, be a random sample from $\mathbf{X}$. Put

$$\widehat{F}_j(x) = \frac{1}{n+1} \sum_{i=1}^{n} I(X_{ij} \le x), \qquad U_{ij} = \widehat{F}_j(X_{ij}). \tag{2.1}$$

Then $\sup_x |\widehat{F}_j(x) - F_j(x)| \xrightarrow{p} 0$, and $\{U_{1j}, \cdots, U_{nj}\}$ may be *approximately* regarded as a sample from $U[0,1]$ when n is large.

It follows from Sklar's theorem that for $\mathbf{x} = (x_1, \cdots, x_p) \in R^p$,

$$\begin{aligned} F(\mathbf{x}) &= P(X_1 \le x_1, \cdots, X_p \le x_p) \\ &= P\{U_1 \le F_1(x_1), \cdots, U_p \le F_p(x_p)\} = C\{F_1(x_1), \cdots, F_p(x_p)\}, \end{aligned} \tag{2.2}$$

where $C(\cdot)$ is the CDF of $\mathbf{U} \equiv (U_1, \cdots, U_p)$, and is called a p-variate copula. In fact $C(\cdot)$ is a distribution function on $[0,1]^p$ with all one-dimensional uniform marginal distributions. We always assume that $C(\cdot)$ admits a probability density function (PDF), denoted by $c(\cdot)$, which is called a copula density function. Then the joint PDF of $\mathbf{X}$ can be written as

$$f(\mathbf{x}) = c\{F_1(x_1), \cdots, F_p(x_p)\} \prod_{j=1}^{p} f_j(x_j), \tag{2.3}$$

where $f_j(\cdot)$ is the PDF of X_j. Hence $c(\cdot) \equiv 1$ if and only if $X_1, \cdots, X_p$ are independent. For more properties on copulas we refer to Nelson (2006). Due to the invariant property with respect to marginals, copula models have become one of the most frequently used tool in risk management; see McNeil, Frey and Embrechts (2005).

2.2 Estimation for $F(\cdot)$

Representations (2.2) and (2.3) separate the dependence among the components of $\mathbf{X}$ from the marginal distributions. They indicate clearly that the dependence is depicted by a copula. A

natural and completely nonparametric estimator for the copula function $C(\cdot)$ is the empirical copula function

$$\widehat{C}(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} I(U_{i1} \le u_1, \cdots, U_{ip} \le u_p), \qquad \mathbf{u} = (u_1, \cdots, u_p) \in [0,1]^p. \tag{2.4}$$

Obviously such a nonparametric estimator $\widehat{C}(\cdot)$ suffers from the so-called 'curse-of-dimensionality' even for moderately large p, though it is still root-n consistent; see, e.g. Fermanian *et al.* (2004). One alternative is to impose the assumption that the unknown copula belongs to a parametric family $\{c(\cdot;\, \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where copula density function $c(\cdot;\, \boldsymbol{\theta})$ is known up to the d unknown parameters $\boldsymbol{\theta}$, the parameter space $\boldsymbol{\Theta}$ is a subset of R^d and $d \ge 1$ is an integer. Then $\boldsymbol{\theta}$ can be estimated by, for example, the pseudo maximum likelihood estimator defined as

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} \log c(U_{i1}, \cdots, U_{ip}; \boldsymbol{\theta}).$$

See also section 2.3 below for further discussion on the specification of $c(\cdot; \boldsymbol{\theta})$. Now by (2.2), an estimator for the CDF of $\mathbf{X}$ is defined as

$$\widehat{F}(\mathbf{x}) = C\{\widehat{F}_1(x_1), \cdots, \widehat{F}_p(x_p); \widehat{\boldsymbol{\theta}}\}, \qquad \mathbf{x} \in R^p, \tag{2.5}$$

where $C(\cdot;\, \boldsymbol{\theta})$ is the CDF corresponding to the PDF $c(\cdot;\, \boldsymbol{\theta})$.

2.3 Copula specification: D-vines

For any integer $p \ge 3$, a p-variate copula function can be effectively specified via pairwise decomposition, leading to various forms of vine copulas (Bedford and Cooke, 2001 and 2002). Different orders of the pairings in the decomposition yield different vines. Nevertheless, only bivariate copula functions are to be specified. When the components of random vector $\mathbf{X}$ (therefore also $\mathbf{U}$) are naturally ordered, such as in the backtesting problems described in section 5 below, the D-vine copulas are particularly easy to use. A copula density function, i.e. a PDF of $\mathbf{U}$, specified by a D-vine admits the form

$$c(\mathbf{u}) = \prod_{j=1}^{p-1}\prod_{i=1}^{p-j} c_{i,\, i+j|i+1,\cdots,i+j-1}\{F(u_i|u_{i+1}, \cdots, u_{i+j-1}),\ F(u_{i+j}|u_{i+1}, \cdots, u_{i+j-1})\}, \tag{2.6}$$

see, for example, (8) of Aas *et al.* (2009), where $F(u_k|u_{i+1}, \cdots, u_{i+j-1})$ denotes the conditional CDF of U_k given $(U_{i+1} = u_{i+1}, \cdots, U_{i+j-1} = u_{i+j-1})$, and $c_{i,\, i+j|i+1,\cdots,i+j-1}(\cdot)$ denotes the copula density for the conditional distribution of (U_i, U_{i+j}) given $U_{i+1}, \cdots, U_{i+j-1}$. Now some remarks are in order.

Remark 1. (i) Only bivariate copula density functions are used in (2.6). See Joe (1997) for various parametric copula families which can be used to specify those copula functions.

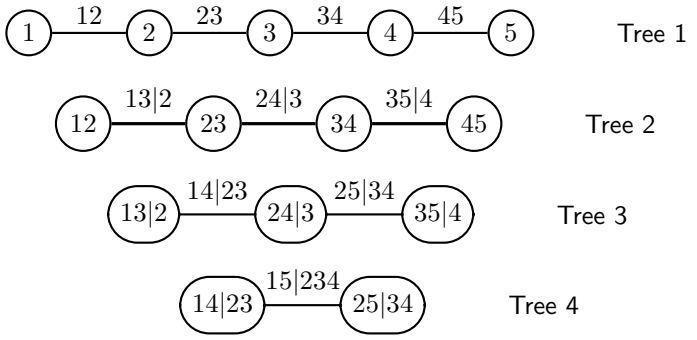

Figure 1: Tree illustration of a D-Vine with 5 variables.

(ii) A p-variate D-vine can be represented as a graph with the maximum p-1 trees, corresponding to $j = 1, \cdots, p-1$ on the RHS of (2.6); see, for example, Aas *et al.* (2009). However the construction of those trees must be done in the order of $j = 1, 2, \cdots, p-1$. For example, the conditional CDF $F(u_i|u_{i+1}, \cdots, u_{i+j-1})$ is required in the j-th tree. By Lemma 1 below, it can be calculated based on a copula constructed in the (j-1)th tree:

$$F(u_i|u_{i+1}, \cdots, u_{i+j-1}) = \frac{\partial C_{i,i+j-1|i+1,\cdots,i+j-2}\{F(u_i|u_{i+1}, \cdots, u_{i+j-2}),\ F(u_{i+j-1}|u_{i+1}, \cdots, u_{i+j-2})\}}{\partial F(u_{i+j-1}|u_{i+1}, \cdots, u_{i+j-2})},$$

where $C_{i,i+j-1|i+1,\cdots,i+j-2}(\cdot)$ is the copula corresponding to the copula density $c_{i,i+j-1|i+1,\cdots,i+j-2}(\cdot)$ specified in the (j-1)th tree. For $j = 1$, $F(u_i) = u_i$. For $j = 2$,

$$F(u_i|u_{i+1}) = \frac{\partial C_{i,i+1}\{F(u_i),\ F(u_{i+1})\}}{\partial F(u_{i+1})} = \frac{\partial C_{i,i+1}(u_i, u_{i+1})}{\partial u_{i+1}}.$$

Figure 1 illustrates the tree structure of a D-vine with $p = 5$ variables.

(iii) U_i and U_{i+j} are conditionally independent given $U_{i+1}, \cdots, U_{i+j-1}$ if and only if

$$c_{i,\, i+j|i+1,\cdots,i+j-1}(\cdot) \equiv 1. \tag{2.7}$$

This follows from (2.3) by letting $f(\mathbf{x})$ be the conditional PDF of (U_i, U_{i+j}) given $U_{i+1}, \cdots, U_{i+j-1}$.

(iv) In applications we often assume that the dependence is of the order $m(< p)$ in the sense that (2.7) holds for all $j > m$. Then (2.6) reduces to

$$c(\mathbf{u}) = \prod_{j=1}^{m} \prod_{i=1}^{p-j} c_{i,\, i+j|i+1,\cdots,i+j-1}\{F(u_i|u_{i+1}, \cdots, u_{i+j-1}),\ F(u_{i+j}|u_{i+1}, \cdots, u_{i+j-1})\}. \tag{2.8}$$

A particular simple case is a Markov D-vine copula which admits the dependence at order $m = 1$ with the copula density function of the form

$$c(\mathbf{u}) = \prod_{i=1}^{p-1} c_{i,i+1}\{F(u_i), F(u_{i+1})\} = \prod_{i=1}^{p-1} c_{i,i+1}(u_i, u_{i+1}),$$

where $c_{i,j}(\cdot)$ are bivariate copulas. For example, when the components of $\mathbf{X}$ are p successive values of a Markov process, $\mathbf{X}$ admits a Markov D-vine copula.

