TAIL DEPENDENCE IN INDEPENDENCE

GERRIT DRAISMA, HOLGER DREES, ANA FERREIRA, AND LAURENS DE HAAN

Abstract. We propose a new estimator of the parameter η, introduced by Ledford and Tawn [1996], governing dependence in bivariate distributions with asymptotically independent componentwise maxima. We prove asymptotic normality of this estimator and two other estimators proposed in the quoted paper. For the latter we develop a weighted approximation result for a two-dimensional rank-process. We compare the estimators and a related test for asymptotic independence in a simulation study. Also we show consistency of the resulting estimator for failure probabilities in this set-up. Our estimator for η is inspired by the work of Peng [1999]. Our less strict second order conditions are satisfied by the normal distribution.

1. Introduction

Suppose a region is protected by a river dam against flooding. The water level is regularly observed at two stations, yielding a sample $(X_i, Y_i), 1 \leq i \leq n$. If there is no other protection within the region, the whole area will be flooded if the water level exceeds the height of the dam at one of both points. Hence the probability of a flooding at a particular date is of the form

$$\Pr\{X_i > u \text{ or } Y_i > v\}. \tag{1.1}$$

We assume that (if necessary, after a suitable declustering) the vectors (X_i, Y_i) are independent and identically distributed with distribution function F, say. If the heights u and v of the dam are large, then multivariate extreme value theory provides a framework which allows a systematic estimation of the probability (1.1). For this, assume that there exist normalizing constants $a_n, c_n > 0$ and $b_n, d_n \in \mathbb{R}$ such that

(1.2)
$$\lim_{n\to\infty} F^n(a_n x + b_n, c_n y + d_n) = \lim_{n\to\infty} \Pr\left\{\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \leq x, \frac{\bigvee_{i=1}^n Y_i - d_n}{c_n} \leq y\right\} = G(x, y)$$

for all but denumerable many vectors (x, y). Here $\bigvee_{i=1}^n X_i$ denotes the maximum of n consecutive water levels at the first station and G is a distribution function with non-degenerate marginals [cf. Resnick, 1987, Chapter 5]. Taking logarithms, one concludes from (1.2) that

$$\lim_{n\to\infty} n \Pr\left\{\frac{X - b_n}{a_n} > x \text{ or } \frac{Y - d_n}{c_n} > y\right\} = -\log G(x, y) \tag{1.3}$$

for a random vector (X, Y) with distribution function F.

For the sake of simplicity, in this introduction we concentrate on the case when both marginals are uniformly distributed; this can be achieved by transforming the random

Date: May 21, 2001.

variables X and Y with their pertaining marginal distribution functions F_i (cf. (3.1)). Then (1.3) simplifies to

$$\lim_{n\to\infty} n \Pr\{1 - X < x/n \text{ or } 1 - Y < y/n\} = -\log G(-x,-y). \tag{1.4}$$

and, in fact, even

$$\lim_{s\to 0} s^{-1} \Pr\{1 - X < sx \text{ or } 1 - Y < sy\} = -\log G(-x,-y) \tag{1.5}$$

with s running through the reals. Dividing the analogous equation where s is replaced with st by (1.5), one sees that

$$\Pr\{1 - X < tx \text{ or } 1 - Y < ty\} \approx t \Pr\{1 - X < x \text{ or } 1 - Y < y\} \tag{1.6}$$

for small x and y, i.e., the function $t \mapsto \Pr\{1-X < tx \text{ or } 1-Y < ty\}$ is regularly varying at 0 with index 1.

Recall that we want to estimate the probability (1.1) with u and v so close to 1 that no or only very few observations lie in the failure region $\{(r,s) \in [0,1]^2 \mid 1-r < 1-u \text{ or } 1-s < 1-v\}$. Now choose a sufficiently small t such that the set

$$\{(r,s) \in [0,1]^2 \mid 1 - r < (1-u)/t \text{ or } 1 - s < (1-v)/t\} \tag{1.7}$$

does contain a considerable number of observations and hence the probability that (X,Y) lies in (1.7) can be estimated using the empirical distribution. Then we can use (1.6) with $x = (1-u)/t$ and $y = (1-v)/t$ to estimate the probability (1.1) we are actually interested in.

However, in many situations one may also be interested in the probability that both thresholds are exceeded, i.e., $\Pr\{X > u \textit{ and } Y > v\}$. This probability is of interest, e.g., if the levels of two different air pollutants, the losses suffered in two different investments or different variables relevant for the probability of a flooding (cf. Section 5) are observed. Convergence (1.3) implies

$$\lim_{n\to\infty} n \Pr\left\{\frac{X-b_n}{a_n} > x \text{ and } \frac{Y-d_n}{c_n} > y\right\} = -\log G(x,y) + \log G_1(x) + \log G_2(y), \tag{1.8}$$

since the marginal distributions converge to the marginals G_1 and G_2 of the limit distribution. Note that if the marginals of the limit distributions are independent, that is, $G(x,y) = G_1(x)G_2(y)$, the limit in (1.8) is identically zero. In that case we say that the maxima of the X_i and those of the Y_i are asymptotically independent. This is a rather common situation; for instance, it holds for nondegenerate bivariate normal distributions.

Unfortunately, in this case the reasoning used above to derive estimators for the probability (1.1) does not lead to anything one can employ for the estimation of the probability of a joint exceedance, since the analog to (1.6) does not hold.

In order to overcome this problem, Ledford and Tawn [1996, 1997, 1998] [see also Coles et al., 1999] introduced a quite general submodel, where the tail dependence is characterized by a coefficient $\eta \in (0,1]$. More precisely, in the setting with uniform marginals, they assumed that the function $t \mapsto \Pr\{1 - X < t \text{ and } 1 - Y < t\}$ is regularly varying at 0 with index $1/\eta$. Then $\eta = 1$ in case of asymptotic dependence, whereas $\eta < 1$ implies asymptotic independence. When η is less than 1, the value of η determines the amount of dependence in asymptotic independence (see (2.1) below and the comments thereafter).

Thus the submodel can also be used to device a test for asymptotic independence in the basic relation (1.2).

Moreover Ledford and Tawn proposed an estimator for η. Peng [1999] presented a theoretical background for their model and proposed a non-parametric estimator for η. Peng proved asymptotic normality of his estimator under second order conditions. The present paper contains the following contributions:

1. Peng's conditions are generalized so that, e.g., the normal distribution is included.
2. Asymptotic normality of two modified versions of estimators introduced by Ledford and Tawn is shown under second order conditions (section 2).
3. A new estimator is introduced and its asymptotic normality is derived (section 2).
4. A procedure is set up to estimate the probability of a failure set that works under asymptotic dependence as well as under asymptotic independence. The estimator is proved to be consistent in our model (section 3).
5. A simulation study compares the behavior of the estimators and their use in testing for asymptotic independence. Also the behavior of the estimator for failure probabilities is studied in a simple situation (section 4).

In Section 5 we examine the dependence between still water level, wave heights and wave periods at a particular point of the Dutch coastal protection. Sections 6 and 7 contain the proofs of the results of section 2 and section 3 respectively. An appendix provides some helpful analytical results.

2. Estimating asymptotic dependence or independence

Let (X, Y) be a random vector whose distribution function F has continuous marginal distribution functions F_1 and F_2. Our basic assumption is that

$$\lim_{t\downarrow 0}\Big(\frac{\Pr\{1-F_1(X)<tx \text{ and } 1-F_2(Y)<ty\}}{q(t)}-c(x,y)\Big)\Big/ q_1(t) =: c_1(x,y) \tag{2.1}$$

exists, for $x, y \geq 0$ (but $x + y > 0$), with q positive, $q_1 \to 0$ as $t \to 0$ and c_1 non-constant and not a multiple of c. Moreover we assume that the convergence is uniform on

$$\{(x,y)\in[0,\infty)^2 \mid x^2+y^2=1\}.$$

It follows that the function q is regularly varying at zero of order $1/\eta$, $\eta \in (0, 1]$; q_1 is also regularly varying at zero, but with order $\tau \geq 0$. Without loss of generality we may take $c(1, 1) = 1$, and we may assume that $q(t) = \Pr\{1 - F_1(X) < t \text{ and } 1 - F_2(Y) < t\}$ (see Appendix A). We also assume that $l := \lim_{t\downarrow 0} q(t)/t, t \to 0$ exists. Since $F_1(X)$ and $F_2(Y)$ are uniformly distributed, obviously $\limsup q(t)/t \leq 1$, and $l = 0$ when $\eta < 1$. Our assumptions imply that (2.1) holds locally uniformly on $(0,\infty)^2$ (see Appendix A). The bivariate normal distribution satisfies these conditions: see the example at the end of this section.

The function c is homogeneous of order $1/\eta$, i.e., $c(tx, ty) = t^{1/\eta}c(x, y)$. The measure ν defined by $\nu([0, x] \times [0, y]) = c(x, y)$ inherits this homogeneity:

$$\nu(tA) = t^{1/\eta}\nu(A) \tag{2.2}$$

for $t > 0$ and all bounded Borel sets $A \subset [0,\infty)^2$.

The parameter η is Ledford's and Tawn's coefficient of asymptotic dependence, cf. Ledford and Tawn [1996, 1997]. Now $l > 0$ implies asymptotic dependence, and $l = 0$ implies asymptotic independence. Hence $\eta < 1$ implies asymptotic independence. Condition (2.1) is somewhat similar to condition (2.8) in Ledford and Tawn [1998].

Now we turn to estimators for η, given an i.i.d. sample $\{(X_1, Y_1), (X_2, Y_2), \dots (X_n, Y_n)\}$. We start with an informal introduction to the estimators of Ledford and Tawn [1996]. They proposed first to standardize the marginals to the unit Fréchet distribution, using either the empirical marginal distributions (that is, using the ranks of the components) or extreme value estimators for the marginal tails, and then to estimate η as the shape parameter of the minimum of the components, e.g. by the maximum likelihood estimator or the Hill estimator. However, since these estimators have larger bias for Fréchet distributions than for Pareto distributions, we prefer to standardize to the unit Pareto distribution using the ranks of the components.

For this consider the random vector

$$T := \frac{1}{1 - F_1(X)} \wedge \frac{1}{1 - F_2(Y)}$$

which is in the domain of attraction of the extreme value distribution with parameter $1/\eta$. Since the marginal d.f.'s F_i are unknown, we replace them with their empirical counterparts. This leads to (with a small modification to prevent division by 0):

$$T_i^{(n)} := \frac{n+1}{n+1-R_i^X} \wedge \frac{n+1}{n+1-R_i^Y}, \quad i = 1, \dots, n,$$

with R_i^X denoting the rank of X_i among $(X_1, X_2, \dots, X_n)$ and R_i^Y that of Y_i. Now η can be estimated by the maximum likelihood estimator in a generalized Pareto model, based on the largest $m = m(n)$ order statistics of the $T_i^{(n)}$. This estimator will be denoted by $\hat{\eta}_1$. Alternatively the Hill estimator can be used:

$$\hat{\eta}_2 := \frac{1}{m} \sum_{i=1}^{m} \log \frac{T_{n,n-i+1}^{(n)}}{T_{n,n-m}^{(n)}}.$$

Note that one important advantage of the maximum likelihood estimator over the Hill estimator in the classical i.i.d. setting, namely its location invariance, is not relevant here: there is no shift after standardizing the marginals to unit Pareto (see Lemma 6.3). Since $\hat{\eta}_2$ has smaller variance, one might expect $\hat{\eta}_2$ to outperform $\hat{\eta}_1$ (however, see Section 4).

Next we introduce Peng's estimator and our new proposal. Equation (2.1) implies for $k/n \to 0$ and $s > 0$

$$\frac{\Pr\{1 - F_1(X) < s\,k/n \text{ and } 1 - F_2(Y) < s\,k/n\}}{\Pr\{1 - F_1(X) < k/n \text{ and } 1 - F_2(Y) < k/n\}} = s^{1/\eta}(1 + o(1)) \tag{2.3}$$

locally uniformly. Denote by $X_{n,i}$ and $Y_{n,i}$ the ith order statistics of the X_j and Y_j, $j = 1, \dots, n$, respectively. To estimate η from the sample we may replace in (2.3) Pr, $1 - F_1$ and $1 - F_2$ by their empirical counterparts. Write

$$S_n(j, k) := \sum_{i=1}^{n} \mathbf{1}\{X_i > X_{n,n-j} \text{ and } Y_i > Y_{n,n-k}\}. \tag{2.4}$$

Note that $S_n(j, k) \approx n \; \Pr\{1 - F_1(X) < j/n \text{ and } 1 - F_2(Y) < k/n\}$.

Using $s = 2$ in (2.3) leads to Peng's estimator [Peng, 1999]:

$$\hat{\eta}_3 = \log 2 / \log\Big(\frac{S_n(2k, 2k)}{S_n(k, k)}\Big).$$

We propose the following estimator, based on integrating (2.3) with respect to s from 0 to 1:

$$\hat{\eta}_4 := \frac{\sum_{j=1}^{k} S_n(j,j)}{kS_n(k,k) - \sum_{j=1}^{k} S_n(j,j)} \tag{2.5}$$

with S_n as in equation (2.4).

Note that $\hat{\eta}_1$ and $\hat{\eta}_2$ are based on the empirical quantile function, and $\hat{\eta}_3$ and $\hat{\eta}_4$ on the empirical distribution function.

We first have to prove the consistency of the new estimator.

Theorem 2.1 (Consistency)**.** *Suppose for* $x, y \geq 0$

$$\lim_{t \to 0} \frac{\Pr\{1 - F_1(X) \leq tx \text{ and } 1 - F_2(Y) \leq ty\}}{q(t)} = c(x,y) \tag{2.6}$$

where q *and* c *are positive functions. Let* $k = k(n)$, $r(n) := n\, q(k/n) \to \infty$ *(this implies* $k \to \infty$*) and* $k/n \to 0$ *for* $n \to \infty$*. Then*

$$\hat{\eta}_4 \to \eta$$

in probability, with η *the reciprocal of the index of regular variation of* q *at 0.*

Remark. Note (2.1) is not needed here. Moreover $\hat{\eta}_1$ and $\hat{\eta}_2$ are consistent too if $m = \lfloor r(n) \rfloor$.

