Statistica Sinica **20** (2010), 365-378

A PRACTICAL WAY FOR ESTIMATING TAIL DEPENDENCE FUNCTIONS

Liang Peng

Georgia Institute of Technology

Abstract: Estimating tail dependence functions is important for applications of multivariate extreme value theory, and only a fraction of the upper order statistics are involved in the estimation. How to choose the sample fraction or threshold is of importance in practice. Motivated by the recent methodologies on threshold selection for a tail index in Guillou and Hall (2001) and Peng (2009a), we apply the idea in Peng (2009a) to obtain a data-driven method for choosing the threshold in estimating a tail dependence function. Further we propose a simple bias-reduction estimator, and the combination of the bias-reduction estimator with the threshold selection procedure gives a satisfactory way of estimating a tail dependence function. This is supported by a simulation study. Moreover, a sub-sample bootstrap method is proposed to construct a confidence interval for a tail dependence function.

Key words and phrases: Bivariate extremes, bootstrap, tail dependence, threshold.

1. Introduction

Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed random variables with continuous distribution function $F(x, y)$. Let $F_1(x) := F(x, \infty)$ and $F_2(y) := F(\infty, y)$ denote the marginal distribution functions. Bivariate extreme value theory is mainly based on the assumption that there exist constants $a_n > 0$, $c_n > 0$, $b_n \in R$, $d_n \in R$ such that

$$\lim_{n\to\infty} P\left(\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \le x, \frac{\bigvee_{i=1}^n Y_i - d_n}{c_n} \le y\right) = G(x, y) \qquad (1.1)$$

for all $x, y \in R$, where G is a continuous distribution function. Under the setup of (1.1), one can estimate the probability of a rare event; see De Haan and de Ronde (1998), and De Haan and Sinha (1999). An important step is to divide (1.1) into marginal conditions and dependence function, that is, (1.1) is equivalent to

$$\begin{cases} \lim_{n\to\infty} P\Big(\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \le x\Big) = G(x, \infty) \\ \lim_{n\to\infty} P\Big(\frac{\bigvee_{i=1}^n Y_i - d_n}{c_n} \le y\Big) = G(\infty, y) \end{cases} \qquad (1.2)$$

and

$$\lim_{t\to 0} t^{-1}\Big\{1 - F(F_1^-(1-tx), F_2^-(1-ty))\Big\} = l(x,y) \tag{1.3}$$

for all $x \geq 0$, $y \geq 0$, where F_i^- denotes the inverse function of F_i and

$$l(x,y) = -\log G\Big((-\log G_1)^-(x), (-\log G_2)^-(y)\Big)$$

(see De Haan and Ferreira (2006)). Hence, estimating the tail dependence function $l(x,y)$ is of importance. Note that the tail copula is defined as

$$\lim_{t\to 0} t^{-1} P(F_1(X_1) \geq 1-tx, F_2(Y_1) \geq 1-ty),$$

which equals $x + y - l(x,y)$. Thus methods for estimating the tail dependence function can be used to estimate the tail copula straightforwardly. For the applications of tail copulas to risk management, we refer to McNeil, Frey and Embrechts (2005).

There are some studies on estimation and construction of confidence intervals and bands for a tail dependence function $l(x,y)$. For example, Huang (1992), Einmahl, de Haan and Huang (1993), De Haan and Resnick (1993), and Schmidt and Stadtmüller (2006) estimated $l(x,y)$ nonparametrically; Einmahl, de Haan and Piterbarg (2001) and Einmahl, de Haan and Sinha (1997) estimated a so-called spectral measure nonparametrically, this is closely related to $l(x,y)$; Drees and Huang (1998) obtained the best convergence rate for estimating $l(x,y)$ nonparametrically; Tawn (1988) studied parametric models and estimation for $l(x,y)$; Peng and Qi (2007) studied the estimation of partial derivatives of $l(x,y)$ and constructed confidence intervals for $l(x,y)$; Peng and Qi (2008) constructed a bootstrap confidence band for $l(x,y)$; Einmahl, de Haan and Li (2006) provided a weighted approximation for $l(x,y)$; De Haan, Neves and Peng (2008) studied the maximum likelihood estimation and goodness-of-fit tests for a parametric model of a tail dependence function. All these studies focus on the case of asymptotic dependence, i.e., $l(x,y) \neq x+y$. In case of asymptotic independence, i.e., $l(x,y) = x+y$, more conditions than (1.1) are needed in order to estimate the probability of a rare event. We refer to Ledford and Tawn (1997) for parametric models and inference on the case of asymptotic independence.

Since the tail dependence function $l(x,y)$ is defined as a limit, estimation can only involve a fraction of upper order statistics, for example, the tail empirical distribution function

$$\hat{l}_n(x,y;k) = \frac{1}{k}\sum_{i=1}^{n} I\Big(X_i \geq X_{n,n-[kx]} \quad \text{or} \quad Y_i \geq Y_{n,n-[ky]}\Big), \tag{1.4}$$

where $X_{n,1} \leq \cdots \leq X_{n,n}$ denote the order statistics of $X_1,\ldots,X_n$, $Y_{n,1} \leq \cdots \leq Y_{n,n}$ denote the order statistics of $Y_1,\ldots,Y_n$, and $k = k(n) \to \infty$ and $k/n \to 0$

as $n \to \infty$. An important question is how to choose the sample fraction k. Peng (1998) proposed a sub-sample bootstrap method to choose the optimal k in terms of asymptotic mean squared error; this was applied by Einmahl, Li and Liu (2009) to some data sets.