(v) We may apply some goodness-of-fit statistics to choose among different specifications or to test a particular model. The goodness-of-fit can be measured in terms of the difference between the empirical copula $\widehat{C}(\cdot)$ defined in (2.4) and the fitted parametric copula $C(\cdot;\boldsymbol{\theta})$ in (2.5). This leads to the Kolmogorov-Smirnov and Cramér-von Mises statistics

$$T_n = n\int_{[0,1]^p} \big\{C(\mathbf{u};\widehat{\boldsymbol{\theta}}) - \widehat{C}(\mathbf{u})\big\}^2 d\mathbf{u}, \qquad S_n = \sup_{\mathbf{u}\in[0,1]^p} \sqrt{n}\big|C(\mathbf{u};\widehat{\boldsymbol{\theta}}) - \widehat{C}(\mathbf{u})\big|.$$

Genest and Rémillard (2008) showed that both the above statistics lead to a consistent test in the sense that if the true copula is not within the specified parametric family, the model will be rejected with probability converging to 1. Unfortunately their asymptotic null distributions depend on the underlying distribution. In practice the parametric bootstrap method described in Appendix A of Genest *et al.* (2009) can be used to evaluate the P-values. The validity of the bootstrap method is established by Genest and Rémillard (2008).

(vi) The D-vine decomposition (2.6) is valid for any continuous distribution on $[0,1]^p$ with uniform marginal distributions. On the other hand, with any bivariate copula density functions used on the RHS of (2.6), the D-vine constructed in the manner described in (ii) above is a valid p-variate copula, i.e. (2.6) is a proper PDF on $U[0,1]^p$ with uniform marginals. Both these assertions can be established by mathematical induction.

(vii) When the components of $\mathbf{X}$ are not naturally ordered as a time series, other vine copula families such as C-vine could be used. We refer to Czado, Brechmann and Gruber (2013) for a survey on the selection of vine copulas.

Lemma 1. Let Y and Z be two random variables, $\mathbf{W}$ be a random vector, and $\mathbf{Z} = (Z, \mathbf{W})$. Denoted by, respectively, $F_{\mathbf{W}}$ and $C_{\mathbf{W}}$ the CDF and the copula of $\mathbf{W}$. Then it holds that

$$F_{Y|Z}(y|z) = \frac{\partial C_{Y,Z}\{F_Y(y), F_Z(z)\}}{\partial F_Z(z)}, \quad F_{Y|\mathbf{Z}}(y|\mathbf{z}) = \frac{\partial C_{Y,Z|\mathbf{W}}\{F_{Y|\mathbf{W}}(y|\mathbf{w}), F_{Z|\mathbf{W}}(z|\mathbf{w})\}}{\partial F_{Z|\mathbf{W}}(z|\mathbf{w})}. \tag{2.9}$$

First equality in (2.9) follows from calculus. The second equality follows from the first by applying it to the conditional distribution of (Y, Z) given $\mathbf{W}$. Those relationships were first established by Joe (1996).

2.4 Estimation for extreme quantiles

With the estimated distribution (2.5) for $\mathbf{X}$, in principle we can deduce an estimator for the distribution of $\xi = h(\mathbf{X})$. Unfortunately in most applications such an estimator cannot be evaluated explicitly. We propose to draw a bootstrap sample $\mathbf{X}_1^\star, \cdots, \mathbf{X}_m^\star$ from (2.5), and to estimate the

extreme quantile $Q_\xi(\alpha)$ of ξ (see (1.1)) by the corresponding sample quantile of $\{\xi_i^\star = h(\mathbf{X}_i^\star)\}$, i.e.

$$\widehat{Q}_\xi(\alpha) = \xi^\star_{[m\alpha]}, \tag{2.10}$$

where $\xi^\star_{[j]}$ denotes the j-th largest value among $\xi_1^\star, \cdots, \xi_m^\star$. We require m sufficiently large such that, for example, $m\alpha \geq 20$.

We apply the inverse of the Rosenblatt transformation to draw $u_1, \cdots, u_p$ from D-vine copula density (2.6). Then we let

$$x_j = \widehat{F}_j^{-1}(u_j), \quad j = 1, \cdots, p, \tag{2.11}$$

where $\widehat{F}_j$ defined in (2.1). To this end, draw $v_1, \cdots, v_p$ independently from $U[0,1]$. Let $u_1 = v_1$, and

$$u_i = F^{-1}(v_i | u_1, \cdots, u_{i-1}) \qquad \text{for } i = 2, \cdots, p,$$

where $F^{-1}(\cdot\,|u_1, \cdots, u_{i-1})$ denotes the inverse function of the conditional CDF of U_i given $(U_1 = u_1, \cdots, U_{i-1} = u_{i-1})$ which is determined by the D-vine copula density (2.6). It follows from Lemma 1 that

$$F(u_i | u_1, \cdots, u_{i-1}) = \frac{\partial C_{1,\, i|2,\cdots,i-1}\{F(u_1|u_2, \cdots, u_{i-1}),\ F(u_i|u_2, \cdots, u_{i-1})\}}{\partial F(u_1|u_2, \cdots, u_{i-1})},$$

where $C_{1,\, i|2,\cdots,i-1}(\cdot)$ is the copula function corresponding to the copula density $c_{1,\, i|2,\cdots,i-1}$ contained on the RHS of (2.6). Aas *et al.* (2009) outlined an algorithm to implement the above scheme.

Remark 2. When all the components of $\mathbf{X}$ are known to be independent with each other, our approach still applies. In this case, $\mathbf{X}_i^\star = (X_{i1}^\star, \cdots, X_{ip}^\star)$ can be obtained with $X_{ij}^\star$ resampled independently from $\{X_{1j}, \cdots, X_{nj}\}$.

3 Asymptotic properties

In this section we present the consistency for our extreme quantile estimation. Recall $C(\cdot) = C(\cdot; \boldsymbol{\theta})$ is the CDF of $\mathbf{U} = (U_1, \cdots, U_p)$. The target quantile, as a function of $\boldsymbol{\theta}$, can be expressed as

$$Q_\xi(\alpha; \boldsymbol{\theta}) = \min\big\{x : P_{\boldsymbol{\theta}}(\xi > x) \leq \alpha\big\},$$

where $\xi = h(\mathbf{X}) = h\{F_1^{-1}(U_1), \cdots, F_p^{-1}(U_p)\}$; see (1.1). Put

$$
\begin{aligned}
A(x) &= \big\{(u_1, \cdots, u_p) : h\{F_1^{-1}(u_1), \cdots, F_p^{-1}(u_p)\} > x\big\}, \\
A_n(x) &= \big\{(u_1, \cdots, u_p) : (u_1, \cdots, u_p) \in A(x),\ \frac{1}{n+1} \le u_1, \cdots, u_p \le \frac{n}{n+1}\big\}, \\
B_n(x) &= \Big\{(u_1, \cdots, u_p) : h\{F_1^{-1}(\widehat{G}_1^{-1}(u_1)), \cdots, F_p^{-1}(\widehat{G}_p^{-1}(u_p))\} > x, \\
&\qquad\qquad \frac{1}{n+1} \le u_1, \cdots, u_p \le \frac{n}{n+1}\Big\},
\end{aligned}
$$

where $\widehat{G}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} I(U_{ij} \le x)$, and U_{ij} is defined in (2.1). Let $\boldsymbol{\theta}_0$ denote the true value of $\boldsymbol{\theta}$. Hence $Q_\xi(\alpha) = Q_\xi(\alpha; \boldsymbol{\theta}_0)$ is the true quantile to be estimated. As we estimate extreme quantiles, we assume $\alpha \equiv \alpha_n \to 0$ as $n \to \infty$.

Some regularity conditions are now in order.

A1. $||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| = O_p(\Delta_n)$ for some $\Delta_n \to 0$ as $n \to \infty$.

A2. For any constant $M > 0$, if

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\alpha_n^{-1} \int_{A(x_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p - 1\Big| \to 0
$$

and

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\alpha_n^{-1} \int_{A(y_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p - 1\Big| \to 0
$$

for sequences $x_n(\boldsymbol{\theta})$ and $y_n(\boldsymbol{\theta})$ as $n \to \infty$, then $\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} |x_n(\boldsymbol{\theta})/y_n(\boldsymbol{\theta}) - 1| \to 0$ as $n \to \infty$.

A3. For any constant $M > 0$, if

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\alpha_n^{-1} \int_{B_n(x_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p - 1\Big| \xrightarrow{p} 0
$$

and

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\alpha_n^{-1} \int_{B_n(y_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p - 1\Big| \xrightarrow{p} 0
$$

for sequences $x_n(\boldsymbol{\theta})$ and $y_n(\boldsymbol{\theta})$ as $n \to \infty$, then $\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} |x_n(\boldsymbol{\theta})/y_n(\boldsymbol{\theta}) - 1| \xrightarrow{p} 0$ as $n \to \infty$.

