# A Method of Moments Estimator of Tail Dependence in Meta-Elliptical Models

Andrea Krajina*

*Institute for Mathematical Stochastics, Göttingen University, Göttingen, Germany*

**Abstract**

A meta-elliptical model is a distribution function whose copula is that of an elliptical distribution. The tail dependence function in such a bivariate model has a parametric representation with two parameters: a tail parameter and a correlation parameter. The correlation parameter can be estimated by robust methods based on the whole sample. Using the estimated correlation parameter as plug-in estimator, we then estimate the tail parameter applying a modification of the method of moments approach proposed in the paper by J.H.J. Einmahl, A. Krajina and J. Segers [Bernoulli 14(4), 2008, 1003-1026]. We show that such an estimator is consistent and asymptotically normal. Further, we derive the joint limit distribution of the estimators of the two parameters. We illustrate the small sample behavior of the estimator of the tail parameter by a simulation study and on real data, and we compare its performance to that of the competitive estimators.

*Keywords:* asymptotic normality, elliptical copula, elliptical distribution, meta-elliptical model, method of moments, semi-parametric model, tail dependence

*JEL codes:* C13, C14, C16

## 1. Introduction

The bivariate elliptical distributions, see for example [13, 4], are frequently used in various areas of statistical application, mainly in different branches of financial mathematics, such as risk management, see [10, 23, 18]. They are a natural extension of Gaussian and t-distributions, and a family wide enough to capture many traits of real-life problems. A number of recent papers have studied the tail behavior of bivariate elliptical distributions, see [1, 2, 14, 7]. An estimator of the tail dependence function of elliptical distributions was suggested in [21]. To model the tail dependence, a wider class of meta-elliptical models can be considered instead of the elliptical distributions, since the (tail) dependence structure does not depend on the marginal distributions. The distribution function of a meta-elliptical model is a distribution function which has the copula of an elliptical distribution, see [12]. The tail dependence of such models was estimated in [22].

Let $(X, Y)$ be a random vector with continuous distribution function $F$ and marginals $F_1, F_2$. To study the upper tail dependence structure, the *tail dependence function* of $(X, Y)$ is defined as

$$R(x, y) = \lim_{t \downarrow 0} t^{-1} \mathbb{P}\left(1 - F_1(X) \leq tx, 1 - F_2(Y) \leq ty\right),$$

where $x \geq 0$ and $y \geq 0$, see for example [3, 5, 11, 16]. The function $R$ is concave; $1 \leq R(x, y) \leq \min\{x, y\}$, for all $x \geq 0$ and $y \geq 0$; and $R$ is homogeneous of order one: $R(tx, ty) = tR(x, y)$, for all $x \geq 0$, $y \geq 0$ and $t \geq 0$. The (upper) tail dependence coefficient, $R(1, 1)$, is often used as a simple measure of tail dependence.

*Tel. +49 551 39 13 520.

*Email address:* `Andrea.Krajina@mathematik.uni-goettingen.de` (Andrea Krajina)

For a meta-elliptical model the tail dependence function depends only on the distribution function through its copula. Since the copula of an elliptical distribution, and hence the tail dependence function of a meta-elliptical model too, belongs to a two-parameter family, the estimation of the tail dependence function reduces to the estimation of the two copula parameters: the correlation parameter and the tail parameter.

The correlation can be estimated using the whole sample, from the rank correlations, which are independent of the precise model. In [22], the tail parameter was estimated by matching the empirical tail dependence function and the theoretical one, after plugging in the estimated correlation. Using the estimated correlation coefficient as plug-in estimator, we adapt and apply the method of moments procedure from [9] to estimate the tail parameter. The method provides a computationally straightforward estimator which is obtained as a solution of a single equation. The estimator is consistent and asymptotically normal. An interesting result that does not appear in the similar literature, namely the joint limit distribution of the tail parameter and the correlation parameter, is derived.

A simulation study shows that the small sample behavior of the estimator of the tail parameter is comparable to and competitive with the small sample behavior of the estimator derived in [22]. Further, we address the estimation in both fully elliptical and meta-elliptical models. When the model is elliptical with regularly varying generating variable $G$ (see next section for details), the tail parameter is equal to the reciprocal value of the extreme value index of $G$. In this case the usual estimators of the extreme value index, see for example, [15, 6], can be used, and they indeed lead to better estimators than the introduced estimator. However, if the model is meta-elliptical, and not elliptical, the transformation of the marginals results in large bias in the estimators of the extreme value index, and the proposed method works much better. Using an insurance data set for which in [25] the assumption of the meta-elliptical model for the tail dependence was not rejected, we illustrate the use of the method on real data.

The paper is organized as follows. In Section 2 we state and describe the model. We formulate the problem and present the estimation method in Section 3. The main results are given in Section 4. In Section 5 the performance of the estimator is illustrated on simulated and real data. All proofs are deferred to Section 6.

## 2. Tail dependence in meta-elliptical models

Let $(Z_1, Z_2)$ be an *elliptically distributed* random vector, that is, it satisfies the distributional equality

$$(Z_1, Z_2) \stackrel{d}{=} G\boldsymbol{A}\boldsymbol{U},$$

where $G > 0$ is the generating random variable, $\boldsymbol{A}$ is a $2 \times 2$ matrix such that $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top$ is of full rank, and $\boldsymbol{U}$ is a two-dimensional random vector independent of $G$ and uniformly distributed on the unit circle $\{(z_1, z_2) \in \mathbb{R}^2 : z_1^2 + z_2^2 = 1\}$. In this case, the matrix $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$. The parameter $\rho$ is called the correlation coefficient and coincides with the usual correlation, if second moments exist.

A distribution function $F$ follows a *meta-elliptical model* if the copula of $F$ is the same as the copula of some elliptical distribution with generating random variable $G$ and correlation coefficient $\rho$. This model is also known as the *elliptical copula model*, as used in [22] and [25]. If $G$ is regularly varying with index $\nu > 0$ and if $|\rho| < 1$, the expression (2.1) for the tail dependence function $R$ was derived in [21]. (Recall that a random variable $G$ is regularly varying with index $\nu > 0$ if $\mathbb{P}(G > x) = x^{-\nu}L(x)$, and $L$ is a slowly varying function.) Setting $f(x, y; \rho, \nu) = \arctan(((x/y)^{1/\nu} - \rho)/\sqrt{1-\rho^2}) \in [-\arcsin\rho, \pi/2]$ for $x, y > 0$, that expression reads

$$R(x, y; \rho, \nu) = \frac{x\int_{f(x,y;\rho,\nu)}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi + y\int_{-\arcsin\rho}^{f(x,y;\rho,\nu)}(\sin(\phi + \arcsin\rho))^\nu \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi}, \tag{2.1}$$

and equivalently,

$$R(x,y;\rho,\nu) = \frac{x\int_{f(x,y;\rho,\nu)}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi + y\int_{f(y,x;\rho,\nu)}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi} \tag{2.2}$$

$$= \frac{\int_{-\arcsin\rho}^{\pi/2}\min\left\{x(\cos\phi)^\nu, y(\sin(\phi+\arcsin\rho))^\nu\right\}\mathrm{d}\phi}{\int_{-\pi/2}^{\pi/2}(\cos\phi)^\nu \mathrm{d}\phi}. \tag{2.3}$$

The expression in (2.2) was derived in [22]. The one in (2.3) is easily obtained from the above formulas.

An expression for Pickands dependence function $A(x) := 1 - R(1-x,x)$ of the bivariate $t$-distribution was derived in [7],

$$A(x) = xF_{t(\nu+1)}\left(\frac{\left(\frac{x}{1-x}\right)^{\frac{1}{\nu}} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right) + (1-x)F_{t(\nu+1)}\left(\frac{\left(\frac{1-x}{x}\right)^{\frac{1}{\nu}} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right),$$

where $F_{t(\nu+1)}$ is the distribution function of $t$-distributed random variable with $\nu+1$ degrees of freedom. It was shown in [2] that Pickands dependence function of an elliptical distribution for which the generating variable $G$ is regularly varying with index $\nu > 0$ is the same. Despite the different appearance, expressions (2.1)-(2.3) lead to the same Pickands dependence function.