Motivated by the recent approaches in Guillou and Hall (2001) and Peng (2009a) on selecting threshold for a tail index, we propose a similar method for selecting the sample fraction in estimating a tail dependence function, and further propose a bias-reduction estimator with k selected by the new method; see Section 2 for details. Moreover, Section 2 gives a sub-sample bootstrap method to construct confidence intervals for a tail dependence function. Although the idea presented here is designed for tail dependence functions, extension to broader settings such as bandwidth selection for nonparametric smoothing, is possible. In Section 3, we investigate the finite sample behavior of the proposed methods. Proofs are given in Section 4.

2. Methodologies

We focus on the simple estimator $\hat{l}_n(x, y; k)$ given in (1.4). Like tail index estimation, when k is small, the variance of $\hat{l}_n(x, y; k)$ is large. On the other hand, the bias of $\hat{l}_n(x, y; k)$ becomes large when k is big. Hence, the optimal choice of k is to minimize the asymptotic mean squared error of $\hat{l}_n(x, y; k)$. To obtain an expression of the asymptotic mean squared error, we assume there exists a regular variation $A(t)$ at zero with index $\rho > 0$ (notation: $A(t) \in RV_\rho^0$), i.e., $\lim_{t\to 0} A(tx)/A(t) = x^\rho$ for all $x > 0$, such that

$$\lim_{t\to 0} \frac{t^{-1}\{1 - F(F_1^-(1 - tx), F_2^-(1 - ty))\} - l(x, y)}{A(t)} = \sigma(x, y) \tag{2.1}$$

holds uniformly on $\mathcal{F} = \{(x, y) : x \geq 0, y \geq 0, x^2 + y^2 = 1\}$, where $\sigma(x, y)$ is non-constant and not a multiple of $l(x, y)$. Under (2.1) and the assumptions that

$$k \to \infty, \quad \frac{k}{n} \to 0, \quad \sqrt{k}A\Big(\frac{k}{n}\Big) \to \lambda \in (-\infty, \infty) \tag{2.2}$$

as $n \to \infty$, one can show that

$$\sqrt{k}\Big\{\hat{l}_n(x, y; k) - l(x, y)\Big\} \xrightarrow{D} \lambda\sigma(x, y) + B(x, y) \tag{2.3}$$

in $D([0, \infty)^2)$. Here

$$B(x, y) = W(x, y) - l_1(x, y)W(x, 0) - l_2(x, y)W(0, y),$$
$$l_1(x, y) = \frac{\partial}{\partial x} l(x, y), \quad l_2(x, y) = \frac{\partial}{\partial y} l(x, y)$$

and $W(x, y)$ is a Gaussian process with zero mean and covariance structure

$$\begin{aligned} E\{W(x_1, y_1)W(x_2, y_2)\} = & \, l(x_1 \wedge x_2, y_1) + l(x_1 \wedge x_2, y_2) + l(x_1, y_1 \wedge y_2) \\ & + l(x_2, y_1 \wedge y_2) - l(x_1, y_2) - l(x_2, y_1) \\ & - l(x_1 \wedge x_2, y_1 \wedge y_2). \end{aligned} \tag{2.4}$$

See the proofs in Huang (1992) or Schmidt and Stadtmüller (2006).

Based on (2.3), one could minimize the asymptotic mean squared error of $\hat{l}_n(x, y; k)$ to obtain an optimal choice of k. Since this choice depends on several unknown quantities and estimating those quantities is not easy, Peng (1998) proposed a two-step sub-sample bootstrap method, that requires a large sample size. Instead of achieving the minimal asymptotic mean squared error, Guillou and Hall (2001) proposed a different way of choosing a sample fraction. The key idea is to find a sample fraction as large as possible while keeping the rate of the optimal one. This can result in a larger asymptotic mean squared error. Moreover their procedure is designed for Hill's estimator because a special structure property of it is employed. Recently, Peng (2009a) generalized the idea in Guillou and Hall (2001) to any statistic whose minimal asymptotic second moment is of the same order as the minimal asymptotic mean squared error. Since this choice results in a larger bias, it was further proposed to use a bias-reduction estimator. Therefore, a bias-reduction estimator with the data-driven chosen k can be used for both point estimation and interval estimation for a tail index. It is known that a tail index estimator with its optimal choice of k can't be employed to obtain a confidence interval directly.

Here we propose to apply the idea in Peng (2009a) to estimating the tail dependence function $l(x, y)$. First we need to construct a statistic whose second moment has the same order as the minimal asymptotic mean squared error of $\hat{l}_n(x, y; k)$.

Under conditions (2.1) and (2.2), it follows from (2.3) that

$$\begin{aligned} & \sqrt{k}\Big\{\hat{l}_n(x, y; k) - 2\hat{l}_n\Big(\frac{x}{2}, \frac{y}{2}; k\Big)\Big\} \\ & \xrightarrow{D} \lambda\sigma(x, y) + B(x, y) - 2\lambda\sigma\Big(\frac{x}{2}, \frac{y}{2}\Big) - 2B\Big(\frac{x}{2}, \frac{y}{2}\Big) \\ & = \{1 - 2^{-\rho}\}\lambda\sigma(x, y) + B(x, y) - 2B\Big(\frac{x}{2}, \frac{y}{2}\Big), \end{aligned} \tag{2.5}$$

since $A(t) \in RV_\rho^0$ implies that

$$\sigma(ax, ay) = a^{\rho+1}\sigma(x, y) \quad \text{for any} \quad a > 0. \tag{2.6}$$