A4. As $n \to \infty$, it holds for any constant $M > 0$ that

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\frac{\int_{B_n(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p}{\int_{A_n(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p} - 1\Big| \xrightarrow{p} 0.
$$

A5. As $n \to \infty$, it holds for any constant $M > 0$ that

$$
\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big|\frac{\int_{A_n(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p}{\int_{A(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p} - 1\Big| \to 0.
$$

A6. As $n \to \infty$, it holds for any constant $M > 0$ that

$$\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \left| \alpha_n^{-1} \int_{A(Q_\xi(\alpha))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p - 1 \right| \to 0.$$

Theorem 1. Under Conditions A1–A6, $\widehat{Q}_\xi(\alpha)/Q_\xi(\alpha) \xrightarrow{p} 1$ as $n \to \infty$.

Proof. Note that

$$\alpha = \int_{A(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p \tag{3.1}$$

and $\widehat{Q}_\xi(\alpha)$ satisfies

$$\int_{B_n(\widehat{Q}_\xi(\alpha))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p/\alpha = 1 + o_p(1). \tag{3.2}$$

Write

$$\begin{aligned}
& \textstyle\int_{B_n(\widehat{Q}_\xi(\alpha))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p - \alpha \\
= \; & \textstyle\int_{B_n(\widehat{Q}_\xi(\alpha))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p - \int_{B_n(Q_\xi(\alpha;\widehat{\boldsymbol{\theta}}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p \\
& + \textstyle\int_{B_n(Q_\xi(\alpha;\widehat{\boldsymbol{\theta}}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p - \int_{A_n(Q_\xi(\alpha;\widehat{\boldsymbol{\theta}}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p \\
& + \textstyle\int_{A_n(Q_\xi(\alpha;\widehat{\boldsymbol{\theta}}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p - \int_{A(Q_\xi(\alpha;\widehat{\boldsymbol{\theta}}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p.
\end{aligned}$$

Then it follows from (3.1), (3.2) and Conditions A1, A4, A5 that

$$\begin{aligned}
& \frac{1}{\alpha} \int_{B_n(\widehat{Q}_\xi(\alpha))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\, du_1 \cdots du_p \xrightarrow{p} 1, \quad \text{and} \\
& \frac{1}{\alpha} \int_{B_n(Q_\xi(\alpha;\widehat{\theta}))} c(u_1, \cdots, u_p; \widehat{\boldsymbol{\theta}})\; du_1 \cdots du_p \xrightarrow{p} 1
\end{aligned} \tag{3.3}$$

as $n \to \infty$. By (3.3) and Condition A3, we have

$$\widehat{Q}_\xi(\alpha)/Q_\xi(\alpha;\widehat{\theta}) \xrightarrow{p} 1 \tag{3.4}$$

as $n \to \infty$. It follows from (3.1), Conditions A1, A2 and A6 that

$$Q_\xi(\alpha;\widehat{\theta})/Q_\xi(\alpha) \xrightarrow{p} 1. \tag{3.5}$$

Hence, the theorem follows from (3.4) and (3.5). ■

Remark 3. Condition A1 holds with $\Delta_n = 1/\sqrt{n}$ under some regularity conditions as in Genest, Ghoudi and Rivest (1995). Condition A2 implies that the extreme quantile is asymptotically uniquely determined. Condition A3 implies that the extreme quantile is still asymptotically uniquely determined when the marginal distributions are replaced by their empirical counterparts. Condition A4 ensures that sets A_n and B_n are close enough. Condition A5 ensures that there

is no need to extrapolate the marginal distributions below $\widehat{G}_i^{-1}(\frac{1}{n+1})$ and above $\widehat{G}_i^{-1}(\frac{n}{n+1})$. We illustrate those conditions in two examples below.

Example 1: Gumbel Copula. Suppose the distribution of $\mathbf{X}$ is the Gumbel copula

$$C(x_1, \cdots, x_p; \theta) = \exp\big\{ - \big(\sum_{i=1}^{p} (-\log x_i)^{\theta} \big)^{1/\theta} \big\},$$

where $\theta > 0$. Consider $h(\mathbf{X}) = \{\max_{1 \le i \le p} X_i\}^{-1}$ and $\alpha = n^{-\gamma}$ for some $\gamma > 1$. Then $Q_\xi(\alpha; \theta) = n^{\gamma / p^{1/\theta}}$ and $Q_\xi(\alpha) = Q_\xi(\alpha; \theta_0)$. It is easy to check that for any $i = 1, \cdots, p$

$$P\big\{ X_i \le n^{-1}, X_j \le Q_\xi^{-1}(\alpha; \theta) \quad \text{for} \quad j = 1, \cdots, i-1, i+1, \cdots, p \big\} = n^{-(1+\gamma^\theta (p-1)/p)^{1/\theta}}.$$

So when $\gamma < p^{1/\theta}$, we have

$$P(X_i \le n^{-1}, X_j \le Q_\xi^{-1}(\alpha; \theta) \quad \text{for} \quad j = 1, \cdots, i-1, i+1, \cdots, p)/\alpha \to 0,$$

which can be used to prove Condition A5. It is straightforward to verify Conditions A1, A2 and A6 when $\gamma \in (1,\, p^{1/\theta})$. Use the fact that

$$\sup_u \Big| \frac{\sqrt{n}(\widehat{G}_i^-(u) - u)}{u^\delta (1-u)^\delta} I\big(\frac{1}{n+1} \le u \le \frac{n}{n+1} \big) \Big| = O_p(1) \tag{3.6}$$

for any $\delta \in (0, 1/2)$, we can show that for any $\epsilon \in (0, 1)$, the following relation

$$A_n\{(1-\epsilon) x_n(\theta)\} \supset B_n(x_n(\theta)) \supset A_n\{(1+\epsilon) x_n(\theta)\}$$

holds with probability tending to one for any sequence $x_n(\theta)/Q_\xi(\alpha; \theta)$ converging to a positive constant. By the above relation, one can show Conditions A3 and A4 hold when $\gamma \in (1,\, p^{1/\theta})$.

Example 2: Clayton copula. Suppose the distribution of $\mathbf{X}$ is

$$F(x_1, \cdots, x_p; \theta, \beta) = (1 - p + \sum_{i=1}^{p} x_i^{-\beta\theta})^{-1/\theta}$$

for some $\theta > 0$ and $\beta > 0$. Then the copula of $\mathbf{X}$ is the Clayton copula

$$C(u_1, \cdots, u_p; \theta) = (1 - p + \sum_{i=1}^{p} u_i^{-\theta})^{-1/\theta}.$$

Consider $h(\mathbf{X}) = \{\max_{1 \le i \le p} X_i\}^{-1}$ and $\alpha = n^{-\gamma}$ for some $\gamma > 1$. Then $Q_\xi(\alpha; \theta) = (\frac{n^{\gamma\theta} - 1 + p}{p})^{1/(\beta\theta)}$ and $Q_\xi(\alpha) = Q_\xi(\alpha; \theta_0)$. It is easy to check that for any $i = 1, \cdots, p$

$$\begin{aligned} & P(X_i \le n^{-1}, X_j \le Q_\xi^{-1}(\alpha; \theta) \quad \text{for} \quad j = 1, \cdots, i-1, i+1, \cdots, p) \\ = \; & \{1 - p + n^{\beta\theta} + \tfrac{(p-1)(n^{\theta\gamma} - 1 + p)}{p}\}^{-1/\theta}. \end{aligned} \tag{3.7}$$

When $\gamma < \beta$, the right hand side of (3.7) is $o(n^{-\gamma})$, which can be used to show Condition A5 holds. The rest conditions can be verified as Example 1 when $1 < \gamma < \beta$. When the distribution of $\mathbf{X}$ is Clayton copula, i.e., $\beta = 1$ for the above distribution, the right hand side of (3.7) is the same order as $n^{-\gamma}$, which implies that Condition A5 does not hold. That is, the marginals have to be modeled parametrically for estimating this extreme quantile with $\alpha = n^{-\gamma}$ in this case.

Theorem 1 above is generic, imposing the conditions directly on the closeness between the quantile set A and its truncated version A_n, the empirical approximation B_n for A_n. When the copula of $\mathbf{X}$ is of multivariate regular variation (i.e. Condition B2 below) and the quantile set A is asymptotically scalar-invariant (see Condition B1 below), Theorem 2 below shows that the consistency still holds.