## 3. Estimation

Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from a continuous distribution function $F$ with marginals $F_1$ and $F_2$. Assume that $F$ follows a meta-elliptical model with underlying generating variable $G > 0$ and correlation coefficient $\rho$. To estimate the tail dependence function $R$, we will estimate the unknown parameters, namely the correlation coefficient $\rho$ and the tail parameter $\nu$, under the assumptions that $|\rho| < 1$ and that $G$ is regularly varying with index $\nu > 0$.

The above assumption corresponds to asymptotic dependence. If $\rho = 1$ or $\rho = -1$, we get complete dependence, $R(1-x,x) = \min\{1-x,x\}$, for any $\nu$. In case of $-1 < \rho < 1$ and $\nu \downarrow 0$ we have a mixture between complete dependence and independence, $R(1-x,x) = \pi^{-1}(\pi/2 + \arcsin\rho)\min\{1-x,x\}$. If $\nu \uparrow \infty$, then for any $\rho$ we are in the case of asymptotic independence, since then $R(1-x,x) \downarrow 0$.

The estimation consists of two steps. We first estimate the correlation coefficient $\rho$ using Kendall's $\tau$, see [19, 20], and the relation $\tau = (2/\pi)\arcsin\rho$ obtained in [26], see also Theorem 4.2 in [17]. Then, using expression (2.1) with the consistent estimator $\hat\rho$ from the previous step plugged in for the true correlation coefficient $\rho$, we apply the method of moments estimation procedure introduced in [9] to estimate $\nu$. A similar approach appears in [22], where the tail parameter is estimated using the pointwise inverse of $R(x,y;\rho,\nu)$ with respect to $\nu$, after the correlation coefficient $\rho$ in $R$ has been replaced by the same consistent estimator as above.

### *3.1. Estimation of the correlation parameter*

Kendall's $\tau$ of two random variables $X$ and $Y$ is defined by

$$\tau = \mathbb{P}\left((X-X')(Y-Y') > 0\right) - \mathbb{P}\left((X-X')(Y-Y') < 0\right),$$

where $(X',Y')$ is independent of and identically distributed as $(X,Y)$. To estimate $\tau$, we will use the classical estimator

$$\hat\tau = \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\operatorname{sign}\left((X_i-X_j)(Y_i-Y_j)\right),$$

and define the estimator of $\rho$ by

$$\hat\rho := \sin\left(\frac{\pi}{2}\hat\tau\right).$$

This is a consistent and asymptotically normal estimator of $\rho$, with rate of convergence $1/\sqrt{n}$, which follows from the corresponding properties of $\hat{\tau}$, see for instance [24].

*3.2. Estimation of the tail parameter*

Denote by $R_i^X$ and $R_i^Y$ the rank of $X_i$ among $X_1, \ldots, X_n$ and the rank of $Y_i$ among $Y_1, \ldots, Y_n$, respectively. Then for $1 \leq k \leq n$,

$$\hat{R}_n(x, y) := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1} \left\{ R_i^X > n + \frac{1}{2} - kx,\ R_i^Y > n + \frac{1}{2} - ky \right\}$$

is a nonparametric estimator of $R$. When studying the asymptotic properties of this estimator, $k = k_n$ is an intermediate sequence, that is, $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

Denote the parameter space by $\bar{\Theta} := \bar{\Theta}_\rho \times \bar{\Theta}_\nu$, with $\bar{\Theta}_\rho = (-1, 1)$ and $\bar{\Theta}_\nu = (0, \infty)$. Its elements are pairs $\bar{\theta} := (\rho, \nu)$. The tail dependence function of a meta-elliptical model belongs to a parametric family $\{R(\cdot, \cdot; \bar{\theta}) \colon \bar{\theta} \in \bar{\Theta}\}$. Given the correlation parameter $\rho$, it reduces to a single-parameter family $\{R(\cdot, \cdot; \rho, \nu) \colon \nu \in \bar{\Theta}_\nu\}$. We use the approach from [9] to estimate $\nu$: for a given $\rho$ and an integrable function $g \colon [0, 1]^2 \to \mathbb{R}$, the method of moments estimator of $\nu$ is defined as the solution to

$$\iint_{[0,1]^2} g(x, y) \hat{R}_n(x, y) \mathrm{d}x\mathrm{d}y = \iint_{[0,1]^2} g(x, y) R(x, y; \rho, \hat{\nu}_n) \mathrm{d}x\mathrm{d}y. \tag{3.1}$$

We can simplify the above equation by an appropriate choice of the function $g$. Choosing $g(x, y) = \mathbf{1}\{x + y \leq 1\}$, $(x, y) \in [0, 1]^2$, reduces the area of integration from the unit square to the triangle $\{(x, y) \in [0, 1]^2 \colon x + y \leq 1\}$. Due to homogeneity of $R$, see for instance [3, 5], we get that

$$\iint_{[0,1]^2} \mathbf{1}\{x + y \leq 1\} R(x,y; \rho, \nu) \mathrm{d}x\mathrm{d}y = \frac{1}{3} \int_{[0,1]} R(1 - x, x; \rho, \nu) \mathrm{d}x.$$

Instead of solving the equation (3.1), for a given $\rho$ we define the estimator of $\nu$ as the solution to

$$\int_{[0,1]} \hat{R}_n(1 - x, x) \mathrm{d}x = \int_{[0,1]} R(1 - x, x; \rho, \hat{\nu}) \mathrm{d}x.$$

That is, for a given $\rho \in \bar{\Theta}_\rho$, we define the estimator of $\nu$ as the inverse of the function $\bar{\varphi}_\rho \colon \bar{\Theta}_\nu \to \mathbb{R}$, defined by

$$\bar{\varphi}_\rho(\nu) := \int_{[0,1]} R(1 - x, x; \rho, \nu) \mathrm{d}x,$$

in the point $\int_{[0,1]} \hat{R}_n(1 - x, x) \mathrm{d}x$. However, if $\rho > 0$ this is not possible for all $\nu$, since for positive $\rho$ the function $\bar{\varphi}_\rho$ is not invertible on its whole domain $(0, \infty)$. For each $\rho > 0$ there exists a point $\nu^* = \nu^*(\rho)$, such that the function $\nu \mapsto \bar{\varphi}_\rho(\nu)$ is increasing on $(0, \nu^*(\rho))$ and decreasing on $(\nu^*(\rho), \infty)$, see Figure 1.

We will restrict the parameter space so to avoid the fact that $\nu^*$ changes with $\rho$, while retaining as much flexibility as possible. We choose some $\rho^* < 1$, numerically approximate the value of $\nu^*(\rho^*)$, and restrict $\bar{\Theta} = (-1, 1) \times (0, \infty)$ to $(-1, \rho^*) \times (\nu^*, \infty) =: \Theta_\rho \times \Theta_\nu =: \Theta$. For example, if $\rho^* = 0.9$, we can take $\nu^* = 0.66$, what leads to a parameter space that is appropriate for applications. In practice this should not present a difficulty. Usually it holds that $\nu \geq 1$, in which case the estimator is well-defined whenever $\rho < 0.97$.

For every $\rho \in \Theta_\rho$ denote by $\varphi_\rho$ the restriction of $\bar{\varphi}_\rho$ to $\Theta_\nu$, that is, for every $\rho \in \Theta_\rho$,

$$\varphi_\rho(\nu) := \int_{[0,1]} R(1 - x, x; \rho, \nu) \mathrm{d}x\,, \quad \nu \in \Theta_\nu.$$

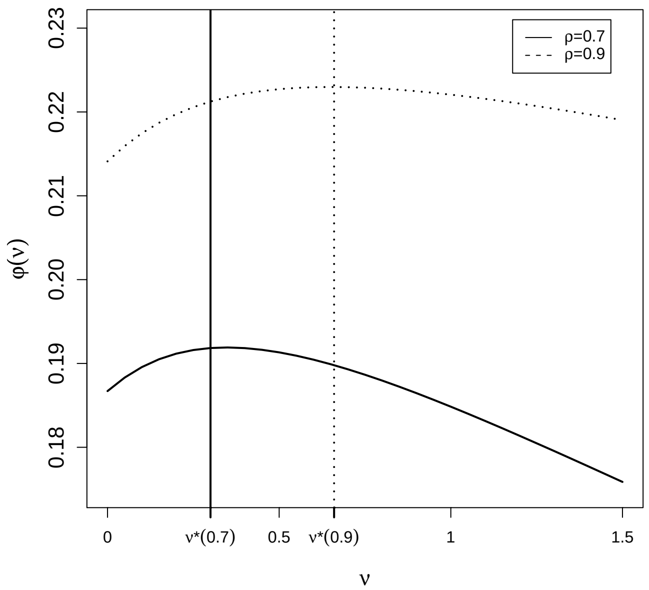



Figure 1: Graph of $\nu \mapsto \bar{\varphi}_\rho(\nu)$ for $\rho \in \{0.7, 0.9\}$. The function is decreasing on the interval $(\nu^*(\rho), \infty)$, as indicated by the vertical lines.