Hence, for fixed $x, y > 0$

$$\sqrt{k}\Big\{\hat{l}_n(x, y; k) - 2\hat{l}_n\Big(\frac{x}{2}, \frac{y}{2}; k\Big)\Big\} \xrightarrow{d} N\Big((1 - 2^{-\rho})\lambda\sigma(x, y), r_1(x, y)\Big), \tag{2.7}$$

where

$$\begin{aligned} r_1(x,y) &= l(x,y) + x l_1^2(x,y) + y l_2^2(x,y) \\ &\quad + l_1(x,y) l_2(x,y) \Big\{ -6l(x,y) + 4l\Big(x, \frac{y}{2}\Big) + 4l\Big(\frac{x}{2}, y\Big) \Big\} \\ &\quad + l_1(x,y) \Big\{ 2l(x,y) - 4l\Big(x, \frac{y}{2}\Big) \Big\} + l_2(x,y) \Big\{ 2l(x,y) - 4l\Big(\frac{x}{2}, y\Big) \Big\}. \end{aligned} \tag{2.8}$$

By (2.3) and (2.7), we conclude that minimal asymptotic second moment of $\hat{l}_n(x,y;k) - 2\hat{l}_n(x/2,y/2;k)$ has the same order as the minimal asymptotic mean squared error of $\hat{l}_n(x,y;k)$.

Next we have to standardize the statistic $\hat{l}_n(x,y;k) - 2\hat{l}_n(x/2,y/2;k)$ to estimate the asymptotic variance. For estimating $l_1(x,y)$ and $l_2(x,y)$, we employ the estimators, via spectral measure, in Peng and Qi (2007) that are defined as

$$\begin{cases} \hat{l}_1(x,y;k) = \int_{\arctan(y/x)}^{\pi/2} \min\{1, \tan\theta\} \hat{\Phi}(d\theta;k), \\ \hat{l}_2(x,y;k) = \int_0^{\arctan(y/x)} \min\{1, \cot\theta\} \hat{\Phi}(d\theta;k), \end{cases} \tag{2.9}$$

where

$$\hat{\Phi}(\theta;k) = \frac{1}{k} \sum_{i=1}^n I\Big(R(X_i) \vee R(Y_i) \ge n+1-k, n+1-R(Y_i) \le (n+1-R(X_i)) \tan\theta \Big).$$

Here $R(X_i)$ denotes the rank of X_i among $X_1, \ldots, X_n$, and $R(Y_i)$ denotes the rank of Y_i among $Y_1, \ldots, Y_n$. Then we can estimate $r_1(x,y)$ by $\hat{r}_1(x,y;k)$, which replaces $l(x,y)$, $l(x/2,y)$, $l(x,y/2)$, $l_1(x,y)$, and $l_2(x,y)$ in (2.8) by $\hat{l}_n(x,y;k)$, $\hat{l}_n(x/2,y;k)$, $\hat{l}_n(x,y/2;k)$, $\hat{l}_1(x,y;k)$ and $\hat{l}_2(x,y;k)$, respectively. For a given $\gamma \in (0,1)$, define z_γ by $P(|N(0,1)| \le z_\gamma) = \gamma$. Then, it follows from (2.7) that

$$P\left(\left| \sqrt{k} \frac{\hat{l}_n(x,y;k) - 2\hat{l}_n(x/2,y/2;k)}{\sqrt{\hat{r}_1(x,y;(\log n)^2)}} \right| < z_\gamma \right) \to \gamma$$

when $\lambda = 0$. Obviously, when $\lambda = \infty$, i.e., k is very large, we have

$$P\left(\left| \sqrt{k} \frac{\hat{l}_n(x,y;k) - 2\hat{l}_n(x/2,y/2;k)}{\sqrt{\hat{r}_1(x,y;(\log n)^2)}} \right| > z_\gamma \right) \to 1.$$

Then, starting with $k = n-1$ until

$$\left| \sqrt{k} \frac{\hat{l}_n(x,y;k) - 2\hat{l}_n(x/2,y/2;k)}{\sqrt{\hat{r}_1(x,y;(\log n)^2)}} \right| < z_\gamma$$

may ensure that k satisfies (2.2). So, like Peng (2009a), we choose k as

$$\hat{k}=\inf\Big\{k:\Big|\sqrt{m}\frac{\hat{l}_n(x,y;m)-2\hat{l}_n(x/2,y/2;m)}{\sqrt{\hat{r}_1(x,y;(\log n)^2)}}\Big|\geq z_\gamma \ \text{ for all } \ m\geq k$$
$$\text{and } m\in\Big[n^{2\rho_n/(1+2\rho_n)}\wedge(0.01n)+1, n^{0.99}\vee\Big(n^{2\rho_n/(1+2\rho_n)}\log n\Big)\wedge n-1\Big]\Big\},\tag{2.10}$$

where, for $\delta\in(0,1)$,

$$\rho_n=(\log 2)^{-1}\bigg|\log\bigg|\frac{\hat{l}_n(x,y;n\exp\{-(\log n)^\delta\})-2\hat{l}_n(\frac{x}{2},\frac{y}{2};n\exp\{-(\log n)^\delta\})}{\hat{l}_n(x,y;2^{-1}n\exp\{-(\log n)^\delta\})-2\hat{l}_n(\frac{x}{2},\frac{y}{2};2^{-1}n\exp\{-(\log n)^\delta\})}\bigg|\bigg|$$

is a consistent estimator of the regular variation index of function $A(t)$ defined in (2.1).