B1. Let $S \subset (0,1]^p$ be a set independent of n. When $Q_\xi(0;\boldsymbol{\theta}) = a < \infty$, put $\bar{a}_n(\boldsymbol{\theta}) = a - Q_\xi(\alpha;\boldsymbol{\theta})$ and assume that for any $\epsilon > 0$, there exist $t_0 > 0$ and a positive function $L(t) \to 0$ as $t \to 0$ such that for all $t \le t_0$

$$(1-\epsilon)S \subset A(a-t)/L(t) \subset (1+\epsilon)S.$$

When $Q_\xi(0;\boldsymbol{\theta}) = \infty$, put $\bar{a}_n(\boldsymbol{\theta}) = 1/Q_\xi(\alpha;\boldsymbol{\theta})$ and assume that for any $\epsilon > 0$, there exist $t_0 > 0$ and a positive function $L(t) \to 0$ as $t \to 0$ such that for all $t \le t_0$

$$(1-\epsilon)S \subset A(t)/L(t) \subset (1+\epsilon)S.$$

B2. For any $M > 0$, there exists N such that, as $t \to 0$

$$\sup_{n \ge N} \sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big| \frac{c(tu_1, \cdots, tu_p;\boldsymbol{\theta})}{c(t, \cdots, t;\boldsymbol{\theta})} - l(u_1, \cdots, u_p;\boldsymbol{\theta}) \Big| \to 0$$

for $u_1, \cdots, u_p > 0$, and

$$\sup_{n \ge N} \sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big| \frac{l(tu, \cdots, tu;\boldsymbol{\theta})}{l(t, \cdots, t;\boldsymbol{\theta})} - u^\gamma \Big| = 0$$

for $u > 0$ and some $\gamma \in R$. Further

$$\sup_{n \ge N} \sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \int_S l(u_1, \cdots, u_p;\boldsymbol{\theta})\, du_1 \cdots du_p < \infty.$$

B3. For any $M > 0$,

$$\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \Big| \frac{c(L(\bar{a}_n(\boldsymbol{\theta}_0)), \cdots, L(\bar{a}_n(\boldsymbol{\theta}_0));\boldsymbol{\theta})}{c(L(\bar{a}_n(\boldsymbol{\theta}_0)), \cdots, L(\bar{a}_n(\boldsymbol{\theta}_0));\boldsymbol{\theta}_0)} - 1 \Big| \to 0$$

as $n \to \infty$.

B4. $\lim_{n\to\infty} \inf_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \le M\Delta_n} \{n^\delta L(\bar{a}_n(\boldsymbol{\theta}))\} > 0$ for some $\delta \in (0,1)$.

Theorem 2. Under Conditions A1 and B1–B4, $\widehat{Q}_\xi(\alpha)/Q_\xi(\alpha) \xrightarrow{p} 1$ as $n \to \infty$.

Proof. We shall verify conditions A2–A6 in Theorem 1. By B1, we can write

$$\begin{aligned}\alpha &= \int_{A(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p \\ &\geq \int_{(1-\epsilon)SL(\bar{a}_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p \\ &= \int_{(1-\epsilon)S} c(L(\bar{a}_n(\boldsymbol{\theta}))u_1, \cdots, L(\bar{a}_n(\boldsymbol{\theta}))u_p; \boldsymbol{\theta}) L^p(\bar{a}_n(\boldsymbol{\theta}))\, du_1 \cdots du_p\end{aligned}$$

and

$$\begin{aligned}\alpha &= \int_{A(Q_\xi(\alpha;\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p \\ &\leq \int_{(1+\epsilon)SL(\bar{a}_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p \\ &= \int_{(1+\epsilon)S} c(L(\bar{a}_n(\boldsymbol{\theta}))u_1, \cdots, L(\bar{a}_n(\boldsymbol{\theta}))u_p; \boldsymbol{\theta}) L^p(\bar{a}_n(\boldsymbol{\theta}))\, du_1 \cdots du_p\end{aligned}$$

as n large enough. Hence it follows from A1 and B2 that

$$\frac{\alpha}{c(L(\bar{a}_n(\boldsymbol{\theta})), \cdots, L(\bar{a}_n(\boldsymbol{\theta})); \boldsymbol{\theta}) L^p(\bar{a}_n(\boldsymbol{\theta}))} = \int_S l(u_1, \cdots, u_p; \boldsymbol{\theta})\, du_1 \cdots du_p. \tag{3.8}$$

Like the proof of (3.8), condition A2 can be shown by using B2. Note that B1 and B4 imply that $A_n(Q_\xi(\alpha;\boldsymbol{\theta})) = A(Q_\xi(\alpha;\boldsymbol{\theta}))$ for $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq M\Delta_n$ and large n. Hence Condition A5 holds. Using (3.6) we can show condition A4. Note that $\alpha^{-1} \int_{B_n(x_n(\boldsymbol{\theta}))} c(u_1, \cdots, u_p; \boldsymbol{\theta}) \to 1$ implies that $x_n(\boldsymbol{\theta}) \to Q_\xi(\alpha;\boldsymbol{\theta})$. Hence, like the proof of (3.8), we can show A3 by using (3.6), B1 and B2. Condition A6 follows from B2 and B3. Hence, Theorem 2 follows from Theorem 1. ■

Remark 4. Condition B1 relates the set A to a fixed set S by a scaling factor depending on the sample size n. This idea appeared in Drees and de Haan (2013). Condition B2 assumes the copula density is of a multivariate regular variation. We refer to Resnick (1987) for more details on multivariate regular variation. It follows from Condition B2 that $c(L(\bar{a}_n(\boldsymbol{\theta}_0)), \cdots, L(\bar{a}_n(\boldsymbol{\theta}_0)); \boldsymbol{\theta}) = O(L^{\gamma-\epsilon}(\bar{a}_n(\boldsymbol{\theta}_0)))$ for any $\epsilon > 0$. Hence, (3.8) implies

$$\alpha = \alpha_n = O(L^{\gamma+p-\epsilon}(\bar{a}_n(\boldsymbol{\theta}_0)))$$

for any $\epsilon > 0$. This reflects the fact that how small α_n can be depends on the geometry of the set A (i.e., $L(\bar{a}_n(\boldsymbol{\theta}_0))$), the property of the copula (i.e., γ) and the dimension (i.e., p). It is straightforward to check that Conditions B1–B4 hold for the above two examples on Gumbel copula and Clayton copula with $L(t) = t^\beta$ and $\beta > \gamma_0$ for $\alpha = n^{-\gamma_0}$.

4 Numerical properties

In this section we illustrate the proposed method by simulation. We let $\mathbf{X} = (X_1, \cdots, X_p)'$, where

$$X_j = 1.2X_{j-1} - 0.6X_{j-2} + \varepsilon_j, \qquad j = 1, \cdots, p, \tag{4.1}$$

and ε_j are independent and identically distributed random variables. We estimate the extreme quantiles of the following four functions:

$$h_1(\mathbf{X}) = X_{(p)} + X_{(p-1)} + X_{(p-2)}, \qquad h_2(\mathbf{X}) = \min_{1\le j\le p} F_j(X_j),$$
$$h_3(\mathbf{X}) = \frac{1}{p}\sum_{j=1}^{p} X_j, \qquad h_4(\mathbf{X}) = \frac{1}{p}\sum_{j=1}^{p}\{1 - F_j(X_j)\},$$

where $X_{(1)} \le \cdots \le X_{(p)}$ are the order statistics of the components of $\mathbf{X}$, $F_j(\cdot)$ is the CDF of the j-th component of $\mathbf{X}$, and hence $F_j(X_j) \sim U(0,1)$.

We consider two distributions for ε_t in (4.1), namely the standard normal $N(0,1)$, and Student's t-distribution with 4 degrees of freedom t_4. With a sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ drawn from the distribution of $\mathbf{X}$, we estimate the $(1-\alpha)$-th quantile with $\alpha = 0.05$, 0.01, 0.005, 0.001 and 0.0005. We set the sample size $n = 500$ or 1,000, and the dimension $p = 20$ or 40. For each sample, we fit the data with three D-vine copulas:

Copula I: two trees only (i.e. $m = 2$ in (2.8)) with Gaussian binary copulas.

Copula II: two trees only with all binary copulas selected by the AIC.

Copula III: the number of trees and all binary copulas are selected by the AIC.

Since $X_t \sim$ AR(2) (see (4.1)), X_t and X_{t+3} are independent conditionally on X_{t+1} and X_{t+2}. Hence the dependence structure of $\mathbf{X}$ can be represented by a D-vine with two trees, i.e. Copula II reflects the underlying dependence structure correctly. Furthermore Copula I specifies the correct parametric model when $\varepsilon_t \sim N(0,1)$ in (4.1).

The computation was carried out using the R-package `CDVine` which selected binary copulas from a large number of copula families; see `cran.r-project.org/web/packages/CDVine/CDVine.pdf`. We let $m = 40,000$ in (2.10).

For each setting, we drew 400 samples, i.e. replicated the estimation 400 times. We calculate the Mean Absolute Relative Error (MARE):

$$\text{MARE} = \frac{1}{400}\sum_{\text{i}=1}^{400}\Big|\frac{\widehat{\text{Q}}_{\text{i}} - \text{Q}}{\text{Q}}\Big|, \tag{4.2}$$

where Q denotes the true quantile value, and $\widehat{Q}_1, \cdots, \widehat{Q}_{400}$ denote its estimated values over 400 replications. The true values of the extreme quantiles for $h_1(\mathbf{X}), \cdots, h_4(\mathbf{X})$ were calculated by a simulation with a sample of size 500,000. For the comparison purpose, we also include the simple sample quantile estimate $\xi_{[n\alpha]}$ from an original samples, where $\xi_{[j]}$ denotes the j-th largest value among $\xi_k \equiv h_i(\mathbf{X}_k)$ for $k = 1, \cdots, n$, and $i = 1, \cdots, 4$.