The next Theorem states the asymptotic normality of all estimators considered.

Theorem 2.2 (Asymptotic normality)**.** *Assume* (2.1)*. Additionally assume that* c *has first order partial derivatives* $c_x = \frac{\partial}{\partial x} c(x,y)$ *and* $c_y = \frac{\partial}{\partial y} c(x,y)$*. Suppose* $k = k(n)$, $r(n) = n\, q(k/n) \to \infty$ *(this implies* $k \to \infty$*),* $k/n \to 0$, $\sqrt{r(n)}\, q_1(k/n) \to 0$ *as* $n \to \infty$*, and* $m = m(n) = \lfloor r(n) \rfloor$*.*

Under these conditions $\sqrt{r(n)}\big(\hat{\eta}_i - \eta\big)$ *are asymptotically normal with mean 0 and variance* σ_i^2, $i = 1, 2, 3, 4$*. The variances are*

$$\sigma_1^2 = (1+\eta)^2 (1-l)(1 - 2l c_x(1,1) c_y(1,1)) \tag{2.7}$$

$$\sigma_2^2 = \eta^2 (1-l)(1 - 2l c_x(1,1) c_y(1,1)) \tag{2.8}$$

$$\begin{aligned} \sigma_3^2 = \eta^4 (\log 2)^{-2} (1 - 2^{-1/\eta}) \Big[&\frac{1}{2} (1-3l)(1 - 2l c_x(1,1) c_y(1,1)) \\ &+ l c(1,2) c_x(1,1)(1 - l c_y(1,1)) + l c(2,1) c_y(1,1)(1 - l c_x(1,1)) \Big] \end{aligned} \tag{2.9}$$

$$\begin{aligned} \sigma_4^2 = \frac{(1+\eta)^2 \eta^2}{2\eta + 1} \Big[&(1-3l)(1 - 2l c_x(1,1) c_y(1,1)) \\ &+ 4l c_x(1,1)(1 - l c_y(1,1)) \int_0^1 c(u,1)\, du \\ &+ 4l c_y(1,1)(1 - l c_x(1,1)) \int_0^1 c(1,u)\, du \Big]. \end{aligned} \tag{2.10}$$

Remark. The assertion for $\hat{\eta}_3$ is a generalization of Peng's, since our conditions are weaker.

Remark. Note that instead of (2.1) the weaker condition $\lim_{t\to 0} \Pr\{1 - F_1(X) < tx$ and $1 - F_2(Y) < ty\}/q(t) - c(x,y) = O(q_1(t))$ is sufficient for Theorem 2.2. However, under (2.1) similar results can be easily deduced if the intermediate sequence k is such that $\sqrt{r(n)}q_1(k/n) \to c \geq 0$. In that case, usually a non-negligible bias occurs if $c > 0$ (and the present results correspond to the simpler case $c = 0$).

Theorem 2.2 may be stated without the unknown sequence $r(n)$ entering explicitly the formulation, as in the following Corollary.

Corollary 2.1. *Assume the conditions of Theorem 2.2. For $i = 3, 4$*

$$\sqrt{S_n(k,k)}(\hat{\eta}_i - \eta)$$

has the limiting distribution of Theorem 2.2, with $S_n(j,k)$ as in equation (2.4).

Remark. When using $\hat{\eta}_1$ or $\hat{\eta}_2$, the choice of the number $m = \lfloor r(n) \rfloor$ of largest order statistics from $T_{n,i}^{(n)}$ is up to the statistician, so there is no need to estimate $r(n)$.

Corollary 2.1, together with consistent estimators for the unknown quantities in the asymptotic variances in Theorem 2.2, can be used to construct a confidence interval for η or to test the hypothesis $\eta = 1$. The following Theorem provides these estimators.

Theorem 2.3. (i) *Define*

$$\hat{c}_x(1,1) := k^{1/4} \frac{S_n(\lfloor k(1+k^{-1/4}) \rfloor, k) - S_n(k,k)}{S_n(k,k)},$$

$$\hat{c}_y(1,1) := k^{1/4} \frac{S_n(k, \lfloor k(1+k^{-1/4}) \rfloor) - S_n(k,k)}{S_n(k,k)},$$

$$\hat{d}_1 := \frac{\sum_{j=1}^{k} S_n(j,k)}{k\ S_n(k,k)}, \qquad \hat{d}_2 := \frac{\sum_{j=1}^{k} S_n(k,j)}{k\ S_n(k,k)},$$

$$\hat{l} := \frac{S_n(k,k)}{k}$$

with $S_n(i,j)$ as in equation (2.4). *If the conditions of Theorem 2.2 hold then*

$$\hat{l} \xrightarrow{p} l.$$

If, in addition, $\eta > 1/2$ then

$$\hat{c}_x(1,1) \xrightarrow{p} c_x(1,1), \qquad \hat{c}_y(1,1) \xrightarrow{p} c_y(1,1),$$

$$\hat{d}_1 \xrightarrow{p} \int_0^1 c(u,1)du, \qquad \hat{d}_2 \xrightarrow{p} \int_0^1 c(1,u)du.$$

Moreover, let

$$\hat{\sigma}_1^2 := (1+\hat{\eta})^2(1-\hat{l})(1-2\hat{l}\hat{c}_x(1,1)\hat{c}_y(1,1))$$

and define $\hat{\sigma}_i^2$, $i = 2,3,4$, likewise by (2.8)–(2.10) with $\eta, l, c_x(1,1), c_y(1,1)$, $\int_0^1 c(u,1)\,du$ and $\int_0^1 c(1,u)\,du$ replaced by their respective estimator. Then $\hat{\sigma}_i^2$, $i = 1, \ldots, 4$, are consistent estimators of σ_i^2 for all $\eta \in (0,1]$.

(ii) *The analogous assertion to (i) holds for the estimators*

$$\tilde{l} := \frac{m}{n} T^{(n)}_{n,n-m}$$
$$\tilde{c}_x(1,1) := \frac{\hat{k}^{5/4}}{n} \big(T^{(n,\hat{k}^{-1/4})}_{n,m} - T^{(n)}_{n,m} \big)$$

with $m := m(n) := \lfloor r(n) \rfloor$, $\hat{k} := m/\tilde{l}$, *and* $T^{(n,u)}_{n,i}, i = 1, \dots, n$, *the order statistics of*

$$T^{(n,u)}_i := \min \Big(\frac{n+1}{n+1-R^X_i}(1+u), \frac{n+1}{n+1-R^Y_i} \Big), \quad i = 1, \dots, n,$$

and $\tilde{c}_y(1,1)$ *defined analogously to* $\tilde{c}_x(1,1)$.

Remark. Note that $\tilde{c}_y(1,1)$ may also be estimated as $1/\hat{\eta} - \tilde{c}_x(1,1)$, provided $\eta > 1/2$.

Example 2.1. The bivariate normal distribution with mean 0, variance 1 and correlation coefficient $\rho \notin \{1, -1\}$, satisfies (2.1) with

$$\eta = (1+\rho)/2, \qquad c(x,y) = (xy)^{1/(1+\rho)},$$
$$q(t) = k_1(\rho) t^{2/(1+\rho)} (-\log t)^{-\rho/(1+\rho)} \left\{ 1 - k_2(\rho) \frac{\log(-\log t)}{2 \log t} \right\},$$
$$c_1(x,y) = -k_3(\rho) - k_4(x,y,\rho), \qquad q_1(t) = \frac{1}{2\log t},$$

where

$$k_1(\rho) = \frac{(1-\rho^2)^{3/2}}{(1-\rho)^2} (4\pi)^{-\rho/(1+\rho)}, \qquad k_2(\rho) = \frac{\rho}{1+\rho},$$
$$k_3(\rho) = \frac{\rho \log(4\pi) + 2}{1+\rho} - \frac{(1+\rho)(2-\rho)}{1-\rho},$$
$$k_4(x,y,\rho) = \log x + \log y + \frac{(\rho-1)(\log x + \log y) + \rho(\log x)(\log y) - \rho\left((\log x)^2 + (\log y)^2\right)/2}{(1-\rho^2)}.$$

This can be checked using the tail expansion of the bivariate normal distribution by Ruben [1964] as given in [Ledford and Tawn, 1997], combined with a sufficiently precise expansion of the function f, the inverse function of $1/(1-\Phi)$ where Φ is the standard univariate normal distribution function:

$$f^2(t) = 2\log t - \log(\log t) - \log(4\pi) + \frac{\log(\log t)}{2\log t} + \frac{\log(4\pi) - 2}{2\log t} + \frac{1}{2}\left(\frac{\log(\log t)}{2\log t}\right)^2 + o\left(\left(\frac{\log(\log t)}{2\log t}\right)^2\right), \text{ as t} \to \infty.$$

3. Estimation of failure probabilities

Throughout this section we assume that the marginal distribution functions F_i of F are continuous and belong to the domain of attraction of a univariate extreme value distribution. Moreover, condition (2.6) and further conditions ensuring $\hat{\eta} - \eta = O_P((r(n))^{-1/2})$ shall hold (cf. Section 2).

Recall from (1.6) that, if we want to estimate the probability of an extreme set of the form $\{X > x \text{ or } Y > y\}$ and we assume that F belongs to the domain of attraction of a bivariate extreme value distribution, then we can use the approximate equality

$$
\begin{aligned}
(3.1)\quad & \Pr\{1 - F_1(X) < 1 - F_1(x) \text{ or } 1 - F_2(Y) < 1 - F_2(y)\} \\
& \approx t \Pr\{1 - F_1(X) < (1 - F_1(x))/t \text{ or } 1 - F_2(Y) < (1 - F_2(y))/t\}
\end{aligned}
$$

since for small t the right hand side can be estimated using the empirical distribution function [de Haan and Sinha, 1999]. However, if the marginals are asymptotically independent and the failure set is e.g. of the form $\{X > x \textit{ and } Y > y\}$ then a different approximate equality holds under condition (2.1) or (2.6):

$$
\begin{aligned}
(3.2)\quad & \Pr\{1 - F_1(X) < 1 - F_1(x) \text{ and } 1 - F_2(Y) < 1 - F_2(y)\} \\
& \approx t^{1/\eta} \Pr\{1 - F_1(X) < (1 - F_1(x))/t \text{ and } 1 - F_2(Y) < (1 - F_2(y))/t\}.
\end{aligned}
$$

We develop an estimation procedure which works in this situation.

More generally, we aim at the estimation of the failure probability $p_n = \Pr\{(X,Y) \in C_n\}$ for failure regions $C_n \subset [x_n, \infty] \times [y_n, \infty]$ for some $x_n, y_n \in \mathbb{R}$ such that

$$
(3.3)\qquad (x,y) \in C_n \quad \Longrightarrow \quad [x,\infty] \times [y,\infty] \subset C_n.
$$

The latter property means that if an observation (x, y) causes a failure (e.g., a flooding of a dike) then an event with both components larger will do so, too. Asymptotically we let both x_n and y_n converge to the right endpoint of the pertaining marginal distribution to ensure that $p_n \to 0$, i.e., that indeed we are estimating the probability of an extremal event.

The basic idea is to use a generalized version of the scaling property (3.2) to inflate the transformed failure set $(1 - F_1, 1 - F_2)(C_n) := \{(1 - F_1(x), 1 - F_2(y)) \mid (x,y) \in C_n\}$ such that it contains sufficiently many observations and hence the empirical probability gives an accurate estimate. Since the marginal distribution functions F_i are unknown, their tails are estimated by suitable generalized Pareto distributions.

To work out this program, first recall from univariate extreme value theory that there exist normalizing constants $a_i(n/k) > 0$ and $b_i(n/k) \in \mathbb{R}$ such that the following generalized Pareto approximation is valid:

$$
1 - F_i(x) \approx \frac{k}{n}\Big(1 + \gamma_i \frac{x - b_i(n/k)}{a_i(n/k)}\Big)^{-1/\gamma_i} =: \frac{k}{n}(1 - F_{a_i,b_i,\gamma_i}(x)), \quad i = 1, 2,
$$

for x close to the right endpoint $F_i^{-1}(1)$. Here a_i and b_i are abbreviations for $a_i(n/k)$ and $b_i(n/k)$, respectively, and $(1 + \gamma x)^{-1/\gamma}$ is defined as ∞ if $\gamma > 0$ and $x \le -1/\gamma$, and it is defined as 0 if $\gamma < 0$ and $x \ge -1/\gamma$. Dekkers et al. [1989] proposed and analyzed the following estimators of the parameters a_i, b_i and γ_i. Define

$$\begin{aligned} M_r(X) &:= \frac{1}{k}\sum_{j=0}^{k-1}(\log X_{n,n-j} - \log X_{n,n-k})^r,\ r = 1,2; \\ \hat{\gamma}_1 &:= M_1(X) + 1 - \frac{1}{2}\left(1 - \frac{(M_1(X))^2}{M_2(X)}\right)^{-1}, \\ \hat{b}_1\left(\frac{n}{k}\right) &:= X_{n,n-k}, \\ \hat{a}_1\left(\frac{n}{k}\right) &:= \frac{X_{n,n-k}\sqrt{3M_1(X)^2 - M_2(X)}}{\sqrt{(1-4\hat{\gamma}_1^-)/((1-\hat{\gamma}_1^-)^2(1-2\hat{\gamma}_1^-))}} \quad \text{with} \quad \hat{\gamma}_1^- := \hat{\gamma}_1 \wedge 0; \end{aligned}$$

for $\hat{\gamma}_2$, $\hat{a}_2$ and $\hat{b}_2$ replace X by Y in the previous formulae. The estimator $\hat{\gamma}_i$ for the extreme value index γ_i is often called moment estimator.