Theorem 1. *Suppose* (2.1) *holds and*

$$\sup_{(x,y)\in\mathcal{F}}\bigg|\frac{t^{-1}\{1-F(F_1^-(1-tx),\ F_2^-(1-ty))\}-l(x,y)}{A(t)}-\sigma(x,y)\bigg|=O(t^\beta)\tag{2.11}$$

for some $\beta>0$. Then for any fixed $x,y>0$

$$\frac{\hat{k}}{n^{2\rho/(1+2\rho)}}\xrightarrow{d}\hat{\tau}:=\inf\Big\{t\geq 1:|\hat{Z}(u)|\geq z_\gamma \ \text{ for all } \ u\geq t\Big\},\tag{2.12}$$

$$\sqrt{\hat{k}}\Big\{\hat{l}_n(x,y;\hat{k})-l(x,y)\Big\}\xrightarrow{d}\hat{\tau}^{\rho+1/2}\lambda_0\sigma(x,y)+\hat{\tau}^{-1/2}\Big\{W(x\hat{\tau},y\hat{\tau})$$
$$-l_1(x,y)W(x\hat{\tau},0)-l_2(x,y)W(0,y\hat{\tau})\Big\},\tag{2.13}$$

where

$$\hat{Z}(u)=u^{\rho+1/2}(1-2^{-\rho})\frac{\lambda_0\sigma(x,y)}{\sqrt{r_1(x,y)}}$$
$$+u^{-1/2}\Big\{W(xu,yu)-l_1(x,y)W(xu,0)-l_2(x,y)W(0,yu)-2W(\frac{xu}{2},\frac{yu}{2})$$
$$+2l_1(x,y)W(\frac{xu}{2},0)+2l_2(x,y)W(0,\frac{yu}{2})\Big\}(r_1(x,y))^{-1/2},\tag{2.14}$$

W is given in (2.4), *and* $\lim_{n\to\infty}n^{\rho/(1+\rho)}A(n^{1/(1+2\rho)})=\lambda_0$.

Since (2.12) implies that $P(\hat{k}/n^{2\rho/(1+2\rho)}>1)>0$, it is better to employ a bias-reduction estimator instead of $\hat{l}_n(x,y;\hat{k})$. Here we propose the simple bias-reduction estimator

$$\tilde{l}_n(x,y;\hat{k})=\hat{l}_n(x,y;\hat{k})-\Big\{\hat{l}_n(x,y;\hat{k})-2\hat{l}_n\Big(\frac{x}{2},\frac{y}{2};\hat{k}\Big)\Big\}(1-2^{-\hat{\rho}})^{-1},\tag{2.15}$$

where

$$\hat{\rho} = \frac{\log \hat{k}/2}{\log n - \log \hat{k}}. \tag{2.16}$$

We remark that our approach works for any bias-reduction estimator although we are not aware of others.

Theorem 2. *Under conditions of Theorem* 1*, we have*

$$\begin{aligned}
&\sqrt{\hat{k}}\{\tilde{l}_n(x, y; \hat{k}) - l(x, y)\} \\
&\xrightarrow{d} \hat{\tau}^{-1/2}\Big\{W(x\hat{\tau}, y\hat{\tau}) - l_1(x, y)W(x\hat{\tau}, 0) - l_2(x, y)W(0, y\hat{\tau})\Big\} \\
&\quad -\hat{\tau}^{-1/2}\Big\{W(x\hat{\tau}, y\hat{\tau}) - l_1(x, y)W(x\hat{\tau}, 0) - l_2(x, y)W(0, y\hat{\tau}) \\
&\quad -2W(\frac{x\hat{\tau}}{2}, \frac{y\hat{\tau}}{2}) - 2l_1(x, y)W(\frac{x\hat{\tau}}{2}, 0) - 2l_2(x, y)W(0, \frac{y\hat{\tau}}{2})\Big\}(1 - 2^{-\rho})^{-1}
\end{aligned} \tag{2.17}$$

for any fixed $x, y > 0$.

Remark 1. Recently, Klüppelberg, Kuhn and Peng (2008) proposed the use of elliptical copulas to model tail copulas, which results in an effective way of dealing with the dimensionality of multivariate extremes. The estimation procedure proposed there depends on the estimation of $l(1, 1)$, and the choice of threshold is taken as the same one in estimating $l(1, 1)$. Hence the estimator $\tilde{l}(1, 1; \hat{k})$ above can be applied to solve the open issue of threshold selection in Klüppelberg, Kuhn and Peng (2008). More details are given in Peng (2009b).

Like Peng (2009a), based on Theorem 2, we propose the following bootstrap method to construct a confidence interval for $l(x, y)$ with level γ_0, as follows.

Take n_1 ($= n_1(n) \to \infty$ and $n_1/n \to 0$ as $n \to \infty$) and draw a random sample of size n_1 from $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ with replacement, say (X_1^*, Y_1^*), $\ldots, (X_{n_1}^*, Y_{n_1}^*)$. Based on this bootstrap sample, we compute the corresponding $\hat{k}$ and $\tilde{l}_n(x, y; \hat{k})$, say $\hat{k}^*$ and $\tilde{l}_n^*(x, y; \hat{k}^*)$. Hence, we have the bootstrap statistic $S^* = \sqrt{\hat{k}^*}\{\tilde{l}_n^*(x, y; \hat{k}^*) - \tilde{l}_n(x, y; \hat{k})\}$. Repeating this procedure M times, we obtain M bootstrap statistics, say $S_1^*, \ldots, S_M^*$. Due to the randomness of $\hat{k}$ and the skewness of the limit in (2.17), we propose to employ $\log((S^*)^2)$ instead of S^*. Put $T_i^* = \log((S_i^*)^2)$ for $i = 1, \ldots, M$, and let $T_{M,1}^* \le \cdots \le T_{M,M}^*$ denote the order statistics of $T_1^*, \ldots, T_M^*$. Hence a confidence interval for $l(x, y)$ with level γ_0 is