Table 1 lists the MARE with sample size $n = 500$ and $\mathbf{X}$ consisting of $p = 20$ successive values of the AR(2) process defined by (4.1) with standard normal innovations. Since Copula I is the

Table 1: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 500$, $p = 20$ and $\varepsilon_t \sim N(0,1)$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0161	.0259	.0337	.0587	.0721
	Copula II	.0167	.0256	.0327	.0603	.0720
	Copula III	.0169	.0258	.0327	.0597	.0709
	sample quantile	.0231	.0373	.0476	.0841	n/a
$h_2(\mathbf{X})$	Copula I	.0082	.0103	.0119	.0151	.0199
	Copula II	.0128	.0125	.0126	.0169	.0189
	Copula III	.0138	.0132	.0130	.0168	.0213
	sample quantile	.0404	.0586	.0718	.1069	n/a
$h_3(\mathbf{X})$	Copula I	.0260	.0216	.0204	.0215	.0227
	Copula II	.0277	.0253	.0258	.0287	.0291
	Copula III	.0289	.0257	.0262	.0283	.0293
	sample quantile	.0463	.0572	.0632	.1020	n/a
$h_4(\mathbf{X})$	Copula I	.0028	.0035	.0041	.0050	.0064
	Copula II	.0035	.0045	.0050	.0056	.0063
	Copula III	.0042	.0051	.0057	.0066	.0074
	sample quantile	.0097	.0167	.0196	.0328	n/a

true parametric family for the underlying distribution, it yields the better estimates than Copulas II and III. Note that both Copulas II and III are still correct models with more parameters to be specified. The differences from using three copulas are not substantial; indicating that the AIC worked well in choosing binary copula functions (for Copulas II and III) as well as specifying the number of trees (for Copula III). Also the MARE tends to increase when α decreases; indicating the increasing difficulty in estimating more extreme quantiles. In fact we reported in the table the MARE which is defined as the mean absolute error (MAE) divided by the true quantile value; see (4.2). In fact the MAE strictly increases when α decreases. Figure 2 displays the boxplots of the estimation errors (i.e. $\widehat{Q}_i - Q$, $i = 1, \cdots, 400$; see (4.2)) for the estimation with Copula I, $n = 500$ and $p = 20$. It shows clearly that both the bias and variance of the estimators increase when α decreases. Note that $n\alpha$ ranges from 25 to 0.25 for $0.05 \geq \alpha \geq 0.0005$. For the most extreme case with $\alpha = 0.0005$, we extrapolate far out of the range covered by data $\{h_i(\mathbf{X}_t),\ t = 1, \cdots, n\}$. Still the maximum MARE is under 8% with function $h_1(\mathbf{X})$, is under 3% with $h_3(\mathbf{X})$, and is even smaller with $h_2(\mathbf{X})$ and $h_4(\mathbf{X})$. We also notice that the extreme quantiles of $h_2(\mathbf{X})$ and $h_4(\mathbf{X})$ can be estimated much more accurately than those of $h_1(\mathbf{X})$ and $h_3(\mathbf{X})$. This is due to the fact that h_2 and h_4 are the function of the marginal distribution functions of $\mathbf{X}$. Therefore they are effectively the functions of a p random vector with all the marginal distributions being $U(0,1)$. Furthermore, their estimates do not suffer from the errors due to the inverse empirical transformations (2.11) in the bootstrap resampling. Overall with normal $\mathbf{X}$, the proposed estimation method works

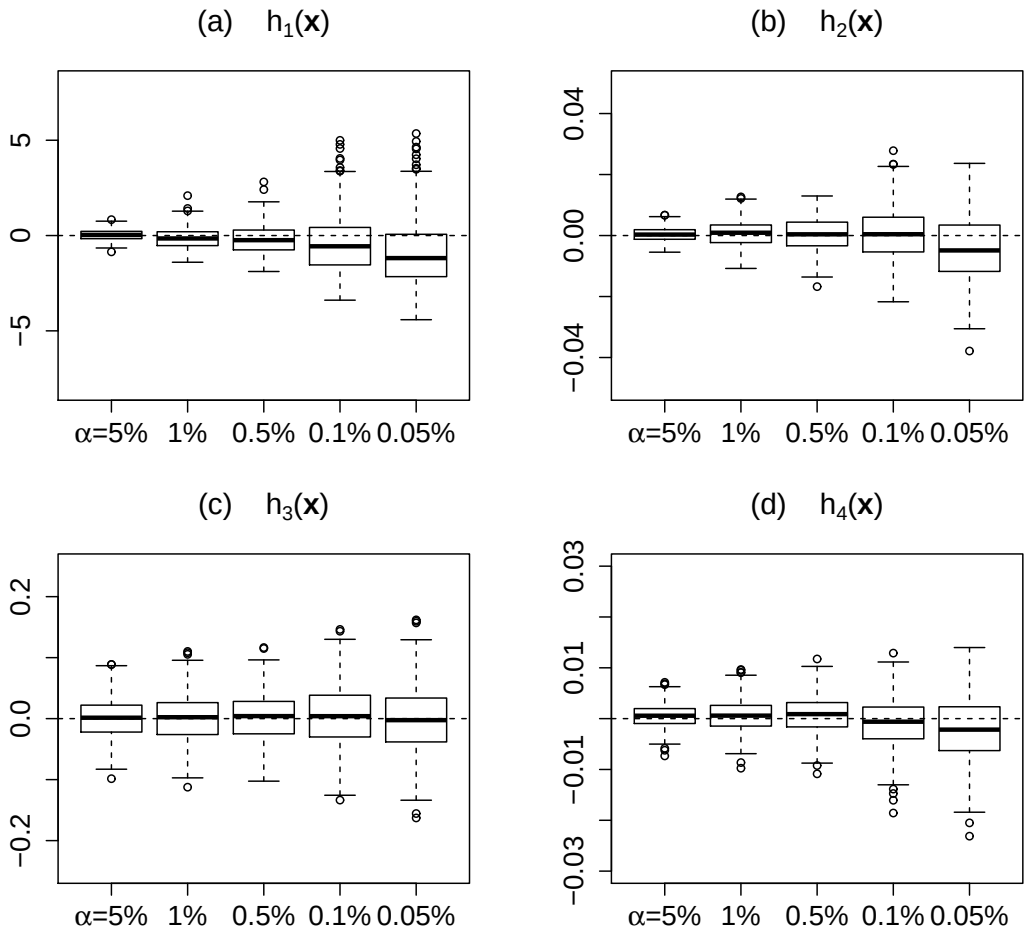

Figure 2: Boxplots of the errors in estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $\varepsilon_t \sim N(0,1)$. Copula I was used in estimation with $n = 500$ and $p = 20$.

very well. It provides much more accurate estimates than the simple sample quantiles even for $\alpha = 0.05$ when there are $n\alpha = 25$ data points in the top α-tails. With sample size $n = 500$ or 1000, the sample quantiles at the $(1-\alpha)$-th level when $\alpha = 0.0005$ are not available.

Tables 2–5 list the MARE when $\varepsilon_t \sim t_4$ in (4.1). Now components of $\mathbf{X}$ are heavy-tailed with $E(||\mathbf{X}||^4) = \infty$. The extreme quantiles to be estimated are more likely to be impacted by the extreme values of the components of $\mathbf{X}$ than the cases with $\varepsilon_t \sim N(0,1)$. The MARE with $\alpha = 0.001$ and 0.0005 in Tables 2–5 tend to be too large with functions $h_1(\mathbf{X})$ and $h_3(\mathbf{X})$, while the estimation for the extreme quantiles of $h_2(\mathbf{X})$ and $h_4(\mathbf{X})$ remains accurate with the MARE smaller than 3%. Nevertheless when the sample size increases from $n = 500$ to $n = 1000$, the MARE decreases. When the number of components of $\mathbf{X}$ increases from $p = 20$ to $p = 40$, the MARE with $h_1(\mathbf{X})$ or $h_2(\mathbf{X})$ increases while that with $h_3(\mathbf{X})$ and $h_4(\mathbf{X})$ decreases. Note that $h_1(\mathbf{X})$ or $h_2(\mathbf{X})$ are extreme functions of the components $\mathbf{X}$, and they become more extreme when

Table 2: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 500$, $p = 20$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0277	.0601	.0918	.2070	.2766
	Copula II	.0277	.0582	.0853	.1887	.2383
	Copula III	.0277	.0576	.0808	.1866	.2399
	sample quantile	.0332	.0703	.0950	.2456	n/a
$h_2(\mathbf{X})$	Copula I	.0264	.0114	.0112	.0197	.0260
	Copula II	.0094	.0181	.0224	.0210	.0214
	Copula III	.0104	.0188	.0227	.0227	.0220
	sample quantile	.0401	.0579	.0657	.1046	n/a
$h_3(\mathbf{X})$	Copula I	.0309	.0289	.0334	.0626	.0848
	Copula II	.0357	.0702	.0897	.1339	.1405
	Copula III	.0370	.0696	.0904	.1343	.1377
	sample quantile	.0496	.0651	.0738	.1569	n/a
$h_4(\mathbf{X})$	Copula I	.0063	.0080	.0075	.0078	.0089
	Copula II	.0029	.0045	.0048	.0067	.0090
	Copula III	.0037	.0051	.0054	.0075	.0093
	sample quantile	.0089	.0163	.0192	.0305	n/a