Using these definitions, $\frac{n}{k}(1 - F_i(x))$ may be estimated by

$$1 - F_{\hat{a}_i,\hat{b}_i,\hat{\gamma}_i}(x) = \left(1 + \hat{\gamma}_i \frac{x - \hat{b}_i(n/k)}{\hat{a}_i(n/k)}\right)^{-1/\hat{\gamma}_i}.$$

Write $\mathbf{1} - \boldsymbol{F}(x,y)$ as a short form for $(1 - F_1(x), 1 - F_2(y))$, and likewise $\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}} = (1 - F_{a_1,b_1,\gamma_1}, 1 - F_{a_2,b_2,\gamma_2})$ and $\mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}} = (1 - F_{\hat{a}_1,\hat{b}_1,\hat{\gamma}_1}, 1 - F_{\hat{a}_2,\hat{b}_2,\hat{\gamma}_2})$ are functions from $\mathbb{R}^2$ to $[0,\infty]^2$. Then the transformed failure set $\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))$ can be approximated by

$$D_n := \mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(C_n)$$

which in turn is estimated by

$$\hat{D}_n := \mathbf{1} - \boldsymbol{F}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(C_n).$$

Now we may argue heuristically as follows, using a generalization of the scaling property (3.2) to inflate the transformed failure set by the factor $1/c_n$ for some $c_n \to 0$ chosen in a suitable way by the statistician:

$$\begin{aligned} p_n &= \Pr\{\mathbf{1} - \boldsymbol{F}(X,Y) \in \mathbf{1} - \boldsymbol{F}(C_n)\} \\ &\approx \Pr\left\{\frac{n}{k}\big(\mathbf{1} - \boldsymbol{F}(X,Y)\big) \in D_n\right\} \\ &\approx c_n^{1/\eta}\Pr\left\{\frac{n}{k}\big(\mathbf{1} - \boldsymbol{F}(X,Y)\big) \in \frac{D_n}{c_n}\right\} \qquad (3.4) \\ &\approx c_n^{1/\hat{\eta}}\Pr\{(X,Y) \in B\}_{|B = \boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - \frac{\hat{D}_n}{c_n})} \\ &\approx c_n^{1/\hat{\eta}}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\Big\{(X_i,Y_i) \in \boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\Big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\Big)\Big\} \qquad (3.5) \\ &=: \hat{p}_n \qquad (3.6) \end{aligned}$$

where $\hat{\eta}$ denotes one of the estimators examined in Section 2.

In the sequel we state the exact conditions under that we will prove consistency of the estimator $\hat{p}_n$, that is, $\hat{p}_n/p_n \to 1$ in probability as $n \to \infty$. In order not to overload the paper, we will not determine the nondegenerate limit distribution of the standardized estimation error. However, employing the ideas of de Haan and Sinha [1999], one may establish asymptotic normality of $\hat{p}_n$ under more complex conditions.

To study the asymptotic behavior of $\hat{p}_n$, we have to impose a regularity condition on the sequence of failure sets C_n, or rather on the transformed sets D_n. Note that D_n shall shrink towards the origin because we are interested in extremal events. We assume that, after a suitable standardization, D_n converges in the following sense:

(D) There exist a sequence $d_n \to 0$ and a measurable bounded set $A \subset [0,\infty)^2$ with $\nu(A) > 0$ such that for all $\epsilon > 0$ one has for sufficiently large n

$$A_{-\epsilon} \subset \frac{D_n}{d_n} \subset A_{+\epsilon}.$$

Here $A_{+\epsilon} := \{\boldsymbol{x} \in [0,\infty)^2 \mid \inf_{\boldsymbol{y}\in A} \|\boldsymbol{x}-\boldsymbol{y}\| \le \epsilon\}$ and $A_{-\epsilon} := [0,\infty)^2 \setminus (([0,\infty)^2 \setminus A)_{+\epsilon})$ denote the outer and inner ϵ–neighborhood of A with respect to the maximum norm $\|\boldsymbol{x}-\boldsymbol{y}\| = |x_1 - y_1| \vee |x_2 - y_2|$, and ν is the measure corresponding to the function c (cf. Section 2).

Note that d_n and A are not determined by this condition as the former may be multiplied by a fixed factor and the latter divided by the same number. Moreover, even for given d_n the set A is determined only up to its boundary.

Condition (3.3) on C_n implies

$$(x,y) \in D_n \quad \Longrightarrow \quad [0,x]\times[0,y] \subset D_n. \tag{3.7}$$

Example. For $C_n = [x_n,\infty]\times[y_n,\infty]$ we have $D_n = [0, 1-F_{a_1,b_1,\gamma_1}(x_n)]\times[0,1-F_{a_2,b_2,\gamma_2}(y_n)]$. Hence (D) is satisfied with $d_n = 1 - F_{a_1,b_1,\gamma_1}(x_n)$ if $(1-F_{a_2,b_2,\gamma_2}(y_n))/(1-F_{a_1,b_1,\gamma_1}(x_n))$ converges in $(0,\infty)$.

This example demonstrates that essentially (D) means that the convergence of the failure set in the x- and the y-direction is balanced.

Next we need a certain rate of convergence for the marginal estimators to ensure that the transformation of the failure set does not introduce too big an error. For that purpose recall that

$$R_i(t,x) := t\big(1 - F_i(a_i(t)x + b_i(t))\big) - (1+\gamma_i x)^{-1/\gamma_i} \to 0, \quad i = 1,2,$$

locally uniformly for $x \in (0,\infty]$ as $t \to \infty$, since F_i belongs to the domain of attraction of an extreme value distribution. Here we impose the following slightly stricter condition:

$$R_{x_1,x_2}(t) := \max_{i=1,2} \sup_{x_i < x < 1/(-\gamma_i \vee 0)} \Big| R_i(t,x)(1+\gamma_i x)^{1/\gamma_i} \Big| \to 0 \tag{3.8}$$

for some $-1/(\gamma_i \vee 0) < x_i < 1/(-\gamma_i \vee 0)$, $i = 1,2$. Observe that then (3.8) even holds for all such x_i. For example, if F_i satisfies the second order condition

$$\frac{R_i(t,x)}{A_i(t)} \to \Psi(x)$$

for some ρ_i–varying function A_i with $\rho_i < 0$ $(i = 1,2)$, then (3.8) holds true with $R_{x_1,x_2}(t) = O(A_1(t) \vee A_2(t))$. In addition, we require that not too many order statistics are used for estimation of the marginal parameters:

$$k^{1/2} R_{x,x}\big(\frac{n}{k}\big) = O(1) \tag{3.9}$$

for some $x < 0$. Then it follows that the estimators $\hat{a}_i$, $\hat{b}_i$ and $\hat{\gamma}_i$ are $\sqrt{k}$–consistent in the following sense:

$$\Big|\frac{\hat{a}_i}{a_i} - 1\Big| \vee \Big|\frac{\hat{b}_i - b_i}{a_i}\Big| \vee |\hat{\gamma}_i - \gamma_i| = O_P(k^{-1/2}), \quad i = 1,2 \tag{3.10}$$

[cf. Dekkers et al., 1989, de Haan and Resnick, 1993].

We will see that using the estimated parameters instead of the unknown true ones for the transformation of the failure sets does not cause problems provided

$$w_{\gamma_1\wedge\gamma_2}(d_n) = o(k^{1/2}) \quad \text{with} \quad w_\gamma(x) := -x^\gamma \int_x^1 u^{-\gamma-1}\log u\, du. \tag{3.11}$$

Check that

$$w_\gamma(x) \sim \begin{cases} -\frac{1}{\gamma}\log x & ,\gamma > 0 \\ \frac{(\log x)^2}{2} & ,\gamma = 0 \\ \frac{x^\gamma}{\gamma^2} & ,\gamma < 0, \end{cases}$$

as $x \to 0$. Though, at first glance, (3.11) seems rather strict a condition if one of the extreme value indices is negative, it is indeed a natural one; for without it the difference between the transformed set D_n and its estimate $\hat{D}_n$ would be at least of the same order in probability as the typical elements of D_n, namely at least of the order d_n, which of course would render impossible any further statistical inference on the failure probability.

In addition, the scaling factor c_n chosen by the statistician when applying the estimator $\hat{p}_n$ must be related to the actual scaling factor d_n as follows:

$$d_n = O(c_n), \quad w_{\gamma_1\wedge\gamma_2}\big(\frac{c_n}{d_n}\big) = o(k^{1/2}) \quad \text{and} \quad \big(\frac{c_n}{d_n}\big)^{1/\eta} = o\big((r(n))^{1/2}\big). \tag{3.12}$$

In particular, (3.12) is satisfied if c_n and d_n are of the same order. Below the choice of c_n is discussed more thoroughly.

Recall from Section 2 that the scaling property (3.2) is a consequence of approximation (2.6) and the homogeneity of the measure ν. In order to justify (3.4) in the motivation for $\hat{p}_n$ given above, we need the following modification of (2.6), which is suitable for more general sets than just upper quadrants:

$$\sup_{B\in\bar{\mathcal{B}}_n}\left|\frac{\Pr\{\mathbf{1} - \boldsymbol{F}(X,Y) \in \mathbf{1} - \boldsymbol{F}(B)\}}{q(k/n)\nu\big(\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(B))\big)} - 1\right| \to 0 \quad \text{as } n \to \infty \tag{3.13}$$

where

$$\mathcal{B}_n := \Big\{\boldsymbol{F}^{-1}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\mathbf{1} - \boldsymbol{F}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}(C_n)}{c_n}\big)\ \Big|\ \big\|\frac{\tilde{\boldsymbol{a}}}{\boldsymbol{a}} - 1\big\| \vee \big\|\frac{\tilde{\boldsymbol{b}} - \boldsymbol{b}}{\boldsymbol{a}}\big\| \vee \|\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\| \le \epsilon_n\Big\}$$

for some $\epsilon_n \to 0$ such that $k^{1/2}\epsilon_n \to \infty$, and

$$\bar{\mathcal{B}}_n := \mathcal{B}_n \cup \Big\{C_n, \bigcup_{B\in\mathcal{B}_m, m\ge n} B\Big\}.$$

It will turn out (see (7.7)) that for sufficiently large n the denominator in (3.13) is strictly positive.

Notice that the convergence of the absolute value in (3.13) for sets of the type $\mathbf{1} - \boldsymbol{F}(B) = [0, xk/n] \times [0, yk/n]$ is equivalent to convergence (2.6) with $t = k/n$.

Finally, to make approximation (3.5) rigorous, we need a kind of uniform law of large numbers. This is provided by the theory of Vapnik-Cervonenkis (VC) classes of sets as outlined, e.g., in the monograph by Pollard [1984, Section II.4]. For this we require

$$\mathcal{B} = \bigcup_{n \in I\!N} \mathcal{B}_n \quad \text{is a VC class.} \tag{3.14}$$

Theorem 3.1. *Suppose the conditions (D),* (3.3) *(or* (3.7)*),* (3.8)*,* (3.9) *and* (3.11)–(3.14) *are satisfied. If* $\hat{\eta} - \eta = O_P((r(n))^{-1/2})$, $\log c_n = o((r(n))^{1/2})$*, and* $k(n)/n$ *is almost decreasing, which means* $\sup_{m \geq n} k(m)/m = O(k(n)/n)$*, then*

$$\frac{\hat{p}_n}{p_n} \to 1 \quad \textit{in probability.}$$

Remark. (i) In the most important case that np_n is bounded, the conditions (3.11)–(3.13) can be jointly satisfied only if $\gamma_1 \wedge \gamma_2 > -1/2$.

(ii) The sequence $k(n)/n$ is almost decreasing, e.g., if $k(n)$ is regularly varying with exponent less than 1 or, more general, has an upper Matuszewska index $\alpha \leq 1$ [see Bingham et al., 1987, Theorem 2.2.2].

The scaling factor $1/c_n$ by which the transformed failure set is inflated determines the number of large observations taken into account for the empirical probability (3.5). More precisely, according to (7.8) in the proof of Lemma 7.4, this number is of the order $r(n)(d_n/c_n)^{1/\eta}$. Hence if d_n and c_n are of the same order then one uses essentially the same number $S_n(k,k)$ of observations as for the estimation of η, which seems quite natural.

In practice, of course, d_n is not known. However, conversely one may choose c_n such that about $S_n(k,k)$ observations lie in the inflated set $\hat{D}_n/c_n$. To be more concrete, let

$$c_n(\lambda) := \sup\Big\{c > 0 \,\Big|\, \sum_{i=1}^{n} \mathbf{1}\big\{(X_i, Y_i) \in \boldsymbol{F}^{-1}_{\hat{\mathbf{a}},\hat{\mathbf{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c}\big)\big\} \geq \lambda S_n(k,k)\Big\} \tag{3.15}$$

for some $\lambda > 0$. Following the lines of the proof of Theorem 3.1, one may show that $c_n(\lambda)$ and d_n are of the same order in probability, and that the resulting estimator $\hat{p}_n$ is consistent for p_n. Alternatively, one may employ a heuristic approach which is common in univariate extreme value statistics: one plots $\hat{p}_n$ as a function of c_n and choose a value c_n where this graph seems sufficiently stable.

Finally, it is worth mentioning that it is not necessary to use the same number k in the estimation of the marginal parameters and in the definition of $\hat{\eta}_3$ and $\hat{\eta}_4$. In fact, one may prove an analog to Theorem 3.1 in more general settings, provided it is guaranteed that the estimation error introduced when standardizing the marginals is asymptotically negligible, that is, one has (3.10), (3.11) and (3.12) for some k which may differ from the one used in the definition of the estimator for η. Likewise one may use other estimators for the marginal parameters, like e.g. the maximum likelihood estimator examined by Smith [1987], provided these estimators converge with the same rate.