Finally, for $\hat{\rho} \in \Theta_\rho$, we define $\hat{\nu}_n$, the moment estimator of the tail parameter $\nu$, as the solution to

$$\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x = \int_{[0,1]} R(1-x, x; \hat{\rho}, \hat{\nu})\mathrm{d}x,$$

that is,

$$\hat{\nu}_n := \varphi_{\hat{\rho}}^{-1}\left(\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x\right). \tag{3.2}$$

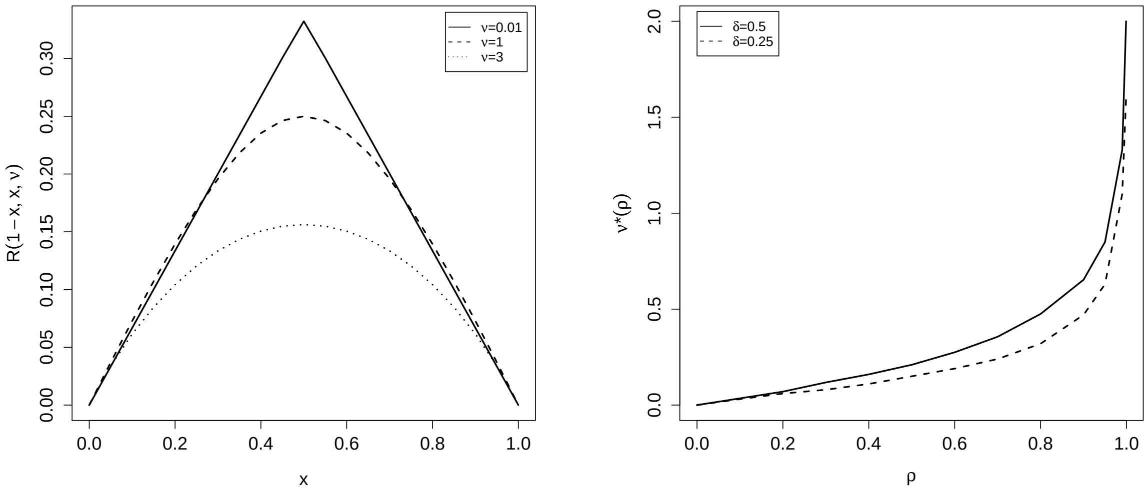
The estimator is well-defined with probability tending to one, as a consequence of the consistency of $\hat{\rho}$ and the uniform consistency of $\hat{R}_n(1-x, x)$. If $\hat{\rho} \notin \Theta_\rho$ or if $\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x \notin \Theta_\nu$, let $\hat{\nu}_n$ be some fixed value in $\Theta_\nu$.

**Remark 3.1.** (i) In the central part of the interval $[0,1]$ the functions $x \mapsto R(1-x, x; \rho, \nu)$, $\rho > 0$, behave in a favorable way, that is, they are decreasing in $\nu$, see Figure 2(a). To keep the parameter space as large as possible, we could restrict the area of integration from $[0,1]$ to $[1/2-\delta, 1/2+\delta]$, for some $\delta \in (0, 1/2]$, see Figure 2(b). However, this may result in a less efficient estimator.

(ii) Note that the set $\Theta = \Theta_\rho \times \Theta_\nu$ is not unique. For any fixed $\rho^*$ we can take $\Theta = (-1, \rho^*) \times (\nu, \infty)$, with $\nu \geq \nu^*(\rho^*)$, see Figure 2(b), the solid line. Also, one could fix $\nu^* > 0$ in advance, and appropriately restrict $\rho$ to the interval $(-1, \rho^*(\nu^*))$.

## 4. Main results

Let $\hat{\rho}$ and $\hat{\nu}_n$ be as in Section 3 and let $\rho_0 \in \Theta_\rho$ and $\nu_0 \in \Theta_\nu$ be the true values of the correlation coefficient and the tail index, respectively. The basic assumption is that




(a) Graphs of $x \mapsto R(1-x, x; \rho, \nu)$ for $\rho = 0.5$ and $\nu \in \{0.01, 1, 3\}$.

(b) Graphs of $\rho \mapsto \nu^*(\rho)$ for $\delta \in \{0.25, 0.5\}$, where $\delta = 0.5$ corresponds to integration over $[0, 1]$.

Figure 2

(C0) $\Theta = \Theta_\rho \times \Theta_\nu$ is such that $\varphi_\rho$ is a homeomorphism between $\Theta_\nu$ and its image, for every $\rho \in \Theta_\rho$.

For some of the results, we will need the following conditions:

(C1) there exists an $\alpha > 0$ such that as $t \to 0$,

$$t^{-1}\mathbb{P}\left(1 - F_1(X_1) \le tx, 1 - F_2(Y_1) \le ty\right) - R(x, y) = O(t^\alpha),$$

uniformly on $\{(x, y) \in (0, \infty)^2 : x + y = 1\}$;

(C2) $k = k_n \to \infty$ and $k = o(n^{2\alpha/(1+2\alpha)})$ as $n \to \infty$, with $\alpha$ from (C1).

**Proposition 4.1.** *Assume a meta-elliptical model in $\mathbb{R}^2$ with $(\rho_0, \nu_0) \in \Theta$. If (C0) holds, then the function $H\colon \Theta \to \Theta_\rho \times \mathbb{R}$ defined by*

$$H(\rho, \nu) := \left(\rho, \int_{[0,1]} R(1 - x, x; \rho, \nu)\, dx\right), \tag{4.1}$$

*is continuously differentiable at $(\rho_0, \nu_0)$ and its differential in this point is regular.*

An application of the inverse mapping theorem yields the following consequence of Proposition 4.1. Let $D_f(x)$ denote the differential of $f$ in $x$.

**Corollary 4.2.** *If the assumptions of Proposition 4.1 are satisfied, then there exist open neighborhoods $U \subseteq \Theta$ of $(\rho_0, \nu_0)$ and $V \subseteq H(\Theta)$ of $H(\rho_0, \nu_0)$ such that the restriction $H|_U\colon U \to V$ is one-to-one. Moreover, its inverse*

$$K := (H|_U)^{-1}\colon V \to U \tag{4.2}$$

*is continuously differentiable and for the differential of $K$ in $H(\rho_0, \nu_0)$ we have*

$$D_K\left(H\left(\rho_0, \nu_0\right)\right) = \left(D_H(\rho_0, \nu_0)\right)^{-1}.$$

Next we present the consistency and asymptotic normality results for $\hat{\nu}_n$ and $(\hat{\nu}_n, \hat{\rho})$, respectively.

**Theorem 4.3 (Consistency of $\hat{\nu}_n$).** *If the assumptions of Proposition 4.1 are satisfied, then*

$$\hat{\nu}_n \xrightarrow{\mathbb{P}} \nu_0, \quad \text{as } n \to \infty,\ k \to \infty,\ k/n \to 0.$$

Denote by $W$ a mean-zero Wiener process on $[0,\infty)^2$ with covariance function

$$\mathbb{E}W(x_1, y_1)W(x_2, y_2) = R(x_1 \wedge x_2, y_1 \wedge y_2; \rho_0, \nu_0), \tag{4.3}$$

and for $x, y \in [0, \infty)$ denote

$$W_1(x) := W(x, \infty), \qquad W_2(y) := W(\infty, y). \tag{4.4}$$

Further, for $(x, y) \in [0,\infty)^2$ let $\dot{R}_1(x, y)$ and $\dot{R}_2(x, y)$ be the partial derivatives of $R$ in the point $(x, y)$ with respect to the first and second coordinates, respectively.