Table 3: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 500$, $p = 40$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0290	.0802	.1234	.2349	.2868
	Copula II	.0287	.0635	.0943	.2095	.2435
	Copula III	.0288	.0635	.0932	.2096	.2424
	sample quantile	.0326	.0748	.1014	.2746	n/a
$h_2(\mathbf{X})$	Copula I	.0613	.0379	.0299	.0235	.0248
	Copula II	.0255	.0144	.0132	.0156	.0202
	Copula III	.0280	.0173	.0157	.0169	.0208
	sample quantile	.0427	.0654	.0769	.1188	n/a
$h_3(\mathbf{X})$	Copula I	.0283	.0248	.0269	.0440	.0604
	Copula II	.0379	.0667	.0861	.1114	.1179
	Copula III	.0377	.0659	.0864	.1147	.1204
	sample quantile	.0486	.0657	.0742	.1350	n/a
$h_4(\mathbf{X})$	Copula I	.0051	.0069	.0075	.0075	.0066
	Copula II	.0018	.0029	.0036	.0045	.0050
	Copula III	.0026	.0041	.0046	.0053	.0058
	sample quantile	.0065	.0114	.0144	.0253	n/a

p increases. In contrast, $h_3(\mathbf{X})$ or $h_4(\mathbf{X})$ are the means of the components of $\mathbf{X}$, they behave more like normal when p increases due the CLT. With $\varepsilon_t \sim t_4$, Copula I misspecified the model while Copula II provides a correct dependence structure (i.e. a D-vine with two trees only). With

Table 4: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 1000$, $p = 20$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0225	.0475	.0689	.1786	.2441
	Copula II	.0210	.0422	.0586	.1419	.2043
	Copula III	.0207	.0424	.0587	.1370	.2039
	sample quantile	.0256	.0516	.0721	.1630	n/a
$h_2(\mathbf{X})$	Copula I	.0282	.0111	.0097	.0186	.0265
	Copula II	.0079	.0144	.0185	.0167	.0168
	Copula III	.0087	.0147	.0186	.0183	.0201
	sample quantile	.0279	.0407	.0499	.0764	n/a
$h_3(\mathbf{X})$	Copula I	.0226	.0210	.0258	.0556	.0808
	Copula II	.0223	.0410	.0563	.0990	.1128
	Copula III	.0237	.0412	.0577	.0988	.1150
	sample quantile	.0351	.0466	.0589	.1065	n/a
$h_4(\mathbf{X})$	Copula I	.0052	.0069	.0061	.0062	.0077
	Copula II	.0021	.0028	.0032	.0050	.0067
	Copula III	.0029	.0037	.0042	.0055	.0073
	sample quantile	.0061	.0102	.0144	.0258	n/a

the functions $h_1(\mathbf{X}), h_2(\mathbf{X})$ and $h_4(\mathbf{X})$, the Gaussian copula (i.e. Copula I) is the least preferable, the estimation with Copula II leads to smaller MARE than those with Copula III across Tables 2–5 although the differences are not substantial, and are certainly smaller than the differences between the estimates based on Copula II and those based on Copula I. However with $h_3(\mathbf{X})$, the estimation with the Gaussian copula is the best. One possible explanation is that with $p = 20$ or $p = 40$, it holds approximately that

$$h_3(\mathbf{X}) = \frac{1}{p}\sum_{t=1}^{p} X_t \;\sim\; N\big(0,\; \frac{1}{p}\mathrm{Var}(X_1) + \frac{2}{p}\sum_{k=2}^{p}(1-\frac{k-1}{p})\mathrm{Cov}(X_1, X_k)\big).$$

Since the Gaussian copula also specifies the correlation among the components of $\mathbf{X}$ correctly, it is an approximately correct parametric model. Overall the proposed method provides more, or much more, accurate estimates than the sample quantiles across Tables 2–5.

For further illustration, we now repeat the above exercise with (4.1) replaced by the MA(2) model:

$$X_j = \varepsilon_j + 1.5\varepsilon_{j-1} - 0.5\varepsilon_{j-2}, \qquad j = 1, \cdots, p,$$

where ε_t are independent t_4 random variables. Now X_{j+3} and X_j are no longer conditional independent given X_{j+2} and X_{j+1}. Hence both Copulas I and II represent misspecified models.

For comparison with the methods based on univariate extreme value theory (EVT), we employ

Table 5: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 1000$, $p = 40$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0191	.0595	.0957	.2167	.2595
	Copula II	.0187	.0426	.0631	.1638	.2292
	Copula III	.0181	.0414	.0617	.1673	.2318
	sample quantile	.0242	.0557	.0754	.1717	n/a
$h_2(\mathbf{X})$	Copula I	.0618	.0388	.0298	.0235	.0242
	Copula II	.0250	.0152	.0136	.0148	.0180
	Copula III	.0262	.0158	.0143	.0163	.0212
	sample quantile	.0309	.0438	.0559	.0871	n/a
$h_3(\mathbf{X})$	Copula I	.0207	.0195	.0209	.0397	.0573
	Copula II	.0218	.0412	.0576	.0832	.0948
	Copula III	.0226	.0408	.0575	.0827	.0990
	sample quantile	.0356	.0471	.0542	.0969	n/a
$h_4(\mathbf{X})$	Copula I	.0047	.0066	.0070	.0073	.0063
	Copula II	.0012	.0020	.0024	.0034	.0042
	Copula III	.0020	.0028	.0032	.0043	.0052
	sample quantile	.0048	.0082	.0106	.0194	n/a

the quantile estimator

$$\widehat{Q}_{\xi}^{EVT}(\alpha) = \xi_{n,n-k} + \widehat{\sigma}_M \frac{(\frac{k}{n\alpha})^{\widehat{\gamma}_M} - 1}{\widehat{\gamma}_M}, \tag{4.3}$$

where

$$M_n^{(j)} = \frac{1}{k}\sum_{i=0}^{k-1}(\log \xi_{n,n-i} - \log \xi_{n,n-k})^j, \quad \widehat{\gamma}_M = M_n^{(1)} + 1 - \frac{1}{2}\{1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\}^{-1},$$

$$\widehat{\gamma}_- = 1 - \frac{1}{2}\{1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}}\}^{-1}, \quad \widehat{\sigma}_M = \xi_{n,n-k} M_n^{(1)}(1 - \widehat{\gamma}_-),$$

and $\xi_{n,1} \leq \cdots \leq \xi_{n,n}$ denote the order statistics of $\xi_1, \cdots, \xi_n$. See Chapter 4.3 of De Haan and Ferreira (2006) for details. Note that this quantile estimator depends on the tuning parameter k, the number of upper order statistics used. Although some data-driven methods for choosing k exist in the literature, we calculate the above estimator for different values of k, i.e. $k = 25, 50, 75, 100, 200, 300$.

As before we report MARE defined in (5.2) in Tables 6 and 7. Now Copula I (i.e. a Gaussian copula with two tree) is a very wrong model in terms of both the tree structure and the distributions at all nodes. It should not be used in practice, as it leads to larger MAREs than those with Copulas II and III. The performances from using Copulas II and III are comparable. This may be due to the fact that for $i \geq 2$, the dependence between X_{j+i} and X_j, conditionally on $X_{j+i-1}, \cdots X_{j+1}$, is not too strong to be overlooked.

The EVT-based method with all chosen values of k provides overall less or much less accurate estimates than the proposed method with Copulas II and III. The exception occurs for the estimation of $h_1(\mathbf{X}) = X_{(p)} + X_{(p-1)} + X_{(p-2)}$ with $\alpha \leq 0.005$. This is due to the fact that the distribution of h_1 can be approximated well by an extreme value distribution. Also the estimation may fluctuate with k, especially for the small values of k.

Table 6: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 1000$, $p = 20$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0361	.0984	.1301	.2204	.2628
	Copula II	.0179	.0391	.0554	.1327	.1603
	Copula III	.0181	.0382	.0553	.1333	.1559
	EVT(k=25)	.0581	.0477	.0623	.1220	.1560
	EVT(k=50)	.0219	.0439	.0577	.1152	.1465
	EVT(k=75)	.0203	.0417	.0561	.1116	.1397
	EVT(k=100)	.0203	.0411	.0557	.1096	.1355
	EVT(k=200)	.0199	.0413	.0555	.1080	.1318
	EVT(k=300)	.0208	.0424	.0545	.1066	.1277
$h_2(\mathbf{X})$	Copula I	.0740	.0565	.0478	.0392	.0414
	Copula II	.0093	.0146	.0151	.0226	.0287
	Copula III	.0116	.0126	.0146	.0177	.0197
	EVT(k=25)	.0760	.0408	.0442	.0677	.0884
	EVT(k=50)	.0300	.0376	.0421	.0723	.0927
	EVT(k=75)	.0271	.0360	.0427	.0765	.0963
	EVT(k=100)	.0266	.0357	.0442	.0800	.0992
	EVT(k=200)	.0247	.0391	.0536	.0944	.1126
	EVT(k=300)	.0241	.0476	.0680	.1189	.1400
$h_3(\mathbf{X})$	Copula I	.0498	.0649	.0751	.1051	.1263
	Copula II	.0335	.0226	.0285	.0617	.0768
	Copula III	.0228	.0393	.0513	.0931	.1071
	EVT(k=25)	.0825	.0433	.0494	.0944	.1272
	EVT(k=50)	.0325	.0397	.0476	.0913	.1191
	EVT(k=75)	.0297	.0390	.0491	.0923	.1174
	EVT(k=100)	.0298	.0393	.0505	.0920	.1142
	EVT(k=200)	.0298	.0441	.0583	.0998	.1225
	EVT(k=300)	.0438	.0662	.0865	.1415	.1716
$h_4(\mathbf{X})$	Copula I	.0067	.0081	.0085	.0086	.0090
	Copula II	.0061	.0063	.0059	.0050	.0053
	Copula III	.0031	.0049	.0058	.0088	.0103
	EVT(k=25)	.0138	.0088	.0106	.0179	.0230
	EVT(k=50)	.0052	.0083	.0100	.0178	.0225
	EVT(k=75)	.0047	.0081	.0100	.0179	.0225
	EVT(k=100)	.0047	.0079	.0100	.0184	.0231
	EVT(k=200)	.0045	.0081	.0108	.0192	.0236
	EVT(k=300)	.0044	.0083	.0115	.0225	.0285

Table 7: MARE for estimating the $(1-\alpha)$-th quantiles of $h_i(\mathbf{X})$ $(i = 1, \cdots, 4)$ with $n = 1000$, $p = 40$ and $\varepsilon_t \sim t_4$.