4. Simulations

4.1. **Methods.** The estimators were tested on 4 different distribution functions:

1. the bivariate Cauchy distribution ($\eta = 1$),
2. the bivariate extreme value distribution (BEV) with a logistic dependence function, with $\alpha = 0.75$ ($\eta = 1$) [Ledford and Tawn, 1996, 1997],
3. the bivariate normal distribution with $\rho = 0.6$ ($\eta = 0.8$) and

4. the Morgenstern distribution with $\alpha = 0.75$ ($\eta = 0.5$) [Ledford and Tawn, 1996, 1997].

From each distribution we generated 250 samples of size 1000. A sample of each the distributions is shown in figure 1. Dependence in these distributions ranges from clear asymptotic dependence (1) through weak asymptotic dependence (2) and non-asymptotic dependence (3) to clear asymptotic independence (4).

The ML-estimator $\hat{\eta}_1$ was estimated by the GAUSS `maxlik` procedure [Schoenberg, 1996].

For each estimator two estimates for the root variance are reported:

$\hat{\sigma}_{(i)}$ the root variance for the general case, calculated as $m^{-1/2}\hat{\sigma}_i$, $i = 1, 2$, resp. $S_n(k,k)^{-1/2}$ $\hat{\sigma}_i$, $i = 3, 4$ (cf. Theorems 2.2 and 2.3), and

$\hat{\sigma}_{(d)}$ the root variance for the case of asymptotic dependence, calculated similarly, with $\eta = 1$.

For comparison the observed empirical standard deviation was calculated from the 250 simulated η estimates.

Correspondingly, one-sided 5% tests for dependence were carried out in two ways. Asymptotic dependence is not rejected when

$$\Phi((1-\hat{\eta})/\hat{\sigma}_{(i)}) \leq 0.95 \quad \text{or alternatively} \quad \Phi((1-\hat{\eta})/\hat{\sigma}_{(d)}) \leq 0.95$$

where Φ represents the standard normal df.

Furthermore, we studied the finite sample behavior of the proposed estimators of a failure probability. For this, failure sets of the type $[a, \infty)^2$ were considered, where a was chosen such that the failure probability equals $p_n = (100n)^{-1} = 10^{-5}$ for the sample size $n = 1000$. We used $\hat{\eta}_4$ as the estimator for the parameter of tail dependence and considered three different estimators of p_n:

$$\begin{aligned} \hat{p}_{\hat{\eta}} &= \hat{p}_n \quad \text{as defined in (3.6),} \\ \hat{p}_1 &= c_n^{1-1/\hat{\eta}}\hat{p}_n, \quad \text{and} \\ \hat{p} &= \begin{cases} \hat{p}_1 & \text{if } \eta = 1 \text{ is not rejected,} \\ \hat{p}_{\hat{\eta}} & \text{if } \eta = 1 \text{ is rejected.} \end{cases} \end{aligned} \tag{4.1}$$

Here $c_n = c_n(1)$ is defined by (3.15) and the test for $\eta = 1$ is based on the variance estimate $\hat{\sigma}_{(d)}$. Note that $\hat{p}_1$ is a natural analog to $\hat{p}_{\hat{\eta}}$ if it is known in advance that $\eta = 1$. In particular, in that case it is a consistent estimator of p_n.

For the normal distribution this resulted in many zero estimates; as this effect was caused by the poor estimates of the marginal parameters, we also considered a distribution with the marginals transformed to standard exponential.

4.2. **Estimating η and testing for asymptotic dependence.** The results are presented in figures 2 and 3 and in tables 1 and 2.

To make the performance of the different estimators for η comparable, m and k were chosen in a range where the overall performance of the estimator under consideration is best. This led to a smaller m for the Hill than for the maximum likelihood estimator, because of the larger bias of former. Recall that Peng's estimator is constructed from $S_n(k,k)$ and $S_n(2k,2k)$, while $\hat{\eta}_4$ is based on $S_n(j,j)$ only up to $j = k$. For that reason we chose k for $\hat{\eta}_4$ double as large as for $\hat{\eta}_3$.

The general picture is that $\hat{\eta}_2$, $\hat{\eta}_3$ and $\hat{\eta}_4$ show a bias (negative for the Cauchy, BEV and normal distributions) that increases with m or k. The ML estimator $\hat{\eta}_1$ shows no clear trend, but it is biased for two of the distributions, though considerably less than the other estimators. Note that although $\eta \leq 1$, the estimates $\hat{\eta}_i$ may be larger than 1.

A comparison of the observed standard deviation with the appropriate estimates (tables 1–2: $\hat{\sigma}_{(d)}$ for the Cauchy and BEV distributions, $\hat{\sigma}_{(i)}$ for Normal and Morgenstern) shows the estimates to be reasonable to good. Note that the standard deviation of $\hat{\eta}_1$ and $\hat{\eta}_2$ on one hand and of $\hat{\eta}_3$ and $\hat{\eta}_4$ on the other hand are not fully comparable as m and k have a different meaning.

Some observations:

- Peng's estimator and ours are not stable at small k leading to missing values for either $\hat{\eta}$ or $\hat{\sigma}$ or both.
- The tests for asymptotic dependence tend to accept dependence for small $k$or m and to reject dependence for larger values. This effect is due to the increasing bias of the estimators for η, which is not taken into account by the tests. Consequently, the effect is weakest for the test based on $\hat{\eta}_1$.
- Hill's estimator has the smallest observed and estimated variances.
- Our estimator has a somewhat smaller observed variance than Peng's, but both have relatively large variance estimates for small k. Overall the ML estimator has a variance comparable to ours but for small k resp. m it is clearly smaller.

To conclude: the outcome of all tests for asymptotic dependence depend on k, the sample fraction used. The test based on the ML estimator $\hat{\eta}_1$ has the great advantage to be less dependent on k, but it is biased to rejecting dependence. Finally, due to the smaller variance of the Hill estimator the corresponding test detects even small deviations from the hypothesis, but on the other hand, due to its considerable bias, for the Cauchy and BEV distribution the hypothesis is much more often wrongly rejected than one would expect from the nominal level of the test. This disappointing behavior indicates that the approximation of the distribution of Hill estimator by a centered normal distribution is rather inaccurate for moderate sample sizes.

All estimators and tests would benefit from a guideline for choosing k.

4.3. **Failure probabilities.** Table 3 summarizes the main results for the failure probability estimators. The empirical distribution of the estimators for three values of k is shown in figure 4.

For the Cauchy distribution we have asymptotic dependence, so $\hat{p}_1$ is appropriate. Figure 4 shows that it is biased for small k, probably related with the negative bias of the γ estimates of the marginal distributions. As expected $\hat{p}_{\hat{\eta}}$ has larger variance; its smaller bias for small k is sort of a surprise.

For the normal distribution the main problem is estimating the marginals. The γ_1 and γ_2 estimates are negative. This implies upper bounds for the marginals and in quite a number of cases the failure area lies outside one or both of the bounds leading to a zero estimated failure probability.

In samples with the marginals transformed to exponential the estimator behaves much better. The marginals are estimated more accurately now with $\hat{\gamma}_i \approx 0$. Still when both $\hat{\gamma}_i$ estimates are negative a number of zero estimates result. The estimator $\hat{p}_1$ assuming asymptotic dependence over-estimates the probability, while $\hat{p}_\eta$ under-estimates it.

The Morgenstern distribution has asymptotically independent marginals. The $\hat{p}_{\hat{\eta}}$ estimate is nearly unbiased for $k = 80, 160$ whereas the $\hat{p}_1$ estimate is strongly biased.

Estimating the marginals does not cause problems here as the Morgenstern distribution has extreme value (Fréchet) marginals.

5. An application: dependence of sea state parameters

In the course of the Neptune project, financed by the European Union (grant MAS2-CT94-0081) the joint distribution of sea state variables was studied and its consequences for the seawall at Petten. The data set, supplied by the Dutch National Institute for Marine and Coastal management, consists of date, time and sea characteristics recorded from 1979 till 1991, at three-hourly intervals at the Eierland station, 20 kilometers off the Dutch coast. After a declustering routine a set of independent observations of waveheight `Hm0`, wave-periode `Tpb` and still water level `SWL` was constructed and analysed. de Haan and de Ronde [1998] concluded that the variables were asymptotically dependent, and estimated the failure probability of the Pettemer zeewering assuming asymptotic dependence between the variables. Figure 5 shows the joint distribution of pairs of these variables and illustrates the estimation of asymptotic dependence. For none of the pairs asymptotic dependence can be rejected although for quite a number of values of k the variances can not be calculated.

6. Proofs for section 2

The first results in this section closely follow Peng [1999]. We first state slightly rephrased versions of his Lemmas 2.1 and 2.2 concerning empirical probability measures. Define uniformly distributed random variables $U_i := 1 - F_1(X_i)$, $V_i := 1 - F_2(Y_i)$ and denote the order statistics by $U_{n,i}$ and $V_{n,i}$, with the convention $U_{n,0} = V_{n,0} = 0$.

We will use the following notation:

$$\begin{aligned} S_1(x,y) &:= \sum_{i=1}^{n} \mathbf{1}\{U_i \le x \text{ and } V_i \le y\}, \\ P_1(x,y) &:= \Pr\{U_1 \le x \text{ and } V_1 \le y\}, \\ S_2(x,y) &:= \sum_{i=1}^{n} \mathbf{1}\{U_i \le x \text{ or } V_i \le y\}, \\ P_2(x,y) &:= \Pr\{U_1 \le x \text{ or } V_1 \le y\}. \end{aligned} \tag{6.1}$$

Note that $S_2(x,y) = S_2(x,0) + S_2(0,y) - S_1(x,y)$, $P_2(x,y) = x + y - P_1(x,y)$, and $S_n(j,k)$ (equation (2.4)) equals $S_1(U_{n,j}, V_{n,k})$ a.s.

Lemma 6.1. *Assume* (2.6). *Let $r(n) = n\, q(k/n) \to \infty$ (which implies $k \to \infty$) and $k/n \to 0$. Then we have*

$$\sqrt{r(n)}\Big(\frac{S_1(\frac{k}{n}x, \frac{k}{n}y)}{r(n)} - \frac{P_1(\frac{k}{n}x, \frac{k}{n}y)}{q(k/n)}\Big) \xrightarrow{D} W_1(x,y).$$

Here, and below, $\xrightarrow{D}$ is convergence in $D([0,\infty)^2)$ and $W_1(x,y)$ is a Gaussian process with mean zero and covariance structure

$$\mathrm{E}\{W_1(x_1,y_1)W_1(x_2,y_2)\} = c(x_1 \wedge x_2, y_1 \wedge y_2).$$

Proof. See [Peng, 1999, Huang, 1992] and [Einmahl, 1997, Theorem 3.1]. □

Corollary 6.1. *Assume* (2.1). *If additionally the sequence $k(n)$ is such that $r(n) = n\,q(k/n) \to \infty$, $k/n \to 0$ and $\sqrt{r(n)}\,q_1(k/n) \to 0$ then*

$$\sqrt{r(n)}\Big(\frac{S_1(\frac{k}{n}x,\frac{k}{n}y)}{r(n)} - c(x,y)\Big) \xrightarrow{D} W_1(x,y).$$

Proof. The extra condition on the sequence $k(n)$ ensures that

$$\sqrt{r(n)}\Big(\frac{P_1(\frac{k}{n}x,\frac{k}{n}y)}{q(k/n)} - c(x,y)\Big) \to 0$$

uniformly on $[0,A]^2$ for any $A > 0$. □

Lemma 6.2. *Assume* (2.1). *Let $k \to \infty$ and $k/n \to 0$. Then we have*

$$\sqrt{k}\Big(\frac{S_2(\frac{k}{n}x,\frac{k}{n}y)}{k} - \frac{n}{k}P_2(\frac{k}{n}x,\frac{k}{n}y)\Big) \xrightarrow{D} W_2(x,y).$$

Here $W_2(x,y)$ is a Gaussian process with mean zero and covariance structure

$$\mathrm{E}\{W_2(x_1,y_1)W_2(x_2,y_2)\} = x_1 \wedge x_2 + y_1 \wedge y_2 - lc(x_1,y_1) - lc(x_2,y_2) + lc(x_1 \vee x_2, y_1 \vee y_2)$$

Proof. See [Peng, 1999, proof of Lemma 2.2] and [Einmahl, 1997, Theorem 3.1]. □

Corollary 6.2. *Assume* (2.1). *Let $k \to \infty$ and $k/n \to 0$. Then*

$$\sqrt{k}\Big(\frac{n}{k}U_{n,\lfloor kx\rfloor} - x\Big) \xrightarrow{D} -W_2(x,0)$$
$$\sqrt{k}\Big(\frac{n}{k}V_{n,\lfloor ky\rfloor} - y\Big) \xrightarrow{D} -W_2(0,y).$$

Proof. We will prove the first equation. Lemma 6.2 implies

$$\sqrt{k}\Big(\frac{1}{k}\sum_{i=1}^{n}\mathbf{1}\{U_i \le \frac{k}{n}x\} - x\Big) \xrightarrow{D} W_2(x,0).$$

Note that the generalized inverse of $x \mapsto 1/k\sum_{i=1}^{n}\mathbf{1}\{U_i \le k/nx\}$ equals $x \mapsto (n/k)U_{n,\lfloor kx\rfloor}$; applying Vervaat's Lemma [Vervaat, 1972] gives the result of the Corollary. □

Corollary 6.3. *Assume the conditions of Theorem 2.2. Then*

$$\sqrt{r(n)}\Big(\frac{S_1(U_{n,\lfloor kx\rfloor},V_{n,\lfloor ky\rfloor})}{r(n)} - c(x,y)\Big) \xrightarrow{D} W(x,y).$$

$W(x,y)$ is a Gaussian process with mean zero and covariance structure depending on l:
in case $l = 0$

$$W(x,y) = W_1(x,y);$$

in case $l > 0$

$$W(x,y) = \frac{1}{\sqrt{l}}\left(W_2(x,0) + W_2(0,y) - W_2(x,y)\right) - \sqrt{l}c_x(x,y)W_2(x,0) - \sqrt{l}c_y(x,y)W_2(0,y),$$

where the term in the first line of the right hand side has the same distribution as $W_1(x,y)$.