Finally, define the stochastic process $B$ on $[0,\infty)^2$ by

$$B(x, y) := W(x, y) - \dot{R}_1(x, y)W_1(x) - \dot{R}_2(x, y)W_2(y). \tag{4.5}$$

Let $N_\rho \sim \mathcal{N}(0, \sigma_\rho^2)$ be the normal limiting random variable of $\sqrt{n}(\hat{\rho} - \rho_0)$ and denote by $N_\nu \sim \mathcal{N}(0, \sigma_\nu^2)$ the normal random variable $N_\nu := c \int_{[0,1]} B(1 - x, x)\mathrm{d}x$, where

$$c := \left( \partial/\partial\nu \int_{[0,1]} R(1 - x, x; \rho_0, \nu)\mathrm{d}x \Big|_{\nu=\nu_0} \right)^{-1}. \tag{4.6}$$

**Theorem 4.4 (Asymptotic normality of $(\hat{\nu}_n, \hat{\rho})$).** *Let $k/n \to 0$. If in addition to the conditions of Proposition 4.1, the conditions (C1) and (C2) hold, then as $n \to \infty$ and $k \to \infty$,*

$$\left( \sqrt{k}(\hat{\nu}_n - \nu_0), \sqrt{n}(\hat{\rho} - \rho_0) \right) \xrightarrow{d} (N_\nu, N_\rho),$$

*where $N_\nu$ and $N_\rho$ are independent.*

**Remark 4.5.** The above results are not tied to the Kendall's tau based estimator of $\rho$. The consistency and the asymptotic normality of $\hat{\nu}_n$ hold whenever the rate of convergence of the estimator of $\rho$ is faster than $1/\sqrt{k}$.

## 5. Simulation study

We simulated 50 random samples of size $n = 1000$ from two meta-elliptical models with correlation coefficient $\rho_0 = 0.3$ and tail parameter $\nu_0 \in \{1, 5\}$.

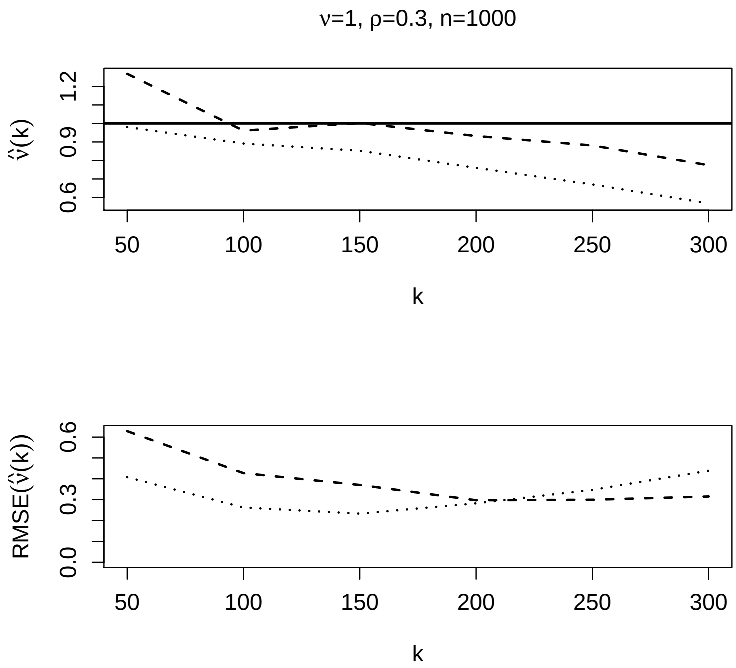
The two estimators that we compare are the MoME, the method of moments estimator $\hat{\nu}_n$ defined in (3.2), and the KKP estimator of tail parameter from [22] with the weight function $m(\psi) = 1 - (4\psi/\pi - 1)^2$, $0 \le \psi \le \pi/2$. The KKP estimator is defined by

$$\hat{\nu}_{KKP} := \frac{1}{M(\hat{Q} \cap \hat{Q}^*)} \int_{\hat{Q} \cap \hat{Q}^*} \tilde{\nu} \left( \sqrt{2}\cos\psi, \sqrt{2}\sin\psi \right) M(\mathrm{d}\psi),$$

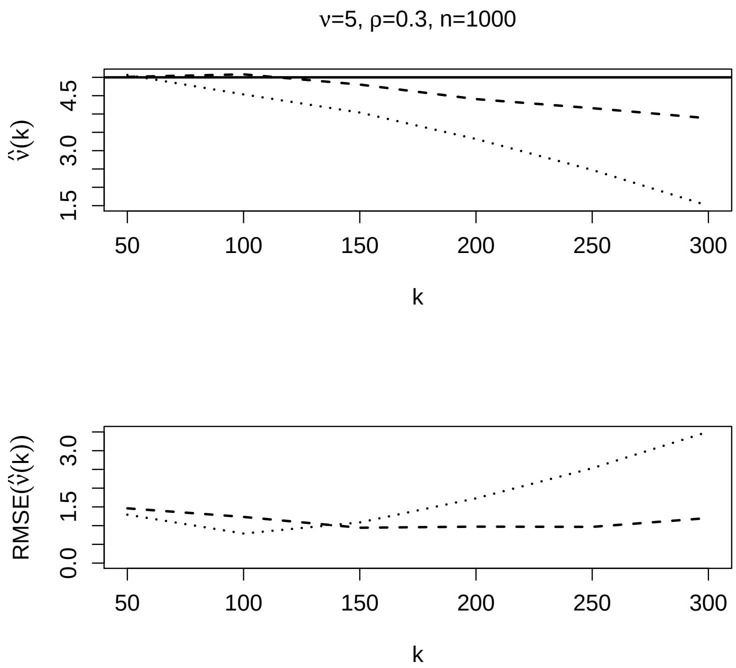
where $M$ is the measure defined by $m$, $\tilde{\nu}(x, y)$ is the inverse of $R(x, y; \hat{\rho}, \nu)$ with respect to $\nu$ in the point $\hat{R}_n(x, y)$, for $x > 0$, $y > 0$, and the sets $\hat{Q}$ and $\hat{Q}^*$ are the subsets of $[0, \pi/2]$ defined in such a way so that $\hat{\nu}_{KKP}$ is well-defined and that it has desired asymptotic properties, see [22] for details.

In Figure 3 we plot for those two estimators the bias and the root mean squared error (RMSE) against the effective sample size $k$.

The plots show that the MoME has much smaller bias than the KKP estimator. Further, it appears to be more robust with respect to the choice of $k$, and better than the KKP estimator for larger values of $k$. Also, the value of $k$ after which the MoME performs better gets smaller as the tail parameter that is estimated gets larger.




(a) Elliptical copula model with $\rho_0 = 0.3$ and $\nu_0 = 1$.




(b) Elliptical copula model with $\rho_0 = 0.3$ and $\nu_0 = 5$.

Figure 3: The bias and the RMSE of two different estimators of tail coefficient $\nu$; MoME (– – – –), KKP (······).

The elliptical distributions form a sub-class of meta-elliptical distributions, hence, the same estimation procedure applies. Assume we have a sample of $n$ independent and identically distributed elliptical random vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$. Under the assumption that their generating random variable $G$ is regularly varying with parameter $\nu > 0$, and since it holds that $G$ and $Z := \sqrt{X_1^2 + Y_1^2}$ have the same distribution, the tail parameter $\nu$ is equal to the reciprocal value of the extreme value index of the random variable $Z$. There are many estimators for the extreme value index, see for example [15, 6, 29]. We use the estimator from [15] in the following simulation study. We simulate 100 samples of size $n = 1500$ from elliptical distribution with $\rho_0 = 0.3$ and with $\nu \in \{1, 5\}$. For both values of the tail parameter we compute the bias and the root mean squared errors for the MoME and for the reciprocal value of the Hill estimator of the extreme value index of $Z$. We repeat the procedure for the meta-elliptical case: we apply a monotone transformation to the marginals, keeping the same copula. The results are intuitively reasonable: when the model is fully elliptical, the moment estimator of the extreme value index performs better, but in the meta-elliptical case, the MoME is better than the estimator of the extreme value index, due to the large bias of the latter, see Figure 4.