$$\Big(a + \tilde{l}_n(x, y; \hat{k}), b + \tilde{l}_n(x, y; \hat{k})\Big) \cup \Big(-b + \tilde{l}_n(x, y; \hat{k}), -a + \tilde{l}_n(x, y; \hat{k})\Big),$$

where

$$a = (\hat{k})^{-1/2} \exp\left\{\frac{1}{2}T^*_{M,[(1-\gamma_0)M/2]}\right\} \text{ and } b = (\hat{k})^{-1/2} \exp\left\{\frac{1}{2}T^*_{M,[(1+\gamma_0)M/2]}\right\}.$$

3. Simulation Study

We investigate the finite sample behavior of the proposed bias-reduction estimator coupled with the proposed threshold selection method. In general, it is not easy to simulate random vectors with a given tail dependence function. Recently, Klüppelberg, Kuhn and Peng (2007) obtained explicit expression for the tail dependence function of an elliptical vector. In particular,

$$l(x,y) = x+y-\frac{x\int_{g((x/y)^{1/\alpha})}^{\pi/2}\cos^\alpha\theta d\theta + y\int_{-\arcsin q}^{g((x/y)^{1/\alpha})}\sin^\alpha(\theta+\arcsin q)d\theta}{\int_{-\pi/2}^{\pi/2}\cos^\alpha\theta d\theta}, \quad (3.1)$$

where $g(t)=\arctan((t-q)/\sqrt{1-q^2})$, when the elliptical vector RAU satisfies

$$AA^T = \begin{pmatrix} \sigma^2 & q\sigma\nu \\ [5pt] q\sigma\nu & \nu^2 \end{pmatrix},$$

$rank(AA^T) = 2$, $-1 < q < 1$, $\sigma^2 > 0$, $\nu^2 > 0$, $R > 0$ is a heavy-tailed random variable with index α (i.e., $\lim_{t\to\infty} P(R > tx)/P(R > t) = x^{-\alpha}$), $U = (U_1, U_2)^T$ is a random vector uniformly distributed on the unit sphere $\{(u_1,u_2)^T : u_1^2+u_2^2 = 1\}$, and U is independent of R. Recently, Li and Peng (2009) proposed a goodness-of-fit test for testing (3.1) and applied it to some financial data sets for which (3.1) cannot be rejected. For other applications of elliptical distributions and elliptical copulas in risk management, we refer to McNeil, Frey and Embrechts (2005).

We drew $1,000$ random samples of size $n = 100, 200, 500$ and $1,000$ from the above elliptical random vector, with

$$A = \begin{pmatrix} \frac{\sqrt{1+q}+\sqrt{1-q}}{2} & \frac{\sqrt{1+q}-\sqrt{1-q}}{2} \\ \frac{\sqrt{1+q}-\sqrt{1-q}}{2} & \frac{\sqrt{1+q}+\sqrt{1-q}}{2} \end{pmatrix}$$

and $P(R \le x) = \exp\{-x^{-\alpha}\}$. Consider $q = 0.5$, $\alpha = 0.5$ and 2, $x = \cos\theta$, $y = \sin\theta$ for $\theta = \pi/8, 2\pi/8, 3\pi/8$, and take $\gamma = 0.9$ or 0.95, and $\delta = 0.1$ in the definition of ρ_n.

For comparisons, we computed the theoretical optimal choice of k, denoted by k_{opt}, which minimizes the asymptotic mean squared error of $\hat{l}_n(x,y;k)$, from Corollary 6 of Klüppelberg, Kuhn and Peng (2007) for the above setup. We used $\hat{k}_1$ and $\hat{k}_2$ to denote the $\hat{k}$ in Theorem 1 for $\gamma = 0.9$ and 0.95, respectively. Moreover we computed the simulated optimal choice of k, denoted by k^*, which minimizes the average of $\{\hat{l}_n(x,y;k)-l(x,y)\}^2$ over those $1,000$ random samples. In Tables 1−4, we report means and root mean squared errors for $\hat{l}_n(x,y;k)$ and $\tilde{l}_n(x,y;k)$ with $k = \hat{k}_1, \hat{k}_2, k^*$. From Tables 1-4, we observe that

Table 1. Estimators, with their root mean squared errors in parentheses, are given for $(x, y) = (\cos\theta, \sin\theta)$ and $n = 100$.