Proof. For $l = 0$ the result follows from Corollaries 6.2 and 6.1: we have $r(n) = o(k)$ and

$$\sqrt{r(n)}\Big(\frac{n}{k}U_{n,\lfloor kx\rfloor} - x\Big) \xrightarrow{p} 0$$
$$\sqrt{r(n)}\Big(\frac{n}{k}V_{n,\lfloor ky\rfloor} - y\Big) \xrightarrow{p} 0.$$

Otherwise, $r(n)/k \to l$ with $l > 0$. Write

$$S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor ky\rfloor}) = \lfloor kx\rfloor + \lfloor ky\rfloor - S_2(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor ky\rfloor})$$
$$P_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor ky\rfloor}) = U_{n,\lfloor kx\rfloor} + V_{n,\lfloor ky\rfloor} - P_2(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor ky\rfloor})$$

and the result follows from Lemma 6.2 and Corollary 6.2 (see Peng [1999]). □

Corollary 6.4. *Assume the conditions of Theorem 2.2. Then*

$$\sqrt{r(n)}\Big(\frac{S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor kx\rfloor})}{S_1(U_{n,k}, V_{n,k})} - x^{1/\eta}\Big) \xrightarrow{d} W(x,x) - x^{1/\eta}W(1,1) =: V(x).$$

Here $\xrightarrow{d}$ *is convergence in* $D[0,1]$. *The process* $V(x)$ *in this equation is Gaussian with mean zero and covariance depending on* η *and* l.

For $l = 0$

$$\mathrm{E}\{V(x)V(y)\} = (x\wedge y)^{1/\eta} - (xy)^{1/\eta}.$$

For $l > 0$

$$\begin{aligned}\mathrm{E}\{V(x)V(y)\} =& \big(1 - 2lc_x(1,1)c_y(1,1)\big)(x\wedge y - (1+l)xy)\\ &+ lc_x(1,1)(1 - lc_y(1,1))\big(yc(x,1) + xc(y,1) - c(x\wedge y, x\vee y)\big)\\ &+ lc_y(1,1)(1 - lc_x(1,1))\big(yc(1,x) + xc(1,y) - c(x\vee y, x\wedge y)\big).\end{aligned}$$

Proof. From Corollary 6.3 we have

$$\begin{aligned}\frac{S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor kx\rfloor})}{S_1(U_{n,k}, V_{n,k})} =& \frac{c(x,x) + r(n)^{-1/2}(W(x,x) + o_p(1))}{c(1,1) + r(n)^{-1/2}(W(1,1) + o_p(1))}\\ =& \frac{c(x,x)}{c(1,1)}\Big[1 + r(n)^{-1/2}\big(\frac{W(x,x)}{c(x,x)} - \frac{W(1,1)}{c(1,1)}\big)\\ &+ r(n)^{-1/2}\big(\frac{o_p(1)}{c(x,x)} - \frac{o_p(1)}{c(1,1)}\big)\Big]\\ =& x^{1/\eta} + r(n)^{-1/2}\frac{W(x,x) - x^{1/\eta}W(1,1)}{c(1,1)} + o_p(r(n)^{-1/2}).\end{aligned}$$

For the proof of the covariance formula in case of $l > 0$, note that then $c_x(1,1) + c_y(1,1) = 1$. □

Remark. For $l = 0$ the process $\{V(x^\eta)\}$ is just a Brownian bridge.

Proof of Theorem 2.1. From Corollary 6.1 we have

$$\lim_{n\to\infty} \frac{S_1(\frac{k}{n}x, \frac{k}{n}y)}{r(n)} = c(x,y)$$

uniformly on say $0 \le x, y \le 2$. Since

$$\frac{n}{k}U_{n,\lfloor kx\rfloor} \to x \text{ and } \frac{n}{k}V_{n,\lfloor ky\rfloor} \to y$$

uniformly on $0 \le x, y \le 2$ by Corollary 6.2,

$$\lim_{n\to\infty} \frac{S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor ky\rfloor})}{r(n)} = c(x,y)$$

uniformly. Hence $S_n(k,k)/r(n) = S_1(U_{n,k}, V_{n,k})/r(n) \to c(1,1)$ and

$$\frac{\frac{1}{k}\sum_{j=1}^{k} S_n(j,j)}{r(n)} = \frac{\int_0^1 S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor kx\rfloor})dx}{r(n)} \to \int_0^1 c(x,x)dx = \frac{1}{1+1/\eta}.$$

□

Proof of Theorem 2.2 (normality of $\hat\eta_4$). By convergence in $D[0,1]$ (Corollary 6.4)

$$\sqrt{r(n)}\Big(\int_0^1 \frac{S_1(U_{n,\lfloor kx\rfloor}, V_{n,\lfloor kx\rfloor})}{S_1(U_{n,k}, V_{n,k})} dx - \int_0^1 x^{1/\eta} dx\Big) \xrightarrow{d} \int_0^1 V(x)dx$$

or equivalently

$$\sqrt{r(n)}\Big(1/k \sum_{j=1}^{k} \big(\frac{S_1(U_{n,j}, V_{n,j})}{S_1(U_{n,k}, V_{n,k})}\big) - \frac{1}{1+1/\eta}\Big) \xrightarrow{d} \int_0^1 V(x)dx. \tag{6.2}$$

The distribution of $\int V(x)dx$ is normal with

$$E\left\{\int_0^1 V(x)dx\right\} = \int_0^1 \mathrm{E}\left\{V(x)\right\} dx = 0$$

and variance

$$\mathrm{E}\left\{\int_0^1 V(x)dx \int_0^1 V(y)dy\right\} = 2\int_0^1\int_0^y \mathrm{E}\left\{V(x)V(y)\right\} dxdy.$$

Using Corollary 6.4, this variance equals

$$\begin{aligned} &2\int_0^1\int_0^y x^{1/\eta}(1-y^{1/\eta})dxdy \\ &\quad = \frac{1/\eta}{(2+1/\eta)(1/\eta+1)^2}, \qquad \text{for } l = 0, \end{aligned}$$

or

$$\begin{aligned} &\frac{1}{3} l c_x(1,1)\big(1 - l c_y(1,1)\big) \int_0^1 c(u,1)du \\ &\quad + \frac{1}{3} l c_y(1,1)\big(1 - l c_x(1,1)\big) \int_0^1 c(1,u)du \\ &\quad + \frac{1}{2} l^2 c_x(1,1) c_y(1,1) - \frac{1}{3} l c_x(1,1) - \frac{1}{3} l c_y(1,1) + \frac{1}{12} \\ &\quad + \frac{1}{12} l (c_x(1,1)^2 + c_y(1,1)^2), \qquad \text{for } l > 0. \end{aligned}$$

Finally

$$\hat\eta_4 - \eta = (1+\eta)^2\big(\frac{1}{1+1/\hat\eta_4} - \frac{1}{1+1/\eta}\big)(1+o(1)).$$

This proves Theorem 2.2 for $\hat\eta_4$. □

Proof of Corollary 2.1 (for $\hat\eta_4$***).*** By Corollary 6.3 we have

$$r(n)^{-1}S_1(U_{n,k},V_{n,k}) \xrightarrow{p} 1;$$

So $S_n(k,k) = S_1(U_{n,k},V_{n,k})$ a.s. is a consistent estimator of $r(n)$ in the Theorem. This proves Corollary 2.1. □

Remark. It is worth mentioning that the asymptotic normality of $\hat\eta_3$ can be derived from Corollary 6.3 in a similar way.

Now we turn to the Ledford and Tawn - type estimators $\hat\eta_1$ and $\hat\eta_2$.

Let $m_n = \lfloor r(n) \rfloor$ and denote by Q_n the tail empirical quantile function pertaining to $T_i^{(n)}$, $1 \le i \le n$, i.e.

$$Q_n(t) := T^{(n)}_{n,n-\lfloor m_n t\rfloor}, \qquad 0 < t < n/m_n.$$

The following lemma is central to the proof of the asymptotic normality of estimators for η based on largest order statistics of $T_i^{(n)}$.

Lemma 6.3. *Under the conditions of Theorem 2.2 there exist suitable versions of* Q_n*, a suitable process* $\bar W$ *equal in distribution to a standard Brownian motion if* $l = 0$ *and to* $x \mapsto W(x,x)$ *if* $l > 0$ *such that for all* $t_0, \varepsilon > 0$

$$\sup_{0<t\le t_0} t^{\eta+1/2+\varepsilon}\Big| m_n^{1/2}\Big(\frac{k}{n}Q_n(t) - t^{-\eta}\Big) - \eta t^{-(\eta+1)}\bar W(t)\Big| = o_P(1).$$

Proof. First check that

$$\begin{aligned}\sum_{i=1}^n 1\{T_i^{(n)} > x\} &= \sum_{i=1}^n 1\{R_i^X > (n+1)(1-1/x) \text{ and } R_i^Y > (n+1)(1-1/x)\}\\ &= \sum_{i=1}^n 1\{U_i < U_{n,\lceil (n+1)/x\rceil} \text{ and } V_i < V_{n,\lceil (n+1)/x\rceil}\} \qquad \text{a.s.}\end{aligned}$$

with the convention $U_{n,n+1} = V_{n,n+1} = 1$. Hence

$$\bar F_n(x) := \frac{1}{n}\sum_{i=1}^n 1\{\tfrac{k}{n+1}T_i^{(n)} > x\} = \frac{1}{n}S_1\Big(U_{n,\lceil k/x\rceil}-, V_{n,\lceil k/x\rceil}-\Big)$$

where $f(x-)$ denotes the left-hand limit of f at x. From Corollary 6.1 one readily obtains that

$$\begin{aligned} & m_n^{1/2}\Big(\frac{\bar F_n(x)}{q(k/n)} - x^{-1/\eta}\Big)_{0<x<\infty} \longrightarrow \Big(W(1/x,1/x)\Big)_{0<x<\infty}\\ \Longrightarrow\ & m_n^{1/2}\Big(\frac{\bar F_n(x^{-\eta})}{q(k/n)} - x\Big)_{0<x<\infty} \longrightarrow \Big(W(x^\eta,x^\eta)\Big)_{0<x<\infty} =: \bar W\\ \Longrightarrow\ & m_n^{1/2}\Big(\Big(\bar F_n^{-1}(q(k/n)t)\Big)^{-1/\eta} - t\Big)_{0<t<\infty} \longrightarrow -\bar W\end{aligned}$$

weakly in $D(0,\infty)$, where in the last step Vervaat's lemma has been used. For this, note that $\bar W$ has a.s. continuous sample paths, because by the definition of W it is a Brownian motion for $l = 0$ and it can be represented as a sum of Brownian motions if $l > 0$.

Consequently for suitable versions

$$\Big(\bar F_n^{-1}(q(k/n)t)\Big)^{-1/\eta} = t - m_n^{-1/2}\bar W(t) + o(m_n^{-1/2})$$

a.s. uniformly on compact intervals bounded away from 0. The δ–method yields

$$F_n^{-1}(q(k/n)t) = t^{-\eta}\Big(1 + m_n^{-1/2}\eta t^{-1}\bar{W}(t) + o(m_n^{-1/2})\Big)$$

uniformly in the same sense. Check that $F_n^{-1}(q(k/n)t) = k/(n+1)Q_n(r(n)t/m_n) = k/nQ_n(t) + O(1/m_n)$ uniformly and $\sup_{0<t\le\vartheta} t^{-1/2+\varepsilon}|\bar{W}(t)| = o_P(1)$ as $\vartheta \downarrow 0$ by the law of the iterated logarithm and the aforementioned representation of $\bar{W}$. Thus it remains to prove that for all $\delta > 0$

$$\lim_{\vartheta\downarrow 0}\limsup_{n\to\infty} \Pr\Big\{ \sup_{0<t\le\vartheta} m_n^{1/2} t^{\eta+1/2+\varepsilon}\Big|\frac{k}{n+1}Q_n(t) - t^{-\eta}\Big| > \delta\Big\} = 0. \tag{6.3}$$

For this, we restrict ourselves to considering

$$\begin{aligned}
&\Pr\Big\{ \sup_{0<t\le\vartheta} m_n^{1/2} t^{\eta+1/2+\varepsilon}\Big(\frac{k}{n+1}Q_n(t) - t^{-\eta}\Big) > \delta\Big\} \\
&\quad\le \Pr\Big\{\exists\, 1 \le i \le m_n\vartheta + 1 : \frac{k}{n+1}T^{(n)}_{n,n-i+1} > x_{i,n}\Big\} \\
&\quad= \Pr\Big\{\exists\, 1 \le i \le m_n\vartheta + 1 : \frac{k}{n+1}T^{(n)}_{n,n-i+1} > x_{i,n} \text{ and } x_{i,n} < k\Big\} \qquad (6.4)
\end{aligned}$$

with

$$x_{i,n} := \Big(\frac{i}{m_n}\Big)^{-\eta} + \delta m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(\eta+1/2+\varepsilon)}.$$

(The other inequality can be treated in a similar way.)

Let $A_i := 1/U_i$, $B_i := 1/V_i$ and

$$\tilde{S}_1(x,y) := \sum_{i=1}^n 1_{\{A_i > x \text{ and } B_i > y\}} = S_1(1/x-, 1/y-).$$

Then the right-hand side of (6.4) equals

$$\Pr\Big\{\exists\, 1 \le i \le m_n\vartheta + 1 : \tilde{S}_1(A_{n,n-\lceil k/x_{i,n}\rceil+1}, B_{n,n-\lceil k/x_{i,n}\rceil+1}) \ge i \text{ and } x_{i,n} < k\Big\}.$$

Now we distinguish two different ranges of i-values.