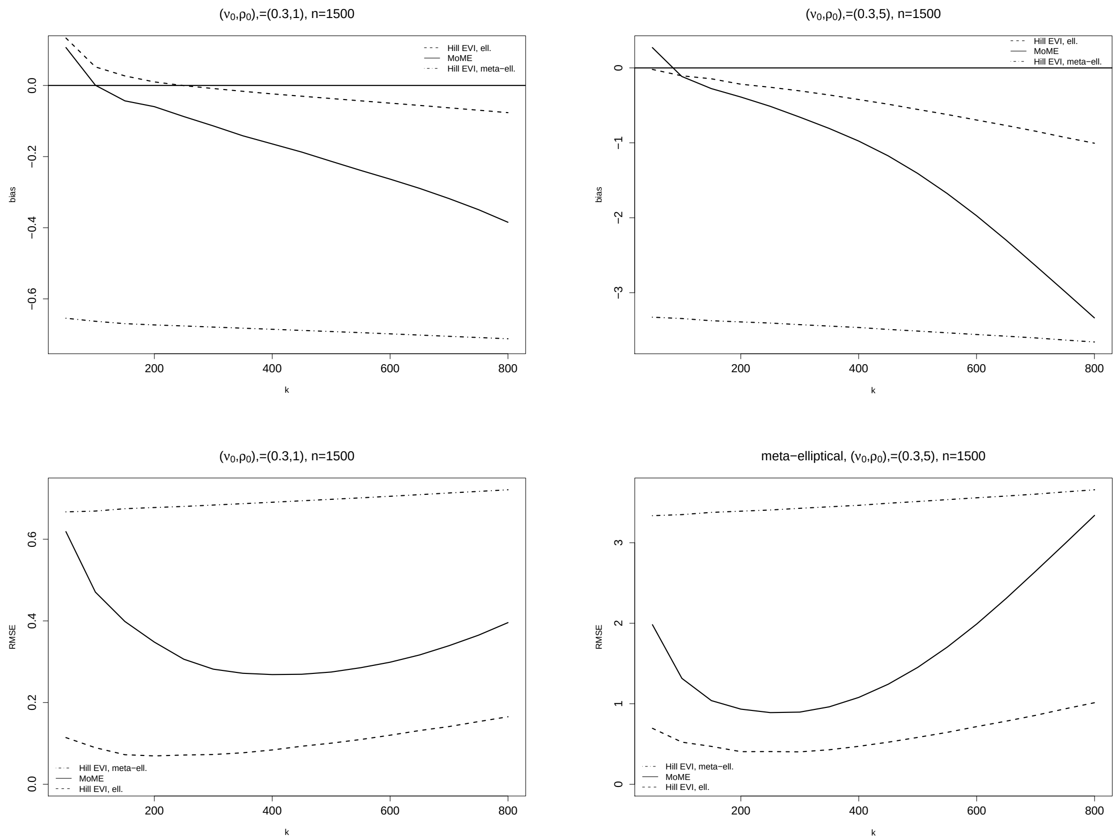



Figure 4: Elliptical vs. meta-elliptical model with the MoME and the Hill estimator; $\nu = 1$ (left panels) and $\nu = 5$ (right panels).

The Loss-ALAE data set consists of 1500 pairs of insurance claims (Loss) and the expenses associated with those claims, e.g. lawyers fees, (ALAE). It was tested whether this data set comes from a distribution with meta-elliptical tail dependence, see [25]. Since such an assumption was not rejected, we assume that the meta-elliptical model is appropriate for this data set. Using the introduced method, we obtain the estimates $(\hat{\rho}, \hat{\nu}_n) = (0.495, 1.69)$. The $k$ we choose here is 320. This choice is empirical, and based on the knowledge gained from the conducted simulations. The choice of an "optimal" $k$ is a difficult problem even in a simpler setting, and falls beyond the scope of this paper. The Hill estimator of $1/\nu$ is, in this case, of the same order as the MoME for most values of $k$, in particular, it is 1.79 for $k = 320$.

## 6. Proofs

*Proof of Proposition 4.1.* To show that the function $H$ is continuously differentiable we will show that its partial derivatives exist and are continuous on $\Theta$. Since $H(\rho,\nu) = (H_1(\rho,\nu), H_2(\rho,\nu))$, where $H_i\colon \Theta \to \mathbb{R}$, $i = 1, 2$, are given by

$$\begin{aligned} H_1(\rho,\nu) &= \rho, \\ H_2(\rho,\nu) &= \int_{[0,1]} R(1-x, x; \rho, \nu)\mathrm{d}x, \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial H_1}{\partial \rho}(\rho,\nu) &= 1, \quad \frac{\partial H_1}{\partial \nu}(\rho,\nu) = 0, \\ \frac{\partial H_2}{\partial \rho}(\rho,\nu) &= c_0^{-1}(1-\rho^2)^{\nu/2} \int_{[0,1]} \frac{x(1-x)}{\left(x^{2/\nu} + (1-x)^{2/\nu} - 2\rho x^{1/\nu}(1-x)^{1/\nu}\right)^{\nu/2}} \mathrm{d}x, \\ \frac{\partial H_2}{\partial \nu}(\rho,\nu) &= c_0^{-2} \int_{[0,1]} (1-x) C\left(\nu, \arctan \frac{\left(\frac{1-x}{x}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}}\right) \mathrm{d}x. \end{aligned}$$

The last partial derivative relies on a similar result in [22]; the notation used above also comes from that paper:

$$c_0 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi, \quad c_1 = \int_{-\pi/2}^{\pi/2} (\cos\phi)^\nu \ln(\cos\phi)\mathrm{d}\phi,$$

$$\begin{aligned} D(\nu, z) &= c_0 \int_z^{\pi/2} (\cos\phi)^\nu \ln(\cos\phi)\mathrm{d}\phi - c_1 \int_z^{\pi/2} (\cos\phi)^\nu \mathrm{d}\phi, \\ C(\nu, z) &= D(\nu, z) + (\rho + \sqrt{1-\rho^2}\tan z)^{-\nu} D(\nu, \arccos\rho - z). \end{aligned}$$

All four partial derivatives exist and are continuous functions on $\Theta$.

It can be shown that the partial derivative $\partial H_2/\partial\nu$ is negative for all $(\rho,\nu) \in \Theta$, which implies that the differential is regular in every point in $\Theta$. □

*Proof of Theorem 4.3.* Let $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ denote the projection onto the second coordinate and let $H$ and $K$ be the mappings introduced in (4.1) and (4.2), respectively. Note that $\nu_0$ can be written as

$$\nu_0 = (p_2 \circ K)\left(\rho_0, \int_{[0,1]} R(1-x, x; \rho_0, \nu_0)\mathrm{d}x\right).$$

Moreover, the estimator $\hat{\nu}_n$ has the representation

$$\hat{\nu}_n = (p_2 \circ K)\left(\hat{\rho}, \int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x\right). \tag{6.1}$$

The uniform consistency of $\hat{R}_n$, see the proof of Theorem 2.2 in [8], and the equation (3.1) in [9] imply

$$\int_{[0,1]} \hat{R}_n(1-x, x)\mathrm{d}x \xrightarrow{\mathbb{P}} \int_{[0,1]} R(1-x, x)\mathrm{d}x. \tag{6.2}$$

Hence the right-hand side of (6.1) is well defined with probability tending to one. Further, from the continuous mapping theorem and [24] we know that

$$\hat{\rho} \xrightarrow{\mathbb{P}} \rho_0, \tag{6.3}$$

as $n \to \infty$. Using (6.2), (6.3) and continuity of $p_2 \circ K$, we obtain $\hat{\nu}_n \xrightarrow{\mathbb{P}} \nu_0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. □

Some more notation and technical results are needed for the proof of Theorem 4.4. For $i = 1, \ldots, n$ denote $U_i := 1 - F_1(X_i)$ and $V_i := 1 - F_2(Y_i)$. Let $U_{1:n} \le \cdots \le U_{n:n}$ and $V_{1:n} \le \cdots \le V_{n:n}$ be the corresponding order statistics and by $\lceil a \rceil$ denote the smallest integer not smaller than $a$. Define

$$\begin{aligned}
\hat{R}_n^1(x, y) &:= \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}\left\{R_i^X > n + 1 - kx,\ R_i^Y > n + 1 - ky\right\},\\
R_n(x, y) &:= \frac{n}{k} \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right),\\
T_n(x, y) &:= \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\},
\end{aligned}$$

and note that

$$\hat{R}_n^1(x, y) = T_n\left(\frac{n}{k} U_{\lceil kx \rceil:n}, \frac{n}{k} V_{\lceil ky \rceil:n}\right).$$

It is easily seen that

$$\sup_{(x,y) \in [0, n/k]^2} \sqrt{k} \left| \hat{R}_n^1(x, y) - \hat{R}_n(x, y) \right| \le \frac{1}{\sqrt{k}} \to 0,$$

as $n \to \infty$.