	$\alpha = 0.5$ $\theta = \frac{\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{2\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{3\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{2\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{3\pi}{8}$
$l(x, y)$	1.0414	1.0090	1.0414	1.0968	1.1377	1.0968
k_{opt}	732	248	732	169	135	169
k^*	65	50	75	35	30	35
$\hat{l}_n(x, y; k^*)$	1.0430	1.0035	1.0273	1.0901	1.1343	1.0901
	(0.0308)	(0.0450)	(0.0303)	(0.0473)	(0.0632)	(0.0438)
$\tilde{l}_n(x, y; k^*)$	1.0756	1.0290	1.0558	1.1727	1.1904	1.1768
	(0.0574)	(0.0786)	(0.0507)	(0.1197)	(0.1491)	(0.1244)
$\hat{k}_1$	80	73	81	62	55	60
$\hat{l}_n(x, y; \hat{k}_1)$	1.0112	0.9460	1.0127	1.0398	1.0467	1.0399
	(0.0527)	(0.0994)	(0.0557)	(0.1061)	(0.1343)	(0.1025)
$\tilde{l}_n(x, y; \hat{k}_1)$	1.0493	1.0027	1.0500	1.0945	1.1274	1.0933
	(0.0478)	(0.0794)	(0.0509)	(0.1006)	(0.1143)	(0.0968)
$\hat{k}_2$	82	76	83	66	59	65
$\hat{l}_n(x, y; \hat{k}_2)$	1.0044	0.9330	1.0066	1.0296	1.0252	1.0247
	(0.0557)	(0.1070)	(0.0584)	(0.1121)	(0.1502)	(0.1060)
$\tilde{l}_n(x, y; \hat{k}_2)$	1.0479	1.0027	1.0496	1.0939	1.1202	1.0851
	(0.0456)	(0.0752)	(0.0501)	(0.1084)	(0.1147)	(0.0965)

Table 2. Estimators, with their root mean squared errors in parentheses, are given for $(x, y) = (\cos\theta, \sin\theta)$ and $n = 200$.

	$\alpha = 0.5$ $\theta = \frac{\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{2\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{3\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{2\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{3\pi}{8}$
$l(x, y)$	1.0414	1.0090	1.0414	1.0968	1.1377	1.0968
k_{opt}	1162	393	1162	268	215	268
k^*	130	85	140	65	60	65
$\hat{l}_n(x, y; k^*)$	1.0393	0.9994	1.0343	1.0871	1.1117	1.0805
	(0.0224)	(0.0376)	(0.0242)	(0.0380)	(0.0540)	(0.0412)
$\tilde{l}_n(x, y; k^*)$	1.0534	1.0414	1.0480	1.1487	1.1938	1.1474
	(0.0372)	(0.0675)	(0.0334)	(0.0897)	(0.1004)	(0.0808)
$\hat{k}_1$	159	144	156	101	100	103
$\hat{l}_n(x, y; \hat{k}_1)$	1.0089	0.09416	1.0164	1.0465	1.0500	1.0443
	(0.0464)	(0.0813)	(0.0457)	(0.0784)	(0.1073)	(0.0874)
$\tilde{l}_n(x, y; \hat{k}_1)$	1.0406	0.9917	1.0475	1.0955	1.1180	1.0936
	(0.0335)	(0.0539)	(0.0411)	(0.0687)	(0.0733)	(0.0751)
$\hat{k}_2$	163	150	161	109	106	114
$\hat{l}_n(x, y; \hat{k}_2)$	1.0031	0.9319	1.0344	1.0341	1.0351	1.0313
	(0.0509)	(0.0894)	(0.0506)	(0.0859)	(0.1190)	(0.0952)
$\tilde{l}_n(x, y; \hat{k}_2)$	1.0415	0.9905	1.0480	1.0907	1.1150	1.0866
	(0.0335)	(0.0528)	(0.0402)	(0.0676)	(0.0724)	(0.0753)

Table 3. Estimators, with their root mean squared errors in parentheses, are given for $(x, y) = (\cos\theta, \sin\theta)$ and $n = 500$.

	$\alpha = 0.5$ $\theta = \frac{\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{2\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{3\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{2\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{3\pi}{8}$
$l(x, y)$	1.0414	1.0090	1.0414	1.0968	1.1377	1.0968
k_{opt}	2141	725	2141	494	396	494
k^*	315	150	315	115	95	130
$\hat{l}_n(x, y; k^*)$	1.0374	1.0056	1.0383	1.0907	1.1291	1.0850
	(0.0152)	(0.0246)	(0.0141)	(0.0284)	(0.0346)	(0.0267)
$\tilde{l}_n(x, y; k^*)$	1.0445	1.0327	1.0448	1.1278	1.1579	1.1067
	(0.0225)	(0.0548)	(0.0212)	(0.0588)	(0.0786)	(0.0505)
$\hat{k}_1$	364	330	361	212	216	212
$\hat{l}_n(x, y; \hat{k}_1)$	1.0230	0.9574	1.0229	1.0552	1.0667	1.0542
	(0.0284)	(0.0571)	(0.0292)	(0.0527)	(0.0810)	0.0528)
$\tilde{l}_n(x, y; \hat{k}_1)$	1.0422	0.9917	1.0418	1.0888	1.1153	1.0890
	(0.0219)	(0.0309)	(0.0219)	(0.0378)	(0.0503)	(0.0349)
$\hat{k}_2$	375	342	370	223	227	222
$\hat{l}_n(x, y; \hat{k}_2)$	1.0189	0.9512	1.0195	1.0476	1.0576	1.0477
	(0.0309)	(0.0627)	(0.0321)	(0.0580)	(0.0890)	(0.0586)
$\tilde{l}_n(x, y; \hat{k}_2)$	1.0419	0.9914	1.0421	1.0870	1.1135	1.0882
	(0.0216)	(0.0305)	(0.0226)	(0.0367)	(0.0509)	(0.0357)

Table 4. Estimators, with their root mean squared errors in parentheses, are given for $(x, y) = (\cos\theta, \sin\theta)$ and $n = 1,000$.