Function	Model	$\alpha = .05$	$\alpha = .01$	$\alpha = .005$	$\alpha = .001$	$\alpha = .0005$
$h_1(\mathbf{X})$	Copula I	.0352	.0993	.1297	.2216	.2600
	Copula II	.0181	.0402	.0609	.1516	.1801
	Copula III	.0174	.0388	.0592	.1533	.1776
	EVT(k=25)	.0534	.0455	.0618	.1267	.1673
	EVT(k=50)	.0222	.1599	.0592	.1188	.1548
	EVT(k=75)	.0204	.0426	.0582	.1160	.1496
	EVT(k=100)	.0202	.0418	.0575	.1129	.1444
	EVT(k=200)	.0199	.0413	.0568	.1053	.1302
	EVT(k=300)	.0198	.0421	.0568	.1025	.1238
$h_2(\mathbf{X})$	Copula I	.0984	.0819	.0779	.0745	.0702
	Copula II	.0075	.0086	.0109	.0196	.0234
	Copula III	.0132	.0166	.0163	.0173	.0216
	EVT(k=25)	.0786	.0437	.0511	.0835	.1058
	EVT(k=50)	.0298	.1825	.0491	.0879	.1117
	EVT(k=75)	.0261	.0384	.0488	.0906	.1158
	EVT(k=100)	.0257	.0387	.0499	.0908	.1146
	EVT(k=200)	.0247	.0433	.0607	.1091	.1355
	EVT(k=300)	.0237	.0518	.0741	.1325	.1642
$h_3(\mathbf{X})$	Copula I	.0553	.0636	.0666	.0936	.1075
	Copula II	.0333	.0237	.0245	.0534	.0671
	Copula III	.0232	.0336	.0440	.0814	.0951
	EVT(k=25)	.0730	.0421	.0502	.0905	.1182
	EVT(k=50)	.0346	.1766	.0501	.0918	.1164
	EVT(k=75)	.0328	.0399	.0513	.0964	.1205
	EVT(k=100)	.0332	.0402	.0536	.0996	.1224
	EVT(k=200)	.0337	.0468	.0619	.1036	.1233
	EVT(k=300)	.0494	.0674	.0840	.1315	.1551
$h_4(\mathbf{X})$	Copula I	.0050	.0067	.0066	.0063	.0074
	Copula II	.0046	.0057	.0054	.0047	.0060
	Copula III	.0021	.0031	.0040	.0063	.0066
	EVT(k=25)	.0098	.0063	.0076	.0134	.0179
	EVT(k=50)	.0038	.0295	.0073	.0133	.0174
	EVT(k=75)	.0034	.0058	.0072	.0138	.0179
	EVT(k=100)	.0035	.0056	.0072	.0141	.0182
	EVT(k=200)	.0035	.0056	.0075	.0142	.0184
	EVT(k=300)	.0034	.0061	.0084	.0165	.0219

5 Application in backtesting

This study was motivated by a real world backtesting problem in financial risk management. Under the current Basel III regulatory framework (see Basel Committee, 2011), investment banks

are required to hold adequate capital to cover the counterparty credit risk (CCR), which is the potential loss in derivative positions due to the default of trading counterparties. The failures of Lehman Brothers and MF Global are recent examples of such disruptive events. The Basel III CCR capital is typically computed by jointly simulating various market risk factors (such as interest rates, equities and foreign exchange rates) into the future, and valuing all the derivative positions of the bank at each time horizon of each simulated market scenario paths to determine the potential loss due to counterparty's default. The market risk factors are typically modeled as stochastic processes that are calibrated to corresponding historical time series.

The Basel III CCR capital is one of the most complicated modeling problems for investment banks, as it typically requires the modeling and simulation of tens of thousands of risk factors, and the valuation of millions of trades under each simulated market scenarios. In practice, it is only possible to run the simulation and valuation for a small number (e.g., hundreds to thousands) of paths, due to the large scale of the problem and banks' IT system limitations.

Backtesting is a critical component in the Basel III regulation; it is the primary analytical tool for a bank and its regulators to monitor the performance of its risk factor simulation and valuation models. Figure 3 is an illustration of the backtesting setup for a risk factor, where the actual realization of the risk factor path is represented by $(X_1, \cdots, X_p)$; and solid curves represent the risk factor distributions at different time horizon according to the risk factor model. If the actual realization of the risk factor path is deemed an extreme event with very small probability (the typical threshold is 0.01%), then a 'red light' is designated, which is a strong indication of misspecified simulation models. In practice, the same backtesting procedure is also applied to the trade or portfolio prices in addition to the market risk factors, in which case the backtesting also covers the pricing models and the correlations between risk factors. Mitigations, such as additional capital add-ons for the affected risk factors and/or trades, are often required to ensure capital adequacy in case the 'red light' persists.

There are two technical issues in the backtesting process described above. First it is difficult to test the multiple distributions along different tenors directly. A common practice is to use an appropriate risk metric $\xi = h(X_1, \cdots, X_p)$ which can be viewed as a test statistic for the original backtesting problem. Therefore we need to evaluate the extreme quantile for ξ under the distribution determined by the simulation and pricing models. Secondly, the simulation and pricing models used in practice do not admit explicit solutions, e.g. the distributions displayed in Figure 3 do not admit explicit formulas. Therefore a small number of paths (typically hundreds to thousands) are drawn from the simulation and pricing models instead, which are regarded as distributions of the risk factors or trade prices. In order to accurately designate the red flag, we need to calculate an extreme quantile of $\alpha = 0.01\%$ for $\xi = h(X_1, \cdots, X_p)$ based on a small

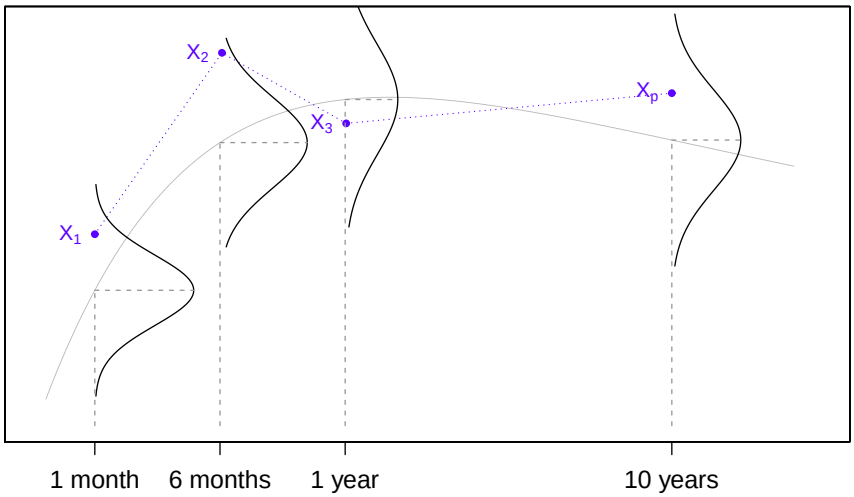

Figure 3: Theoretical risk factor distribution varies with respect to tenor (such as 1 month, 6 months, 1 year, $\cdots$, 10 year). The blue path $(X_1, X_2, \cdots, X_p)$ represents the realized risk factor paths (e.g. equity prices). The goal of a backtesting is to test the hypothesis that the theoretical distributions are correct based on the observed path $(X_1, X_2, \cdots, X_p)$. A 'red light' is designated if the observed path $(X_1, X_2, \cdots, X_p)$ is regarded as extreme event with probability less than 0.01% under the hypothesis.

sample size of hundreds to thousands. The number of tenors p varies, but is typically in the range of 10 to 200, depending on the trade maturities.

One would think that the problem of estimating extreme quantiles can be easily resolved by generating more paths from the simulation and pricing models. However, the sheer volume of trades and risk factors make this brute-force approach infeasible in practice; as we discussed earlier, it is prohibitively expensive to run adequate number of simulation paths and valuations for an extreme quantile of $\alpha = 0.01\%$.