Let $W$, $W_1$ and $W_2$ be as in (4.3) and (4.4). Write

$$v_n(x, y) = \sqrt{k}\left(T_n(x, y) - R_n(x, y)\right), \quad v_{n,1}(x) := v_n(x, \infty) \ \text{ and } \ v_{n,2}(y) := v_n(\infty, y).$$

Proposition 3.1 in [8] shows that for any $T > 0$

$$\begin{aligned}
&\left(v_n(x, y), (x, y) \in [0, T]^2;\ v_{n1}(x), x \in [0, T];\ v_{n2}(y), y \in [0, T]\right)\\
&\quad \xrightarrow{d} \left(W(x, y), (x, y) \in [0, T]^2;\ W_1(x), x \in [0, T];\ W_2(y), y \in [0, T]\right),
\end{aligned}$$

in the topology of uniform convergence, as $n \to \infty$.

Let $F_n(x, y) = (1/n) \sum_{i=1}^{n} \mathbf{1}\{X_i \le x, Y_i \le y\}$ be the empirical distribution function of $F$, and let $F_{n1}$ and $F_{n2}$ be the empirical distribution functions of the marginals $F_1$ and $F_2$, respectively. Define the empirical process $r_n(x, y) := \sqrt{n}(F_n(x, y) - F(x, y))$, $(x, y) \in [-\infty, \infty]^2 =: \bar{\mathbb{R}}^2$, and denote by $W_B$ a Brownian bridge on $[-\infty, \infty]^2$ with covariance structure

$$\mathbb{E} W_B(x_1, y_1) W_B(x_2, y_2) = F(\min\{x_1, x_2\}, \min\{y_1, y_2\}) - F(x_1, y_1) F(x_2, y_2).$$

We know, see e.g. [27], that $r_n \xrightarrow{d} W_B$ in the topology of uniform convergence, as $n \to \infty$. Hence we obtain for the marginal processes, $r_{nj} \xrightarrow{d} W_{Bj}$, where $r_{nj}(x) := \sqrt{n}(F_{nj}(x) - F_j(x))$, $j = 1, 2$, $W_{B1}(x) = W_B(x, \infty)$ and $W_{B2}(x) = W_B(\infty, x)$.

**Lemma 6.1.** *For fixed $(x, y) \in [0, \infty]^2$ and $(t, w) \in [-\infty, \infty]^2$ it holds that as $n \to \infty$, $k \to \infty$, $k/n \to 0$,*

$$\mathbb{E} v_n(x, y) r_n(t, w) \to 0.$$

*Proof.* Fix $(x,y) \in [0,\infty]^2$ and $(t,w) \in \bar{\mathbb{R}}^2$. Then,

$$\begin{aligned}
\mathbb{E} v_n(x,y) r_n(t,w) &= \mathbb{E}\left[\sqrt{k}\,(T_n(x,y) - R_n(x,y)) \cdot \sqrt{n}\,(F_n(x,y) - F(x,y))\right] \\
&= \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum_{i=1}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot \sum_{j=1}^{n} (\mathbf{1}\{X_j \le t, Y_j \le w\} - F(t,w))\right] \\
&= \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum\sum_{i,j=1, i\neq j}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot (\mathbf{1}\{X_j \le t, Y_j \le w\} - F(t,w))\right] \\
&\quad + \frac{1}{\sqrt{kn}} \mathbb{E}\left[\sum_{i=1}^{n}\left(\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right)\right. \\
&\qquad \left. \cdot (\mathbf{1}\{X_i \le t, Y_i \le w\} - F(t,w))\right] \\
&=: E_1 + E_2.
\end{aligned}$$

Using independence of the sample, we get

$$\begin{aligned}
E_1 &= \frac{1}{\sqrt{kn}} \sum_{i,j=1, i\neq j}^{n} \mathbb{E}\left[\mathbf{1}\left\{U_i < \frac{kx}{n},\ V_i < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right] \\
&\qquad \cdot \mathbb{E}\left[\mathbf{1}\{X_j \le t, Y_j \le w\} - F(t,w)\right] \\
&= 0.
\end{aligned}$$

Since the factors in the sum in $E_2$ have the same distribution, we get

$$\begin{aligned}
|E_2| &= \sqrt{\frac{n}{k}}\left|\mathbb{E}\left[\left(\mathbf{1}\left\{U_1 < \frac{kx}{n},\ V_1 < \frac{ky}{n}\right\} - \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right)\right)\right.\right. \\
&\qquad \left.\left. \cdot (\mathbf{1}\{X_1 \le t, Y_1 \le w\} - F(t,w))\right]\right| \\
&\le \sqrt{\frac{n}{k}}\left(\left|\mathbb{E}\left[\mathbf{1}\left\{U_1 < \frac{kx}{n},\ V_1 < \frac{ky}{n}\right\}\mathbf{1}\{X_1 \le t, Y_1 \le w\}\right]\right|\right. \\
&\qquad \left. + \mathbb{P}\left(U_1 \le \frac{kx}{n}, V_1 \le \frac{ky}{n}\right) F(t,w)\right) \\
&\le \sqrt{\frac{n}{k}}\,\mathbb{P}\left(U_1 < \frac{kx}{n}, V_1 < \frac{ky}{n}\right)(1 + F(t,w)) \\
&\le 2\sqrt{\frac{k}{n}}\min\{x,y\} \to 0,
\end{aligned}$$

as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. □

**Lemma 6.2.** *Let $T > 0$. In the topology of uniform convergence, as $n \to \infty$, $k \to \infty$, $k/n \to 0$, the process*

$$(v_n(x,y),(x,y) \in [0,T]^2;\ v_{n1}(x), x \in [0,T];\ v_{n2}(y), y \in [0,T],\ r_n(x,y),(x,y) \in \bar{\mathbb{R}}^2) \tag{6.4}$$

*converges in distribution to*

$$(W(x,y),(x,y) \in [0,T]^2;\ W_1(x), x \in [0,T];\ W_2(y), y \in [0,T],\ W_B(x,y),(x,y) \in \bar{\mathbb{R}}^2), \tag{6.5}$$

*with*

$$(W(x,y),(x,y) \in [0,T]^2;\ W_1(x), x \in [0,T];\ W_2(y), y \in [0,T])$$

*and* $(W_B(x,y),(x,y) \in \bar{\mathbb{R}}^2)$ *independent.*

*Proof.* From the weak convergence, and hence tightness, of

$$\big(v_n(x,y),(x,y) \in [0,T]^2;\ v_{n1}(x), x \in [0,T];\ v_{n2}(y), y \in [0,T]\big)$$

and $(r_n(x,y),(x,y) \in \bar{\mathbb{R}}^2)$, we get the tightness of the process in (6.4).