Case 1: $i <= i_n := \lfloor(\delta m_n^\varepsilon/L)^{1/(1/2+\varepsilon)}\rfloor$, $x_{i,n} < k$

According to Shorack and Wellner [1986, Theorem 10.3.1], for all $\bar{\varepsilon} > 0$ there exists $\bar{\delta} > 0$ such that

$$\limsup_{n\to\infty} \Pr\{\exists\, 2 \le j \le m_n + 1 : \frac{j-1}{n}A_{n,n-j+1} < \bar{\delta}\} \le \bar{\varepsilon},$$

and likewise for $B_{n,n-i+1}$. Thus

$$\begin{aligned}
&\Pr\Big\{\exists\, 1 \le i \le i_n : \tilde{S}_1(A_{n,n-\lceil k/x_{i,n}\rceil+1}, B_{n,n-\lceil k/x_{i,n}\rceil+1}) \ge i \text{ and } x_{i,n} < k\Big\} \\
&\quad\le \Pr\Big\{\exists\, 1 \le i \le i_n : \tilde{S}_1(x_{i,n}\bar{\delta}n/k, x_{i,n}\bar{\delta}n/k) \ge i\Big\} + 2\bar{\varepsilon}.
\end{aligned}$$

Check that

$$\frac{n}{k}x_{i,n}\bar{\delta} \;\ge\; \delta\frac{n}{k}m_n^{-1/2}\Big(\frac{i}{m_n}\Big)^{-(\eta+1/2+\varepsilon)}\bar{\delta} \;\ge\; \bar{\delta}Lk^{-1}m_n^{\eta}n^{1-\eta}(n/i)^{\eta}. \tag{6.5}$$

Denote by F_T the d.f. of $T_i := \min(A_i, B_i)$, i.e. $1 - F_T(x) = P_1(1/x, 1/x)$, so that F_T^{-1} is $(-\eta)$–varying at 1.

In case of $\eta < 1$, we have $k = o(m_n^{\eta+\iota})$ and $F_T^{-1}(1-t) = o(t^{-(\eta+\iota)})$ as $t \downarrow 0$ for all $\iota > 0$, so that the right-hand side of (6.5) is of larger order than $F_T^{-1}(1 - 2i/(\bar{\delta} L n))$, provided $\iota < (1-\eta)/2$.

If $\eta = 1$, then one can show that, in analogy to Lemma 2.1 of Drees [1998b],

$$\sup_{x \le 1} x^{\iota-1} \Big| \frac{P_1(tx, tx)}{P_1(t,t)} - x \Big| = o(q_1(t)).$$

Apply this bound with $t = k/n$ and $x = i/(\bar{\delta} L m_n)$ to obtain $1 - F_T(x_{i,n}\bar{\delta} n/k) \le 2i/(\bar{\delta} L n)$, since $P_1(k/n, k/n) \sim q(k/n) \sim m_n/n$ and $(i/m_n)^{1-\iota} q_1(k/n) = o(m_n^{1/2} q_1(k/n) i/m_n) = o(i/m_n)$ uniformly for $1 \le i \le i_n$.

Hence it follows that

$$\begin{aligned}
&\limsup_{n\to\infty} \Pr\Big\{ \exists\, 1 \le i \le i_n : \tilde{S}_1(x_{i,n}\bar{\delta} n/k, x_{i,n}\bar{\delta} n/k) \ge i \Big\} \\
&\le \limsup_{n\to\infty} \Pr\Big\{ \exists\, 1 \le i \le i_n : T_{n,n-i+1} > \frac{n}{k} x_{i,n} \bar{\delta} \Big\} \\
&\le \limsup_{n\to\infty} \Pr\Big\{ \max_{1 \le i \le m_n+1} \frac{T_{n,n-i+1}}{F_T^{-1}(1 - 2i/(\bar{\delta} L n))} > 1 \Big\} \\
&< \bar{\varepsilon}
\end{aligned}$$

for sufficiently large L, where for the last step again Theorem 10.3.1 of Shorack and Wellner [1986] has been used.

Case 2: $i_n < i \le m_n \vartheta + 1$

In this case we use the convergence

$$\lim_{\vartheta \downarrow 0} \limsup_{n\to\infty} \Pr\Big\{ \sup_{0 < t \le \vartheta} k^{1/2} t^{3/2+\iota} \Big| \frac{k}{n} A_{n,n-\lceil kt \rceil+1} - t^{-1} \Big| > \tilde{\delta} \Big\} = 0$$

for all $\tilde{\delta}, \iota > 0$, which is immediate from Theorem 2.1 of Drees [1998b]. By arguments similar to the ones given above, it suffices to consider $\Pr\{\exists\, i_n < i \le m_n\vartheta + 1 : \tilde{S}_1(y_{i,n}, y_{i,n}) \ge i\}$ with

$$\begin{aligned}
y_{i,n} &:= \frac{n}{k} x_{i,n} - \tilde{\delta} n k^{-3/2} x_{i,n}^{3/2+\iota} \\
&= \frac{n}{k} \Big(\frac{i}{m_n}\Big)^{-\eta} \Big[1 + \delta m_n^{-1/2} \Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)} - \tilde{\delta} k^{-1/2} \Big(\frac{i}{m_n}\Big)^{-\eta(1/2+\iota)} \times \\
&\qquad \times \Big(1 + \delta m_n^{-1/2} \Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)} \Big)^{3/2+\iota} \Big] \\
&\ge \frac{n}{k} \Big(\frac{i}{m_n}\Big)^{-\eta} \Big[1 + \frac{\delta}{2} m_n^{-1/2} \Big(\frac{i}{m_n}\Big)^{-(1/2+\varepsilon)} \Big]
\end{aligned}$$

for $\iota < \varepsilon$ and $\tilde{\delta} \le \delta(1+L)^{-(3/2+\iota)}/2$, since $k \ge m_n$ and $\eta \le 1$. Therefore

$$\begin{aligned}
&\lim_{\vartheta \downarrow 0} \limsup_{n\to\infty} \Pr\{ \exists\, i_n < i \le m_n \vartheta + 1 : \tilde{S}_1(y_{i,n}, y_{i,n}) \ge i \} \\
&\le \lim_{\vartheta \downarrow 0} \limsup_{n\to\infty} \Pr\Big\{ \exists\, i_n < i \le m_n\vartheta + 1 : \\
&\qquad m_n^{1/2} \Big(\frac{i}{m_n}\Big)^{\eta+1/2+\varepsilon} \Big(\frac{k}{n} T_{n,n-i+1} - \Big(\frac{i}{m_n}\Big)^{-\eta} \Big) > \delta/2 \Big\} \\
&= 0,
\end{aligned}$$

again by Theorem 2.1 of Drees [1998b], where (2.1) implies Condition 1 of that paper and $m_n^{1/2} q_1(k/n) \to 0$ ensures that the bias is asymptotically negligible.

Combining both cases one arrives at (6.3). □

Theorem 2.2, (asymptotic normality of $\hat{\eta}_1$ ***and*** $\hat{\eta}_2$***).*** Note that this approximation is analogous to the approximation of the tail empirical quantile function established in Drees [1998b] in the classical situation of i.i.d. random variables. Hence the asymptotic normality of $\hat{\eta}_1$ and $\hat{\eta}_2$ follows from Lemma 6.3 exactly as in [Drees, 1998b, Example 4.1] and [Drees, 1998a, Example 3.1] using the δ–method. The asymptotic variance is given by

$$\int_0^1 \int_0^1 \operatorname{Cov}(\bar{W}(s), \bar{W}(t))(st)^{-(\eta+1)}\, \nu_\eta(ds)\, \nu_\eta(dt)$$

with $\nu_\eta(dt) := (\eta+1)^2(t^\eta - (2\eta+1)t^{2\eta})/\eta\, dt + (\eta+1)\varepsilon_1(dt)$ for the maximum likelihood estimator $\hat{\eta}_1$ and $\nu_\eta(dt) := \eta(t^\eta dt - \varepsilon_1(dt))$ in case of the Hill estimator. (Here ε_1 denotes the Dirac measure at 1.) Now using the homogeneity of order 1 of the covariance function which implies $\int_0^t \operatorname{Cov}(\bar{W}(s), \bar{W}(t))(st)^{-1}\, ds = \int_0^1 \operatorname{Cov}(\bar{W}(u), \bar{W}(1))u^{-1}\, du$, one obtains $(\eta+1)^2 \operatorname{Var}(\bar{W}(1))$ and $\eta^2 \operatorname{Var}(\bar{W}(1))$, respectively, as asymptotic variance and thus the assertion, using $c_x(1,1) + c_y(1,1) = 1/\eta$. □

Proof of Theorem 2.3. Note that according to Corollary 6.3

$$\frac{S_n(i,j)}{r(n)} = c\Big(\frac{i}{k}, \frac{j}{k}\Big) + O_P\Big((r(n))^{-1/2}\Big)$$

uniformly for $1 \le i, j \le 2k$. Hence $\hat{l} \to l$ in probability by the definition of $r(n)$. Moreover,

$$\begin{aligned}
\hat{c}_x(1,1) &= k^{1/4} \frac{c(\lfloor k(1+k^{-1/4})\rfloor/k, 1) - c(1,1) + O_P((r(n)^{-1/2}))}{1 + O_P((r(n)^{-1/2}))} \\
&= c_x(1,1) + O_P\Big(k^{1/4}(r(n))^{-1/2}\Big) \\
&\stackrel{P}{\to} c_x(1,1)
\end{aligned}$$

if $\eta > 1/2$. The consistency of $\hat{c}_y(1,1), \hat{d}_1$ and $\hat{d}_2$ can be proved in a similar way, so that the consistency of $\hat{\sigma}_i^2$ follows readily in that case.

In case of $\eta \le 1/2$, we have

$$\hat{l}^{1/2}\hat{c}_x(1,1) = O_P\Big((r(n)/k)^{1/2}(1 + k^{1/4}(r(n))^{-1/2})\Big) = o_P(1)$$

and likewise $\hat{l}^{1/2}(\hat{c}_y(1,1) + \hat{d}_1 + \hat{d}_2) \to 0$ in probability. Thus the consistency of $\hat{\sigma}_i$ is obvious because of $l = 0$.

Assertion (ii) follows similarly from

$$\frac{k}{n} T^{(n,u)}_{n,n-\lfloor m_n t\rfloor} = \left(\frac{t}{c(1+u,1)}\right)^{-\eta} + O_P(m_n^{-1/2})$$

which in turn can be verified using the same arguments as in the proof of Lemma 6.3. □

7. Proof of Theorem 3.1

The proof of Theorem 3.1 will be established in several steps. The following sequence of equalities and asymptotic in probability equivalences provides an overview over the reasoning:

$$
\begin{aligned}
p_n &= \Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(C_n)\}\\
&\overset{(3.13)}{\sim} q\Big(\frac{k}{n}\Big)\nu\Big(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(C_n))\Big)\\
&\overset{\text{Lemma 7.2}}{\sim} q\Big(\frac{k}{n}\Big)\nu(D_n)\\
&\overset{(2.2)}{=} c_n^{1/\eta} q\Big(\frac{k}{n}\Big)\nu\Big(\frac{D_n}{c_n}\Big)\\
&\overset{\text{Cor. 7.3}}{\sim} c_n^{1/\eta} q\Big(\frac{k}{n}\Big)\nu\Big(\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\hat{D}_n}{c_n}\big)\big)\Big)\\
&\overset{\text{Lemma 7.3}}{\sim} c_n^{1/\eta} q\Big(\frac{k}{n}\Big)\nu\Big(\frac{n}{k}\big(\mathbf{1}-\boldsymbol{F}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\hat{D}_n}{c_n}\big)\big)\big)\Big)\\
&\overset{(3.13)}{\sim} c_n^{1/\eta}\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}_{|B=\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\frac{\hat{D}_n}{c_n})}\\
&\overset{\text{Lemma 7.4}}{\sim} c_n^{1/\eta}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\Big\{(X_i,Y_i)\in\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\hat{D}_n}{c_n}\big)\Big\}\\
&\sim \hat{p}_n. \qquad (7.1)
\end{aligned}
$$

Lemma 7.1. *Let $a=a(n),\tilde{a}>0$, $b,\tilde{b},\gamma,\tilde{\gamma}\in\mathbb{R}$ denote sequences such that*

$$\Big|\frac{\tilde{a}}{a}-1\Big|\vee\Big|\frac{\tilde{b}-b}{a}\Big|\vee|\tilde{\gamma}-\gamma|=O(\epsilon_n)$$

for some $\epsilon_n\downarrow 0$. Suppose that the sequence $\lambda_n>0$ is bounded and satisfies $\epsilon_n\log\lambda_n\to 0$ and $\epsilon_n w_\gamma(\lambda_n)\to 0$ with w_γ defined in (3.11). *Then*

$$1-F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1-x))=x+o(\lambda_n) \qquad (7.2)$$

uniformly for $0\le x\le\lambda_n$.

Proof. First check that

$$T(x):=1-F_{\tilde{a},\tilde{b},\tilde{\gamma}}(F^{-1}_{a,b,\gamma}(1-x))=\Big[1+\tilde{\gamma}\frac{a}{\tilde{a}}\Big(\frac{x^{-\gamma}-1}{\gamma}+\frac{b-\tilde{b}}{a}\Big)\Big]^{-1/\tilde{\gamma}},$$

where, as usual, $(x^{-\gamma}-1)/\gamma:=-\log x$ if $\gamma=0$. Now we distinguish three cases.

$\boldsymbol{\gamma>0}$ **:** Then

$$
\begin{aligned}
T(x) &= \big(1+(1+O(\epsilon_n))(x^{-\gamma}-1+O(\epsilon_n))\big)^{-(1+O(\epsilon_n))/\gamma}\\
&= \big(x^{-\gamma}(1+O(\epsilon_n))+O(\epsilon_n)\big)^{-(1+O(\epsilon_n))/\gamma}\\
&= x\exp(O(\epsilon_n)\log x)(1+o(1)).
\end{aligned}
$$

For $\lambda_n\epsilon_n\le x\le\lambda_n$

$$|\log x|\epsilon_n\le\big(|\log\lambda_n|+|\log\epsilon_n|\big)\epsilon_n\ \to\ 0,$$

so that $T(x)=x(1+o(1))=x+o(\lambda_n)$ uniformly.