Hereby we use a GJR-GARCH model to illustrate how our proposed method provides an adequate solution for such backtesting problems. We fit the following GJR-GARCH(1,1,1) model for the daily return volatility process of the S&P 500 equity index:

$$\varepsilon_t = \sigma_t e_t, \qquad \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I(\varepsilon_t < 0) + \beta\sigma_{t-1}^2, \tag{5.1}$$

where e_t is independent and $N(0,1)$, $\varepsilon_t = \log(P_t/P_{t-1}) - \mu$ is the centered daily logarithmic return, where P_t is the index price and μ is historical average of the daily $\log(P_t/P_{t-1})$, so that ε_t is constructed to have zero mean. Furthermore the parameter ω is assumed to be positive, and α, γ, β are non-negative. It can be viewed as the standard GARCH(1,1) model with an addition of asymmetric term (with coefficient γ) which captures the well established empirical fact that the down movements in stock prices impacts the volatility much more than the up movements. Under the condition $\alpha + \gamma/2 + \beta < 1$, σ_t defined by (5.1) is strictly positive with the long-run

variance

$$\sigma^2 \equiv \mathrm{Var}(\varepsilon_t) = \omega/\{1-(\alpha+\gamma/2+\beta)\} > 0.$$

We refer to Glosten (1993) for the further details of the GJR-GARCH model.

We estimate the parameters in GJR-GARCH(1,1,1) model (5.1) by the maximum likelihood method using the historical daily prices of S&P 500 in 3 January 2005 – 18 June 2014. The estimated values are:

$$\widehat{\sigma}^2 = 0.0100, \quad \widehat{\alpha} = 0.000, \quad \widehat{\gamma} = 0.1696, \quad \widehat{\beta} = 0.8949.$$

Now suppose we use the fitted GJR-GARCH(1,1,1) model to simulate the future evolution of S&P 500 index. To backtest the model's performance, we draw 1000 simulated paths up to 10 years from the fitted GJR-GARCH(1,1,1) model, and we use the average quantile of the quarterly prices over the 10 year period as the test metric:

$$\xi = \frac{1}{40}\sum_{j=1}^{40} \widehat{F}_j(X_j), \tag{5.2}$$

where X_j denotes the price at the j-th quarter in the 10 year period, and $\widehat{F}_j(\cdot)$ is the empirical distribution of X_j (based on the 1000 observations). Different metrics, such as a weighted mean, a geometric mean or some extreme values of $\widehat{F}_j(X_j)$, are used in practice, to test different aspects of the model. We use the proposed method, fitting directly the data $Z_j \equiv \widehat{F}_j(X_j)$ with the three types of copula specification used in section 4, to estimate the extreme quantiles in the top tail of the distribution of ξ. For the comparison purpose, we also report the estimates obtained based on the extreme value theory method using the k extreme observations; see (4.3). All the estimates are listed in Table 8. To assess the goodness of the different estimation method, we also report the true values of the extreme quantiles of ξ which were evaluated based on 500,000 simulations from the fitted GJR-GARCH(1,1,1) model. The proposed method with Copulas II & III provides reasonable estimates. In fact the performances with II and III are close and the method with Copula II is slightly better. When $\alpha \leq 0.01$, the Gaussian copula (i.e. Copula I) performs worse than the two others. Note that X_j is not stationary in j and ξ defined in (5.2) is far from normal. On the other hand the EVT method with all chosen values of k performs worse or substantially worse than the proposed method with Copulas II & III when $\alpha \leq 0.01$. Furthermore the EVT method is sensitive to the choice of k.

Conclusions

We propose in this paper a new method for estimating the extreme quantiles of a function of several random variables. The extreme quantiles concerned are typically outside the range of the

Table 8: Estimated $(1-\alpha)$-th quantiles of ξ defined in (5.2) by the proposed methods with 3 copulas specified in section 4, and also the EVT methods using k extreme observations.

α	.1000	.0100	.0010	.0001
True quantile	.8207	.9481	.9831	.9946
Copula I	.8160	.9406	.9758	.9889
Copula II	.8110	.9432	.9806	.9929
Copula III	.8101	.9425	.9804	.9928
EVT ($k=25$)	.7589	.9428	.9690	.9728
EVT ($k=50$)	.8194	.9398	.9786	.9910
EVT ($k=75$)	.8146	.9402	.9761	.9863
EVT ($k=100$)	.8112	.9392	.9685	.9752
EVT ($k=200$)	.8109	.9379	.9705	.9788
EVT ($k=300$)	.8094	.9329	.9608	.9671

observed data. Unlike the standard methods based on extreme value theory, the new method models the marginal distributions nonparametrically, and fits the high dimensional dependence structure within the observed range with a vine copula model. Hence the new method does not impose any explicit parametric forms on the tails of the underlying distribution, and it avoids the difficulties in choosing a fraction of the sample to be used for estimation.

The underpinning idea of the new method is that it is not necessary to go to extremes along any component variable in order to observe a joint extreme event. This also indicates that the method may fail to handle excessively extreme cases. How extreme it can do depends on the underlying distribution and the number of the variables involved. Nevertheless if the function concerned depends on each random variable through its CDF transformation (such as $h_2(\cdot)$ and $h_4(\cdot)$ used in section 4, the risk metric used in section 5), we effectively deal with the cases when all random variables are bounded. Then the new method can provide accurate estimation for very extreme quantiles.

The proposed method fits the dependence among $X_1, \cdots, X_p$ within the observed range by a vine copula. When those components can be ordered such that the dependence between X_i and X_j decays as $|i-j|$ increases, some parsimonious fitting may be obtained by using a D-vine copula; see Remarks 1(iii-iv). The empirical evidences reported in sections 4 and 5 also indicate that the copulas specified by AIC (such as Copula III) or partially specified by AIC (such as Copula II) often provide satisfactory estimates. On the other hand, Gaussian copulas should not be used in general unless the data are normal or close to normal. However, for the functions in the form $\xi = h(p^{-1}\sum_j g(X_j))$ with moderately large p and stationary X_j (in j), fitting a Gaussian copula to capture the dependence (i.e. the correlation) among $g(X_1), \cdots, g(X_p)$ may leads to a good estimation for the quantiles of ξ. This is due to the fact that $p^{-1}\sum_j g(X_j)$ would then behave

like a normal random variable, the fitted Gaussian copula should provide adequate estimates for its first two moments.

Acknowledgements. We thank Joint Editor Professor Piotr Fryzlewicz, an Associate Editor and two referees for their helpful comments. Jinguo Gong was partially supported by National Social Science Foundation of China (Grant No.12CTJ007). Liang Peng was partially supported by Simons Foundation. Qiwei Yao was partially supported by an EPSRC research grant.

References

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

Basel Committee (2011). Basel III: a global regulatory framework for more resilient banks and banking systems. *Bank for International Settlements*

Bedford, Y. and Cooke, R.M. (2001). Probability density decomposition for conditional dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, **32**, 245-268.

Bedford, Y. and Cooke, R.M. (2002). Vines – a new graphical model for dependent random variables. *The Annals of Statistics*, **30**, 1031-1068.

Beirlant, J., Matthys, G., Guillou, A. and Delafosse, E. (2004). Estimation of catastrophic quantile levels for heavy tailed distributions. *Insurance: Mathematics and Economics*, **34**, 517–537.

Brechmann, E.C. and Schepsmeier, U. (2013). Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine. *Journal of Statistical Software*, **52**, 1-27.

Bruun, J.T. and Tawn, J.A. (1998). Comparison of approaches for estimating the probability of coastal flooding. *Appl. Statist.*, **47**, 405-423.

Coles, S.G. (2001). An Introduction to Statistical Modeling of Extreme Values. *Springer.*

Coles, S.G. and Tawn, J.A. (1994). Statistical methods for multivariate extremes: an application to structure design. *J. Royal Statist. Soc. C*, **43**, 1-31.

Czado, C., Brechmann, E.C. and Gruber, L. (2013). Selection of vine copulas. In *Copulae in Mathematical and Quantitative Finance*. Edited by Jaworski, P., Durante, F. and W.K. Härdle. Springer Lecture Notes in Statistics **213**, pp 17-37.

De Haan, L. and Ferreira, A. (2006). Extreme Value Theory: An Introduction. *Springer.*

De Haan, L. and Sinha, A.K. (1999). Estimating the probability of a rare event. *Ann. Statist.*, **27**, 732-759.

Dekkers, A.L.M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation *The Annals of Statistics*, **17**, 1759-1832.

Drees, H. and de Haan, L. (2013). Estimating failure probabilities. *Bernoulli*, to appear.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Spriner, Berlin.

Fermanian, J.-D., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847-860.

Ferreira, A., de Haan, L. and Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*, **37**, 403-434.

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347-368.

Genest, C., Ghoudi K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Probab. Statist.*, **44**, 1096-1127.

Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insurance Mathematics and Economics*, **44**, 199-213.

Glosten, L. R., Jagannathan, R., and Runkle D. E. (1993). On The Relation between The Expected Value and The Volatility of Nominal Excess Return on stocks. *Journal of Finance*, **48**, 1779-1801.

Gomes, M.I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator. *J. Amer. Statist. Assoc.*, **102**, 280–292.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In *Distributions with Fixed Marginals and Related Topics*. Edited by Rüschendorf, L., Schweizer, B. and Taylor, M.D. IMS Lecture Notes **28**, 120-141.

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

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

Nelson, R.B. (2006). *An Introduction to Copulas*. Springer.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, **23**, 470-472.

Wang, H.J., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy tailed distribution. *Journal of the American Statistical Association*, **107**, 1453-1464.