By the Cramér-Wold device, see for example [28], and the univariate Lindeberg-Feller central limit theorem, using Lemma 6.1, we get convergence of the finite-dimensional distributions. □

Using the Skorohod construction we get a probability space containing all processes $\tilde{v}_n$, $\tilde{v}_{n1}$, $\tilde{v}_{n2}$, $\tilde{r}_n$, $\tilde{W}$, $\tilde{W}_1$, $\tilde{W}_2$ and $\tilde{W}_B$, where

$$\begin{aligned}(\tilde{v}_n, \tilde{v}_{n1}, \tilde{v}_{n2}, \tilde{r}_n) &\overset{d}{=} (v_n, v_{n1}, v_{n2}, r_n),\\ (\tilde{W}, \tilde{W}_1, \tilde{W}_2, \tilde{W}_B) &\overset{d}{=} (W, W_1, W_2, W_B),\end{aligned}$$

and it holds that as $n \to \infty$, $k \to \infty$, $k/n \to 0$,

$$\sup_{(x,y)\in[0,T]^2} |\tilde{v}_n(x,y) - \tilde{W}(x,y)| \to 0 \ \text{ a.s.}, \tag{6.6}$$

$$\sup_{(x,y)\in\bar{\mathbb{R}}^2} |\tilde{r}_n(x,y) - \tilde{W}_B(x,y)| \to 0 \ \text{ a.s.}, \tag{6.7}$$

and the analogous statements hold for marginal processes $v_{n1}$, $v_{n2}$, $r_{n1}$ and $r_{n2}$ as well. We work on this space from now on, but keep the old notation (without tilde's).

**Lemma 6.3.** *Assume the situation as in Theorem 4.4. On the probability space of the Skorohod construction*

$$\left(\sqrt{k}\left(\int_{[0,1]} \hat{R}_n(1-x,x)\,dx - \int_{[0,1]} R(1-x,x)\,dx\right), \sqrt{n}(\hat{\rho}-\rho_0)\right) \\ \overset{\mathbb{P}}{\to} \left(\int_{[0,1]} B(1-x,x)\,dx, N_\rho\right), \tag{6.8}$$

*as $n \to \infty$, $k \to \infty$ and $k/n \to 0$, where $\int_{[0,1]} B(1-x,x)dx$ and $N_\rho$ are independent, and $B$ is the process defined in* (4.5).

*Proof.* By Lemma 6.2 it is sufficient to show that

$$\sqrt{k}\left(\int_{[0,1]} \hat{R}_n(1-x,x)\mathrm{d}x - \int_{[0,1]} R(1-x,x)\mathrm{d}x\right) \overset{\mathbb{P}}{\to} \int B(1-x,x)\mathrm{d}x, \tag{6.9}$$

and

$$\sqrt{n}(\hat{\rho}-\rho_0) \overset{\mathbb{P}}{\to} N_\rho, \tag{6.10}$$

since $N_\rho$ is a functional of $W_B$, by (6.12) and the delta method.

For the convergence in (6.10) we will first show that $\sqrt{n}(\hat{\tau}-\tau_0) \xrightarrow{\mathbb{P}} N_\tau$, where $N_\tau$ is the limiting normal random variable for $\hat{\tau}$, see for example [20] or [24]. By the Hoeffding representation of U-statistics and its properties, see for example [24], we get that

$$\sqrt{n}(\hat{\tau}-\tau_0) = 2\sqrt{n}\left(\iint_{\bar{\mathbb{R}}^2} \Phi(x,y)\mathrm{d}F_n(x,y) - \iint_{\bar{\mathbb{R}}^2} \Phi(x,y)\mathrm{d}F(x,y)\right) + o_{\mathbb{P}}(1),$$

where $\Phi(x,y) = 1-2F_1(x)-2F_2(y)+4F(x,y)$. Let $r_n$, $r_{n1}$, $r_{n2}$, $W_B$, $W_{B1}$ and $W_{B2}$ be as defined before Lemma 6.1. From integration by parts we get

$$\begin{aligned} \sqrt{n}(\hat{\tau}-\tau_0) &= -8\iint_{\mathbb{R}^2} r_n(x,y)\mathrm{d}F(x,y) + 4\int_{\mathbb{R}} r_{n1}(x)\mathrm{d}F_1(x) \\ &\quad +4\int_{\mathbb{R}} r_{n2}(y)\mathrm{d}F_2(y) + o_{\mathbb{P}}(1). \end{aligned} \tag{6.11}$$

Denote

$$N_\tau := -8\iint_{\mathbb{R}^2} W_B(x,y)\mathrm{d}F(x,y) + 4\int_{\mathbb{R}} W_{B1}(x)\mathrm{d}F_1(x) + 4\int_{\mathbb{R}} W_{B2}(y)\mathrm{d}F_2(y). \tag{6.12}$$

The result in (6.7), its marginal versions and (6.11) yield that $\sqrt{n}(\hat{\tau}-\tau_0) \xrightarrow{\mathbb{P}} N_\tau$. Since $\hat{\rho} = \sin((\pi/2)\hat{\tau})$, the delta method yields (6.10), where $N_\rho$ is an appropriate function of $N_\tau$. Note that $N_\rho$ is a normally distributed random variable with mean zero and some variance, $\sigma_\rho^2$, say. □

**Lemma 6.4.** *Assume the situation as in Proposition 4.1. As $n\to\infty$, $k\to\infty$ and $k/n\to 0$,*

$$\frac{\varphi_{\hat{\rho}}^{-1}(\int_{[0,1]} \hat{R}_n(1-x,x)dx) - \varphi_{\hat{\rho}}^{-1}(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)dx)}{\int_{[0,1]} \hat{R}_n(1-x,x)dx - \int_{[0,1]} R(1-x,x;\rho_0,\nu_0)dx} \xrightarrow{\mathbb{P}} c, \tag{6.13}$$

*where $c$ is defined in (4.6), and*

$$\sqrt{k}\left(\varphi_{\hat{\rho}}^{-1}\left(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)dx\right) - \varphi_{\rho_0}^{-1}\left(\int_{[0,1]} R(1-x,x;\rho_0,\nu_0)dx\right)\right) \xrightarrow{\mathbb{P}} 0. \tag{6.14}$$

*Proof.* Throughout the proof we omit writing the region of integration, $[0,1]$. As before, let $H$ be the function on $\Theta$ given by $H(\rho,\nu) = (\rho,\varphi_\rho(\nu))$, let $K$ be its local inverse, and let $p_2$ be the projection onto the second coordinate. Since $K(\rho,\mu) = (\rho,\nu)$, where $\nu$ is such that $\mu = \int_{[0,1]} R(1-x,x;\rho,\nu)\mathrm{d}x$, we see that $(p_2\circ K)(\rho,\mu) = \varphi_\rho^{-1}(\mu)$. Denote $\mu_0 := \int R(1-x,x;\rho_0,\nu_0)\mathrm{d}x$.

First we prove (6.13). Define the function $f\colon [0,1]\to\mathbb{R}$ by

$$f(t) := (p_2\circ K)\left(\hat{\rho}, \mu_0 + t\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0\right)\right).$$

Using the mean value theorem for $f$ on $[0,1]$ we get

$$f(1)-f(0) = (1-0)\cdot f'(t)\big|_{t=t^*}, \quad t^*\in(0,1).$$

Since $f(1) = (p_2\circ K)(\hat{\rho}, \int \hat{R}_n(1-x,x)\mathrm{d}x) = \varphi_{\hat{\rho}}^{-1}(\int \hat{R}_n(1-x,x)\mathrm{d}x)$ and $f(0) = (p_2\circ K)(\hat{\rho},\mu_0) = \varphi_{\hat{\rho}}^{-1}(\mu_0)$, we get

$$\varphi_{\hat{\rho}}^{-1}(\int \hat{R}_n(1-x,x)\mathrm{d}x) - \varphi_{\hat{\rho}}^{-1}(\mu_0) = \frac{\partial}{\partial\mu}(p_2\circ K)(\hat{\rho},\mu)\big|_{\mu=\mu^*}\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0\right),$$

with $\mu^* = \mu_0 + t^*(\int \hat{R}_n(1-x,x)\mathrm{d}x - \mu_0)$. Because $\mu^*$ is between $\int \hat{R}_n(1-x,x)\mathrm{d}x$ and $\mu_0$, the consistency of $\int \hat{R}_n(1-x,x)\mathrm{d}x$ implies that $\mu^* \xrightarrow{\mathbb{P}} \mu_0$, as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. This, together with the consistency of $\hat{\rho}$ and the continuous differentiability of $K$, see Corollary 4.2, implies that the left-hand side of (6.13) converges in probability to $(\partial/\partial\mu)(p_2 \circ K)(\rho_0, \mu_0) = (\partial/\partial\mu)\varphi_{\rho_0}^{-1}(\mu_0)$. By the inverse mapping theorem, this constant equals $c$.