	$\alpha = 0.5$ $\theta = \frac{\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{2\pi}{8}$	$\alpha = 0.5$ $\theta = \frac{3\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{2\pi}{8}$	$\alpha = 2.0$ $\theta = \frac{3\pi}{8}$
$l(x, y)$	1.0414	1.0090	1.0414	1.0968	1.1377	1.0968
k_{opt}	3399	1151	3399	785	630	785
k^*	630	310	550	220	215	155
$\hat{l}_n(x, y; k^*)$	1.0360	0.9975	1.0389	1.0837	1.1203	1.0908
	(0.0115)	(0.0231)	(0.0122)	(0.0238)	(0.0295)	(0.0229)
$\tilde{l}_n(x, y; k^*)$	1.0435	1.0065	1.0445	1.1064	1.1519	1.1032
	(0.0167)	(0.0333)	(0.0183)	(0.0379)	(0.0482)	(0.0476)
$\hat{k}_1$	718	603	710	393	372	403
$\hat{l}_n(x, y; \hat{k}_1)$	1.0243	0.9667	1.0252	1.0581	1.0857	1.0552
	(0.0230)	(0.0472)	(0.0224)	(0.0454)	(0.0595)	(0.0467)
$\tilde{l}_n(x, y; \hat{k}_1)$	1.0404	0.9931	1.0405	1.0878	1.1244	1.0838
	(0.0154)	(0.0280)	(0.0161)	(0.0290)	(0.0361)	(0.0282)
$\hat{k}_2$	735	631	732	412	392	424
$\hat{l}_n(x, y; \hat{k}_2)$	1.0214	0.9610	1.0220	1.0528	1.0777	1.0497
	(0.0256)	(0.0523)	(0.0246)	(0.0498)	(0.0665)	(0.0514)
$\tilde{l}_n(x, y; \hat{k}_2)$	1.0405	0.9920	1.0445	1.0870	1.1221	1.0823
	(0.0157)	(0.0284)	(0.0155)	(0.0289)	(0.0371)	(0.0272)

(i) the theoretical optimal choice of k was larger than sample size in most cases, which implies that estimating k_{opt} is only practical for a large sample size;

(ii) the bias-reduction estimators $\tilde{l}_n(x, y; \hat{k}_1)$ and $\tilde{l}_n(x, y; \hat{k}_2)$ worked well and outperformed $\hat{l}_n(x, y; \hat{k}_1)$ and $\hat{l}_n(x, y; \hat{k}_2)$;

(iii) as γ grew large, $\hat{k}$ grew large as well, but the root mean squared error only changed slightly;

(iv) as sample size increased, the root mean squared errors of $\tilde{l}_n(x, y; \hat{k}_1)$ and $\tilde{l}_n(x, y; \hat{k}_2)$ approached the root mean squared error of $\hat{l}_n(x, y; k^*)$; this is not obtainable in practice;

(v) both $\hat{k}_1$ and $\hat{k}_2$ were larger than k^*, which implies that our method results in a larger bias and the bias-reduction is indeed needed;

(vi) both $\tilde{l}_n(x, y; \hat{k}_1)$ and $\tilde{l}_n(x, y; \hat{k}_2)$ had a smaller bias than $\hat{l}_n(x, y; k^*)$ in most cases;

(vii) $\tilde{l}_n(x, y; k^*)$ was worse than $\tilde{l}_n(x, y; \hat{k}_1)$ and $\tilde{l}_n(x, y; \hat{k}_2)$, which indicates that a bias-reduction estimator based on the simulated optimal choice of k is not better than that based on the sample fraction chosen by the proposed data-driven method.

Next we examined the coverage accuracy of the proposed bootstrap confidence intervals. We drew $1,000$ random samples of size $n = 200$ from the above elliptical distribution with $q = 0.5$ and $\alpha = 0.5$. For each random sample, we drew 200 bootstrap samples with size $n_1 = n^{0.95}$. Consider confidence intervals for $l(\cos\theta, \sin\theta)$ with levels 0.9 and 0.95. For computing $\hat{k}$, we used $\gamma = 0.9$ and $\delta = 0.1$. The coverage probabilities for the 90% confidence intervals were $0.91, 0.885, 0.89$ for $\theta = \pi/8, 2\pi/8, 3\pi/8$, respectively; and those for the 95% confidence intervals were $0.950, 0.94, 0.943$ for $\theta = \pi/8, 2\pi/8, 3\pi/8$, respectively. Hence, this proposed bootstrap method worked well.

In summary, the proposed data-driven method for choosing k worked well for small sample sizes, and the proposed bias-reduction estimator with the k chosen by the data-driven method can be employed for both interval and point estimation of a tail copula or tail dependence function simultaneously.

4. Proofs

Proof of Theorem 1. It follows from (2.3) and (2.11) that

$$\rho_n - \rho = o_p(\log n). \tag{4.1}$$

Since $l(ax, ay) = al(x, y)$ for any $a > 0$, we have

$$l_i(ax, ay) = l_i(x, y) \quad \text{for} \quad a > 0,\ i = 1, 2. \tag{4.2}$$

By (2.3), (2.6) and (4.2), we have

$$\begin{aligned}
&\sqrt{n^{2\rho/(1+2\rho)}s}\Big\{\hat{l}_n\Big(x,y;n^{2\rho/(1+2\rho)}s\Big)-2\hat{l}_n\Big(\frac{x}{2},\frac{y}{2};n^{2\rho/(1+2\rho)}s\Big)\Big\}\\
&\overset{D}{\to} s^{-1/2}\Big\{\lambda_0\sigma(xs,ys)+B(xs,yx)-2\lambda_0\sigma\Big(\frac{xs}{2},\frac{ys}{2}\Big)-2B\Big(\frac{xs}{2},\frac{ys}{2}\Big)\Big\}\\
&= s^{\rho+1/2}(1-2^{-\rho})\lambda_0\sigma(x,y)+s^{-1/2}\Big\{W(xs,ys)-l_1(x,y)W(xs,0)\\
&\quad -l_2(x,y)W(0,ys)-2W\Big(\frac{xs}{2},\frac{ys}{2}\Big)+2l_1(x,y)W\Big(\frac{xs}{2},0\Big)\\
&\quad +2l_2(x,y)W\Big(0,\frac{ys}{2}\Big)\Big\} \qquad (4.3)
\end{aligned}$$