Otherwise, i.e. for $0 < x < \lambda_n \epsilon_n$,

$$T(x) \le T(\lambda_n\epsilon_n) = \lambda_n\epsilon_n(1+o(1)) = o(\lambda_n) = x + o(\lambda_n)$$

by the monotonicity of T.

$\boldsymbol{\gamma < 0}$ **:** Choose $\delta_n \to 0$ such that $\epsilon_n(\lambda_n\delta_n)^\gamma \to 0$ and hence also $\epsilon_n \log \delta_n \to 0$ (e.g. $\delta_n = (\epsilon_n \lambda_n^\gamma)^{-1/(2\gamma)}$). Then uniformly for $\lambda_n\delta_n \le x \le \lambda_n$

$$T(x) = x^{1+O(\epsilon_n)}\big(1 + O(\epsilon_n) + O(\epsilon_n(\lambda_n\delta_n)^\gamma)\big)^{-(1+O(\epsilon_n))/\gamma} = x(1+o(1))$$

and again (7.2) follows from the monotonicity of T.

$\boldsymbol{\gamma = 0}$ **:** Note that $\tilde\gamma|\log x| \to 0$ uniformly for $\lambda_n\epsilon_n \le x \le \lambda_n$. Hence a Taylor expansion of log yields

$$\begin{aligned} T(x) &= \exp\Big(-\frac{1}{\tilde\gamma}\log\big(1 + \tilde\gamma(1+O(\epsilon_n))(-\log x + O(\epsilon_n))\big)\Big) \\ &= \exp\Big(-\frac{1}{\tilde\gamma}\big[\tilde\gamma(1+O(\epsilon_n)(-\log x + O(\epsilon_n)) + O(\tilde\gamma^2(\log x + O(\epsilon_n))^2)\big]\Big) \\ &= x\exp\big(O(\epsilon_n)\log x + O(\epsilon_n) + O(\epsilon_n \log^2 x)\big) \\ &= x(1+o(1)) \end{aligned}$$

and thus the assertion by the aforementioned arguments. □

Remark 7.1. For fixed sequences a, b and γ, assertion (7.2) even holds true uniformly for

$$(7.3)\qquad (\tilde a, \tilde b, \tilde\gamma) \in M(\epsilon_n) := \Big\{(\bar a, \bar b, \bar\gamma) \in (0,\infty)\times\mathbb{R}^2 \;\Big|\; \big|\frac{\bar a}{a} - 1\big| \vee \big|\frac{\bar b - b}{a}\big| \vee |\bar\gamma - \gamma| \le \epsilon_n\Big\}.$$

Corollary 7.1. *If condition (D),* (3.7) *and* (3.10)–(3.11) *are satisfied then, for all* $\delta > 0$,

$$\Pr\big\{A_{-\delta} \subset \frac{\hat D_n}{d_n} \subset A_{+\delta}\big\} \;\to\; 1.$$

Proof. Since the set A is bounded, there exists $L > 0$ such that $D_n \subset [0, d_n L]^2$ for all sufficiently large n. Because of (3.11), one can find a sequence $\epsilon_n \to 0$ such that $k^{-1/2} = o(\epsilon_n)$ and the conditions of Lemma 7.1 hold for $\lambda_n = d_n L$. Then $\Pr\{(\hat{\boldsymbol a}, \hat{\boldsymbol b}, \hat{\boldsymbol\gamma}) \in (M(\epsilon_n))^2\} \to 1$ with $M(\epsilon_n)$ defined in (7.3) and Lemma 7.1 yields

$$(7.4)\qquad \sup_{(x,y)\in D_n} \|\mathbf{1} - \boldsymbol F_{\hat{\boldsymbol a},\hat{\boldsymbol b},\hat{\boldsymbol\gamma}}(\boldsymbol F^{-1}_{\boldsymbol a,\boldsymbol b,\boldsymbol\gamma}(\mathbf{1} - (x,y))) - (x,y)\| \le \frac{\delta}{2} d_n$$

with probability tending to 1. Thus, in view of $\hat D_n = \mathbf{1} - \boldsymbol F_{\hat{\boldsymbol a},\hat{\boldsymbol b},\hat{\boldsymbol\gamma}}(\boldsymbol F^{-1}_{\boldsymbol a,\boldsymbol b,\boldsymbol\gamma}(\mathbf{1} - D_n))$ and condition (D),

$$\Pr\big\{\frac{\hat D_n}{d_n} \subset \big(\frac{D_n}{d_n}\big)_{+\delta/2} \subset A_{+\delta}\big\} \to 1.$$

On the other hand, by the definition of the inner neighborhood of a set, $(x,y) \in (D_n/d_n)_{-\delta/2}$ implies $(x + \delta/2, y + \delta/2) \in D_n/d_n$. Since, in view of (7.4),

$$d_n(x,y) \le \mathbf{1} - \boldsymbol F_{\hat{\boldsymbol a},\hat{\boldsymbol b},\hat{\boldsymbol\gamma}}\Big(\boldsymbol F^{-1}_{\boldsymbol a,\boldsymbol b,\boldsymbol\gamma}\Big(\mathbf{1} - d_n\big(x + \frac{\delta}{2}, y + \frac{\delta}{2}\big)\Big)\Big)$$

componentwise, (3.7) shows that indeed $d_n(x,y) \in \hat D_n$. Hence, again by condition (D),

$$\Pr\big\{A_{-\delta} \subset \big(\frac{D_n}{d_n}\big)_{-\delta/2} \subset \frac{\hat D_n}{d_n}\big\} \to 1.$$

□

Corollary 7.2. *If the conditions of Corollary 7.1 hold and, in addition,* (3.12) *then, for all* $\delta > 0$,

$$\Pr\Big\{A_{-\delta} \subset \frac{c_n}{d_n}\big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big) \subset A_{+\delta}\Big\} \to 1.$$

Proof. According to Corollary 7.1, there exists $L > 0$ such that $\Pr\{\hat{D}_n/c_n \subset [0, \lambda_n]^2\} \to 1$ for $\lambda_n := Ld_n/c_n$. It follows from (3.10) and (3.12) that $\lambda_n^{\hat{\gamma}_i} = \lambda_n^{\gamma_i}(1 + o_P(1))$, $i = 1, 2$. Hence one may apply Lemma 7.1 with $(a, b, \gamma) = (\hat{a}_i, \hat{b}_i, \hat{\gamma}_i)$ and $(\tilde{a}, \tilde{b}, \tilde{\gamma}) = (a_i, b_i, \gamma_i)$ to obtain

$$\sup_{(x,y)\in \hat{D}_n/c_n} \big\|\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1} - (x, y))) - (x, y)\big\| \le \frac{\delta}{2}\frac{d_n}{c_n}$$

with probability tending to 1 for all $\delta > 0$. Now one may conclude the proof following the lines of the preceding proof. □

Corollary 7.3. *Under the conditions of Corollary 7.2*

$$\nu\Big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\Big) = \nu\big(\frac{D_n}{c_n}\big)(1 + o_P(1)).$$

Proof. Denote the boundary of the set A by ∂A. Condition (3.7) implies a slightly weaker version for A, namely $(x, y) \in A \Rightarrow [0, x) \times [0, y) \subset A$. Hence $\lambda \cdot \partial A \subset \partial A$ for all $\lambda \in (0, 1)$ and these sets are pairwise disjoint. Since ν is homogeneous in the sense of (2.2) and $\nu(A) < \infty$ by the boundedness of A, it follows that $\nu(\partial A) = 0$. Moreover, $A_{+\delta} \setminus A_{-\delta} \downarrow \partial A$ as $\delta \downarrow 0$, so that $\nu(A_{+\delta} \setminus A_{-\delta}) \to 0$. Thus Corollary 7.2 and condition (D) yield

$$\nu\Big(\frac{c_n}{d_n}\big(\mathbf{1} - \boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}\big(\boldsymbol{F}^{-1}_{\hat{\boldsymbol{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1} - \frac{\hat{D}_n}{c_n}\big)\big)\big)\Big) \to \nu(A)$$

and $\nu(D_n/d_n) \to \nu(A)$. Now the assertion is an obvious consequence of the homogeneity (2.2). □

Lemma 7.2. *If condition (D),* (3.7) *and* (3.8) *hold, then*

$$\nu(D_n) = \nu\big(\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))\big)(1 + o(1)).$$

Proof. There exists $L > 0$ such that $D_n \subset [0, d_nL]^2$ for all sufficiently large n. Choose arbitrary $-1/(\gamma_i \vee 0) < x_i < 1/(-\gamma_i \vee 0)$, $i = 1, 2$. Then, by (3.8), for all $(x, y) \in D_n$

$$\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(\boldsymbol{F}^{-1}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(\mathbf{1} - (x, y)))) = (x(1 + \delta_x), y(1 + \delta_y)) \tag{7.5}$$

with $|\delta_x| \vee |\delta_y| \le R_{x_1,x_2}(n/k)$ for sufficiently large n. According to (3.7), the left-hand side of (7.5) is an element of $D_n(1 + R_{x_1,x_2}(n/k))$. Thus, by the definition of D_n,

$$\frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n)) \subset D_n\big(1 + R_{x_1,x_2}\big(\frac{n}{k}\big)\big).$$

Likewise, (7.5) together with (3.7) implies

$$D_n\big(1 - R_{x_1,x_2}\big(\frac{n}{k}\big)\big) \subset \frac{n}{k}(\mathbf{1} - \boldsymbol{F}(C_n))$$

eventually. Now the assertion is obvious from the homogeneity property (2.2). □

Lemma 7.3. *Under the conditions (D),* (3.7)*,* (3.8) *and* (3.10)–(3.12) *one has*

$$\nu\Big(\big(\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\gamma}\big(\boldsymbol{F}^{-\mathbf{1}}_{\hat{\mathbf{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\hat{D}_n}{c_n}\big)\big)\big)\Big)=\nu\Big(\frac{n}{k}\big(\mathbf{1}-\boldsymbol{F}\big(\boldsymbol{F}^{-\mathbf{1}}_{\hat{\mathbf{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\hat{D}_n}{c_n}\big)\big)\big)\Big)(1+o_P(1)).$$

Proof. The proof is very much the same as that for Lemma 7.2 with D_n replaced by $\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\gamma}(\boldsymbol{F}^{-\mathbf{1}}_{\hat{\mathbf{a}},\hat{\boldsymbol{b}},\hat{\boldsymbol{\gamma}}}(\mathbf{1}-\hat{D}_n/c_n))$. For this note that, by the boundedness of d_n/c_n and the assertion of Corollary 7.2, this set is eventually bounded. Hence (3.8) is applicable for sufficiently small x_1 and x_2. □

Lemma 7.4. *If the conditions of Theorem 3.1 are satisfied, then*

$$\sup_{B\in\mathcal{B}_n}\Big|\frac{\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{\mathbf{1}-\boldsymbol{F}(X_i,Y_i)\in\mathbf{1}-\boldsymbol{F}(B)\}}{\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}}-1\Big|\ \to\ 0\quad \textit{in probability.}$$

Proof. We will apply Theorem 5.1 of [Alexander, 1987]. To check the conditions of this uniform law of large numbers, first note that every set $B\in\mathcal{B}_n$ can be represented as

$$B=\boldsymbol{F}^{-\mathbf{1}}_{\tilde{\mathbf{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}\big(\mathbf{1}-\frac{\mathbf{1}-\boldsymbol{F}_{\tilde{\mathbf{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}}}(C_n)}{c_n}\big)\tag{7.6}$$

with $(\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\gamma}})\in(M(\epsilon_n))^2$ (cf. (7.3)). Therefore the arguments of the proofs for Lemma 7.3 and Corollary 7.3 show that

(7.7)
$$\nu\big(\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(B))\big)=\nu\big(\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(B)\big)(1+o(1))=\nu\big(\frac{D_n}{c_n}\big)(1+o(1))=\big(\frac{d_n}{c_n}\big)^{1/\eta}\nu(A)(1+o(1))$$

uniformly for $B\in\mathcal{B}_n$ (cf. Remark 7.1). Now (3.13) leads to

$$\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}=q\big(\frac{k}{n}\big)\big(\frac{d_n}{c_n}\big)^{1/\eta}\nu(A)(1+o(1))\tag{7.8}$$

uniformly. In particular, there exists n_0 such that $\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}<1/2$ for all $n\geq n_0$ and all $B\in\mathcal{B}_n$.

Next note that

$$\begin{aligned}\bar{B}_t\ &:=\bigcup_{B\in\mathcal{B}_n,n\geq n_0,\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}(1-\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\})\leq t}B\\ &\subset\bigcup_{B\in\mathcal{B}_n,n\geq n_0,\Pr\{\mathbf{1}-\boldsymbol{F}(X,Y)\in\mathbf{1}-\boldsymbol{F}(B)\}\leq 2t}B.\end{aligned}\tag{7.9}$$

In view of (7.6), one may prove as in Corollary 7.2 that, for all $\delta>0$, eventually $\mathbf{1}-\boldsymbol{F}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{\gamma}}(B)\subset A_{+\delta}d_n/c_n$ for all $B\in\mathcal{B}_n$. Hence it follows as in the proof of Lemma 7.2 that

$$\frac{n}{k}(\mathbf{1}-\boldsymbol{F}(B))\subset\frac{d_n}{c_n}A_{+\delta}(1+o(1))\tag{7.10}$$

uniformly for $B\in\mathcal{B}_n$.