Next we show that (6.14) holds. Similarly, we define the function $f\colon [0,1] \to \mathbb{R}$ by

$$f(t) := (p_2 \circ K)(\rho_0 + t(\hat{\rho} - \rho_0), \mu_0).$$

The mean value theorem applied to $f$ on $[0,1]$ yields

$$f(1) - f(0) = (1-0) \cdot f'(t)\big|_{t=t^*}, \quad t^* \in (0,1).$$

Write $\rho^* := \rho_0 + t^*(\hat{\rho} - \rho_0)$. Since $f(1) = \varphi_{\hat{\rho}}^{-1}(\mu_0)$, and $f(0) = \varphi_{\rho_0}^{-1}(\mu_0)$, the left-hand side of (6.14) is equal to

$$\frac{\partial}{\partial\rho}\varphi_\rho^{-1}(\mu_0)\big|_{\rho=\rho^*}\sqrt{k}(\hat{\rho} - \rho_0). \tag{6.15}$$

By Corollary 4.2, $\rho \mapsto (\partial/\partial\rho)\varphi_\rho^{-1}(\mu_0)$ is continuous, hence it is bounded on a closed neighborhood of $\rho_0$. The consistency of $\hat{\rho}$ then implies that $(\partial/\partial\rho)\varphi_\rho^{-1}(\mu_0)\big|_{\rho=\rho^*}$ is bounded with probability tending to one. Since the rate of convergence of $\hat{\rho}$ is $1/\sqrt{n}$, the expression in (6.15) converges to zero in probability as $n \to \infty$, $k \to \infty$ and $k/n \to 0$. □

*Proof of Theorem 4.4.* Here we again omit writing the region of integration, $[0,1]$, and we write $R(1-x,x)$ instead of $R(1-x,x;\rho_0,\nu_0)$. We have

$$\begin{aligned}\sqrt{k}\,(\hat{\nu}_n - \nu_0) \quad &= \quad \frac{\varphi_{\hat{\rho}}^{-1}(\int \hat{R}_n(1-x,x)\mathrm{d}x) - \varphi_{\hat{\rho}}^{-1}(\int R(1-x,x)\mathrm{d}x)}{\int \hat{R}_n(1-x,x)\mathrm{d}x - \int R(1-x,x)\mathrm{d}x} \\ &\quad \cdot\sqrt{k}\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \int_{[0,1]} R(1-x,x)\mathrm{d}x\right) \\ &\quad + \sqrt{k}\left(\varphi_{\hat{\rho}}^{-1}(\int R(1-x,x)\mathrm{d}x) - \varphi_{\rho_0}^{-1}(\int R(1-x,x)\mathrm{d}x)\right).\end{aligned}$$

By Lemma 6.4 it follows that

$$\sqrt{k}\,(\hat{\nu}_n - \nu_0) = c(1 + o_{\mathbb{P}}(1))\sqrt{k}\left(\int \hat{R}_n(1-x,x)\mathrm{d}x - \int R(1-x,x)\mathrm{d}x\right) + o_{\mathbb{P}}(1). \tag{6.16}$$

Combining (6.8) and (6.16) we conclude that

$$\left(\sqrt{k}(\hat{\nu}_n - \nu_0), \sqrt{n}(\hat{\rho} - \rho_0)\right) \xrightarrow{d} (N_\nu, N_\rho),$$

where $N_\nu$ and $N_\rho$ are independent, and if $\sigma_R^2$ is the variance of $\int B(1-x,x)\mathrm{d}x$, we have that $\sigma_\nu^2 = c^2\sigma_R^2$. □


## Acknowledgments

I gratefully acknowledge financial support from the Open Competition grant from the Netherlands Organisation for Scientific Research (NWO) and from the Deutsches Forschungsgemeinschaft (DFG) grant SNF FOR 916. This research was mainly performed at TUM, Munich and Tilburg University. I would like to thank Claudia Klüppelberg for the hospitality at Technische Universität München, and for her useful comments to the paper, and to John H.J. Einmahl and Johan Segers for their time and valuable help.

[1] Abdous, B., A.-L. Fougères, and K. Ghoudi (2005). Extreme behavior for bivariate elliptical distributions. *The Canadian Journal of Statistics 33*(3), 317–334.

[2] Asimit, A. V. and B. L. Jones (2007). Extreme behavior of bivariate elliptical distributions. *Insurance: Mathematics & Economics 41*(1), 53–61.

[3] Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Chichester: Wiley.

[4] Berman, S. (1992). *Sojourns and extremes of stochastic processes*. Belmont: Wadsworth & Brooks/Cole Advanced Books & Software.

[5] de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. New York: Springer.

[6] Dekkers, A. L. M., J. H. J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics 17*, 1833–1855.

[7] Demarta, S. and A. J. McNeil (2005). The $t$-copula and related copulas. *International Statistical Review 73*(1), 111–129. Preprint version available on http://www.ma.hw.ac.uk/∼mcneil/ftp/tCopula.pdf.

[8] Einmahl, J., L. de Haan, and D. Li (2006). Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics 34*(4), 1987–2014.

[9] Einmahl, J. H. J., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli 14*(4), 1003–1026.

[10] Embrechts, P., A. Mcneil, and D. Straumann (2002). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

[11] Falk, M., J. Hüsler, and R.-D. Reiss (2004). *Laws of Small Numbers: Extremes and Rare Events*. Basel: Birkhäuser-Verlag.

[12] Fang, H.-B., K.-T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*(1), 1–16.

[13] Fang, K.-T., S. Kotz, and K.-W. Ng (1990). *Symmetric Multivariate and Related Distributions*. London: Chapman and Hall.

[14] Hashorva, E. (2005). Extremes of asymptotically spherical and elliptical random vectors. *Insurance: Mathematics & Economics 36*(3), 285–302.

[15] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 1163–1174.

[16] Huang, X. (1992). *Statistics of bivariate extreme values*. Ph. D. thesis, Tinbergen Institute Research Series.

[17] Hult, H. and F. Lindskog (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability 34*(3), 587–608.

[18] Kaynar, B., S. Birbil, and J. Frenk (2007). Application of a general risk management model to portfolio optimization problems with elliptical distributed returns for risk neutral and risk averse decision makers. Preprint version available on http://hdl.handle.net/1765/10151.

[19] Kendall, M. (1938). A new measure of rank correlation. *Biometrika 30*, 81–89.

[20] Kendall, M. (1948). *Rank Correlation Methods*. London: Charles Griffin & Company Limited.

[21] Klüppelberg, C., G. Kuhn, and L. Peng (2007). Estimating tail dependence of elliptical distributions. *Bernoulli 13*(1), 229–251.

[22] Klüppelberg, C., G. Kuhn, and L. Peng (2008). Semi-parametric models for the multivariate tail dependence function - the asymptotically dependent case. *Scandinavian Journal of Statistics 35*(4), 701–718.

[23] Landsman, Z. M. and E. A. Valdez (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal 7*(4), 55–71.

[24] Lee, A. (1990). *U-Statistics*. New York: Marcel Dekker.

[25] Li, D. and L. Peng (2009). Goodness-of-fit test for tail copulas modeled by elliptical copulas. *Statistics and Probability Letters 79*, 1097–1104.

[26] Lindskog, F., A. McNeil, and U. Schmock (2003). Kendall's tau for elliptical distributions. In G. Bol, G. Nakhaeizadeh, S. Rachev, T. Ridder, and K.-H. Vollmer (Eds.), *Credit Risk: Measurement, Evaluation and Management*, pp. 149–156. Heidelberg: Physica-Verlag.

[27] Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics 42*(4), 1285–1295.

[28] Shorack, G. and J. Wellner (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

[29] Smith, R. (1987). Estimating tails of probability distributions. *Annals of Statistics 15*, 1174–1207.