in $D(0,\infty)$. Hence (2.12) follows from (4.1), (4.3) and the fact that

$$\hat{r}_1\Big(x,y;(\log n)^2\Big)\overset{p}{\to} r_1(x,y). \qquad (4.4)$$

Write

$$\begin{aligned}
&\sqrt{\hat{k}}\Big\{\hat{l}_n(x,y;\hat{k})-l(x,y)\Big\}\\
&=\sqrt{\frac{\hat{k}}{n^{2\rho/(1+2\rho)}}\;n^{2\rho/(1+2\rho)}}\;\Big\{\hat{l}_n\Big(x,y;\frac{\hat{k}}{n^{2\rho/(1+2\rho)}}\;n^{2\rho/(1+2\rho)}\Big)-l(x,y)\Big\}. \qquad (4.5)
\end{aligned}$$

It follows from (2.6), (2.3) and (4.2) that

$$\begin{aligned}
&\sqrt{n^{2\rho/(1+2\rho)}s}\Big\{\hat{l}_n\Big(x,y;n^{2\rho/(1+2\rho)}s\Big)-l(x,y)\Big\}\\
&\overset{D}{\to} s^{\rho+1/2}\lambda_0\sigma(x,y)+s^{-1/2}\Big\{W(xs,ys)-l_1(x,y)W(xs,0)-l_2(x,y)W(0,ys)\Big\} \qquad (4.6)
\end{aligned}$$

in $D(0,\infty)$. Since the convergences in (4.3) and (4.6) hold in $D(0,\infty)$, and the limits are expressed in terms of the same Gaussian process W, the joint weak convergence follows immediately; see Billingsley (1999) for details on convergences in $D(0,\infty)$. Using the Lemma on Page 151 of Billingsley (1999), equation (2.13) follows from (2.12), (4.5) and (4.6).

Proof of Theorem 2. This result follows from (2.12), (4.3), (4.6) and the fact that $\hat{\rho}\overset{p}{\to}\rho$.

Acknowledgement

The author thanks an associate editor and two reviewers for their helpful comments. Research was supported by NSF grant SES-0631608.

References

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley, New York.

De Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* **1**, 7-45.

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

De Haan, L., Neves, C. and Peng, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99**, 1260-1275.

De Haan, L. and Resnick, S.I. (1993). Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models* **9**, 275-309.

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

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimates of the stable tail dependence functions. *J. Multivariate Anal.* **64**, 25-47.

Einmahl, J. H. J., de Haan, L. and Huang, X. (1993). Estimating a multidimensional extreme value distribution. *J. Multivariate Anal.* **47**, 35-47.

Einmahl, J. H. J., de Haan, L. and Li, D. (2006). Weighted approximations of tail copula processes with application to testing the multivariate extreme value distribution. *Ann. Statist.* **34**, 1987-2014.

Einmahl, J. H. J., de Haan, L. and Piterbarg, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29**, 1401-1423.

Einmahl, J. H. J., de Haan, L. and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70**, 143 - 171.

Einmahl, J. H. J., Li, J. and Liu, R. Y. (2009). Multivariate thresholds for extremes and simultaneous monitoring of multiple aviation risk indicators. *J. Amer. Statist. Assoc.*, to appear.

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *J. Roy. Statist. Soc. Ser. B* **63**, 293-305.

Huang, X. (1992). *Statistics of Bivariate Extreme Values.* Tinbergen Institute Research Series 22.

Klüppelberg, C., Kuhn, G. and Peng, L. (2007). Estimating the tail dependence of an elliptical distribution. *Bernoulli* **13**, 229-251.

Klüppelberg, C., Kuhn, G. and Peng, L. (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent case. *Scand. J. Statist.* **35**, 701-718.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59**, 475-499.

Li, D. and Peng, L. (2009). Goodness-of-fit test for tail copulas modeled by elliptical copulas. *Statist. Probab. Lett.* **79**, 1097-1104.

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

Peng, L. (1998). *Second Order Condition And Extreme Value Theory.* Tinbergen Institute Research Series 178.

Peng, L. (2009a). A practical method for analyzing heavy tailed data. *Canad. J. Statist.*, to appear.

Peng, L. (2009b). Estimating the probability of a rare event via elliptical copulas. *The North American Actuarial Journal*, to appear.

Peng, L. and Qi, Y. (2007). Partial derivatives and confidence intervals for bivariate tail dependence functions. *J. Statist. Plann. Inference* **137**, 2089-2101.

Peng, L. and Qi, Y. (2008). Bootstrap approximation of tail dependence function. *J. Multivariate Anal.* **99**, 1807-1824.

Schmidt, R. and Stadtmüller, U. (2006). Nonparametric estimation of tail dependence. *Scand. J. Statist.* **33**, 307-335.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397-415.

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, U.S.A.

E-mail: peng@math.gatech.edu

(Received November 2007; accepted November 2008)