Let $n(t) := \min\big\{n \geq n_0 \,|\, q(k/n)(d_n/c_n)^{1/\eta}\nu(A) \leq 3t\big\}$, which tends to ∞ as t tends to 0. Combining (7.8)–(7.10), we arrive at

$$\begin{aligned} \mathbf{1} - \boldsymbol{F}(\bar{B}_t) &\subset \bigcup_{n \geq n(t)} \frac{k(n)d_n}{nc_n} A_{+\delta}(1 + o(1)) \\ &\subset 2 \sup_{n \geq n(t)} \frac{k(n)d_n}{nc_n} A_{+\delta} \end{aligned}$$

for sufficiently small t. By (3.13), the regularity condition on $k(n)$ and the definition of $n(t)$ it follows that

$$\begin{aligned} \Pr\{\mathbf{1} - \boldsymbol{F}(X, Y) \in \mathbf{1} - \boldsymbol{F}(\bar{B}_t)\} &= O\Big(q\big(\frac{k(n(t))}{n(t)}\big)\big(\frac{n(t)}{k(n(t))} \sup_{n \geq n(t)} \frac{k(n)d_n}{nc_n}\big)^{1/\eta}\Big) \\ &= O\Big(q\big(\frac{k(n(t))}{n(t)}\big)\big(\frac{d_n}{c_n}\big)^{1/\eta}\Big) \\ &= O(t). \end{aligned}$$

Since $\mathcal{B}_n$ is a VC class, Theorem 5.1 of Alexander [1987] yields

$$\sup\Big\{\Big|\frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\mathbf{1} - \boldsymbol{F}(X_i, Y_i) \in \mathbf{1} - \boldsymbol{F}(B)\}}{\Pr\{\mathbf{1} - \boldsymbol{F}(X, Y) \in \mathbf{1} - \boldsymbol{F}(B)\}} - 1\Big| \;\Big| \\ B \in \mathcal{B}_n, \Pr\{\mathbf{1} - \boldsymbol{F}(X, Y) \in \mathbf{1} - \boldsymbol{F}(B)\} \geq \epsilon_n\Big\} \to 0,$$

provided $n\epsilon_n \to \infty$. Because of (7.8) and the last assumption of (3.12), the choice $\epsilon_n = q(k/n)(d_n/c_n)^{1/\eta}\nu(A)/2$ leads to the assertion. □

Proof of Theorem 3.1. Now the consistency of $\hat{p}_n$ can be proven as shown in (7.1). For this note that, because of (3.10), $\boldsymbol{F}^{-1}_{\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\boldsymbol{\gamma}}}(\mathbf{1} - \hat{D}_n/c_n)$ belongs to $\mathcal{B}_n$ with probability tending to 1 and that $\log c_n = o((r(n))^{1/2})$ implies $c_n^{1/\hat{\eta}} = c_n^{1/\eta}(1 + o_P(1))$ since $\hat{\eta}$ was assumed $\sqrt{r(n)}$–consistent for η. □

Appendix A. Some analytical results

Write $Q(x, y) = \Pr\{1 - F_1(X) < x \text{ and } 1 - F_2(Y) < y\}$. As in (2.1) suppose

$$\lim_{t \downarrow 0} \frac{\frac{Q(tx, ty)}{q(t)} - c(x, y)}{q_1(t)} =: \tilde{c}_1(x, y) \tag{A.1}$$

exists for $x, y \geq 0$ (but $x + y > 0$) with q positive, $q_1(t) \to 0, (t \downarrow 0)$, $\tilde{c}_1$ non constant and not a multiple of c and w.l.o.g. $c(1, 1) = 1$. Moreover assume that (A.1) holds uniformly on

$$\{(x, y) \mid x^2 + y^2 = 1, x \geq 0, y \geq 0\}.$$

It is easy to see that this implies the same for the limit relation

$$\lim_{t \downarrow 0} \frac{\frac{Q(tx, ty)}{Q(t, t)} - c(x, y)}{q_1(t)} = \tilde{c}_1(x, y) - \tilde{c}_1(1, 1)c(x, y) =: c_1(x, y) \tag{A.2}$$

with $c_1(x, y) \not\equiv 0$. Clearly q_1 is a regularly varying function with non-negative index.

Proposition A.1. *Under the stated conditions, relations* (A.1) *and* (A.2) *hold locally uniformly on* $(0,\infty)^2$. *If the index of the regularly varying function* q_1 *is strictly positive, the relation holds locally uniformly on* $[0,\infty)^2$.

Proof. Relation (A.2) implies that the function $Q(x,x)$ is regularly varying of second order (cf. de Haan and Stadtmüller [1996]), hence we can assume that $c(x,x) = x^{1/\eta}$ and $c_1(x,x) = x^{1/\eta}\frac{x^\tau - 1}{\tau}$ with $1/\eta$ the index of regular variation of q and $\tau \geq 0$ the index of regular variation of q_1.

Let $(x(t), y(t))$ converge to $(x,y) \in (0,\infty)^2$ as $t \downarrow 0$. Write $(x(t), y(t)) = a(t)(u(t), v(t))$ with $u^2(t) + v^2(t) = 1$. Then, as $t \downarrow 0$, $(u(t), v(t)) \to (u, v)$ and $a(t) \to a > 0$, say and

$$\begin{aligned}
\frac{Q(tx(t), ty(t))}{Q(t,t)} &= \frac{Q(ta(t)u(t), ta(t)v(t))}{Q(ta(t), ta(t))} \cdot \frac{Q(ta(t), ta(t))}{Q(t,t)} \\
&= \Big(c(u(t), v(t)) + q_1(ta(t))\big[c_1(u(t), v(t)) + o(1)\big]\Big) \\
&\quad \cdot \Big(a(t)^{1/\eta} + q_1(t)\big[c_1(a(t), a(t)) + o(1)\big]\Big) \\
&= c(u(t), v(t))a(t)^{1/\eta}\Big(1 + q_1(t)a(t)^\tau(1+o(1))\big[\frac{c_1(u(t), v(t))}{c(u(t), v(t))} + o(1)\big]\Big) \\
&\quad \cdot \Big(1 + q_1(t)a(t)^{-1/\eta}\big[c_1(a(t), a(t)) + o(1)\big]\Big) \\
&= c(x(t), y(t))\Big(1 + q_1(t)\big[a(t)^\tau \frac{c_1(u(t), v(t))}{c(u(t), v(t))} \\
&\quad + a(t)^{-1/\eta}(c_1(a(t), a(t)) + o(1))\big]\Big).
\end{aligned}$$

It follows that

$$\lim_{t\downarrow 0} \frac{\frac{Q(tx(t), ty(t))}{Q(t,t)} - c(x(t), y(t))}{q_1(t)} = \Big(a^\tau \frac{c_1(u,v)}{c(u,v)} + \frac{c_1(a,a)}{c(a,a)}\Big)c(x,y).$$

□

The proof shows that the following relation holds.

Corollary A.1. *For* $a, u, v > 0$

$$c_1(au, av) = a^{1/\eta+\tau}c_1(u,v) + a^{1/\eta}\frac{a^\tau - 1}{\tau}c(u,v) \tag{A.3}$$

(remember we have chosen q_1 *in such a way that* $c_1(a,a) = a^{1/\eta}\frac{a^\tau-1}{\tau}$*).*

Remark. Write

$$R(s,t) := \frac{c_1(e^s, e^t)}{c(e^s, e^t)}.$$

Then for all h, s and t

$$R(s+h, t+h) - R(s,t) = R(s,t)(e^{h\tau} - 1) + R(h,h).$$

Hence

$$R_1(s,t) = \lim_{h\to 0} \frac{R(s+h, t+h) - R(s,t)}{h} = \tau R(s,t) + R_1(0,0) = \tau R(s,t) + 1.$$

This means that for $\tau = 0$,

$$R_1(s,t) = 1 \text{ for all } s, t$$

and $\tau > 0$

$$R_1(s,t) = \tau \tilde{R}(s,t)$$

with $\tilde{R}(s,t) := R(s,t) + 1/\tau$. Hence τ and the values of c_1/c on the unit circle determine the values of $R(s,t)$ everywhere.

Acknowledgement: Holger Drees was partly supported by a Heisenberg fellowship, granted by the DFG. Ana Ferreira's research was partially supported by PRAXIS XXI/FEDER.

References

K. S. Alexander. Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Relat. Fields*, 75:379–423, 1987.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analysis. *Extremes*, 2:339–366, 1999.

L. de Haan and J. de Ronde. Sea and wind: Multivariate extremes at work. *Extremes*, 1: 7–45, 1998.

L. de Haan and S. Resnick. Estimating the limit distribution of multivariate extremes. *Stochastic Models*, 9:275–309, 1993.

L. de Haan and A. K. Sinha. Estimating the probability of a rare event. *Annals of Statistics*, 27:732–759, 1999.

L. de Haan and U. Stadtmüller. Generalized regular variation of second order. *J. Australian Mathematical Society*, 61:381–395, 1996.

A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme value distribution. *The Annals of Statistics*, 17:1833–1855, 1989.

H. Drees. A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference*, 66:95–112, 1998a.

H. Drees. On smooth statistical tail functionals. *Scandinavian Journal of Statistics*, 25: 187–210, 1998b.

J. H. Einmahl. Poisson and Gaussian approximation of weighed local empirical processes. *Stoch. Proc. Appl.*, 70:31–58, 1997.

X. Huang. *Statistics of bivariate extreme values.* PhD thesis, Erasmus University, Rotterdam, 1992.

A. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *J. Royal Statistical Society, B*, 59:475–499, 1997.

A. W. Ledford and J. A. Tawn. Concomitant tail behavior for extremes. *Advances in Applied Probability*, 30:197–215, 1998.

L. Peng. Estimation of the coefficient of tail dependence in bivariate extremes. *Statistics & Probability Letters*, 43:399–409, 1999.

D. Pollard. *Convergence of stochastic processes.* Springer, New York, 1984.

S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied probability.* Springer Verlag, 1987.

H. Ruben. An asymptotic expansion for the multivariate normal distribution and Mill's ratio. *Journal of Research of the National Bureau of Standards*, 68B:3–11, 1964.

R. Schoenberg. *Maximum likelihood estimation.* Aptech Systems Inc., 1996. Gauss maxlik application.

G. Shorack and J. Wellner. *Empirical processes with applications to Statistics.* Wiley, New York, 1986.

R. L. Smith. Estimating tails of probability distributions. *Annals of Statistics*, 15:1174–1207, 1987.

W. Vervaat. Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. und verw. Gebiete*, 23:249–253, 1972.

(Draisma, de Haan) Econometrisch Instituut, Erasmus Universiteit Rotterdam
E-mail address: draisma@mgz.fgg.eur.nl, ldhaan@few.eur.nl

(Drees) Institute for Applied Mathematics, Ruprecht-Karls-University, Heidelberg
E-mail address: hdrees@statlab.uni-heidelberg.de

(Ferreira) EURANDOM, Eindhoven, on leave from Universidade Técnica de Lisboa
E-mail address: ferreira@eurandom.tue.nl

Tables and Figures

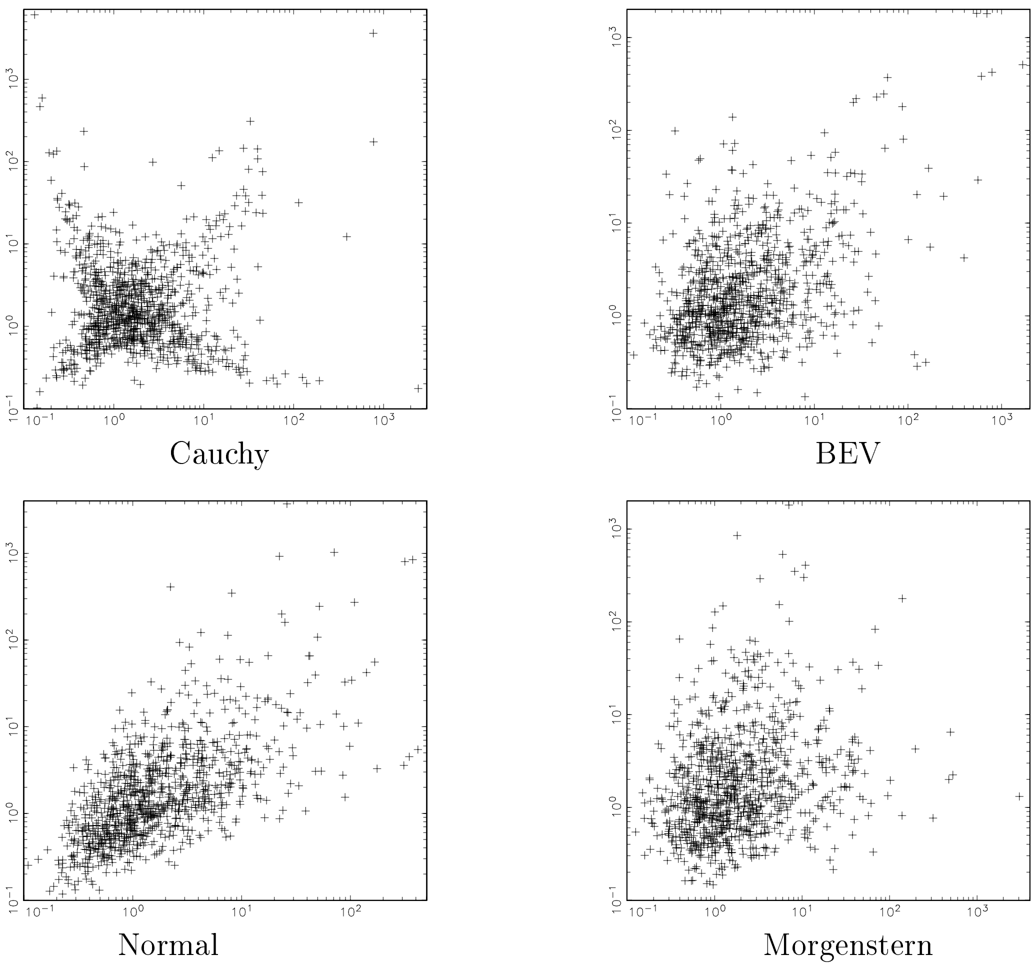
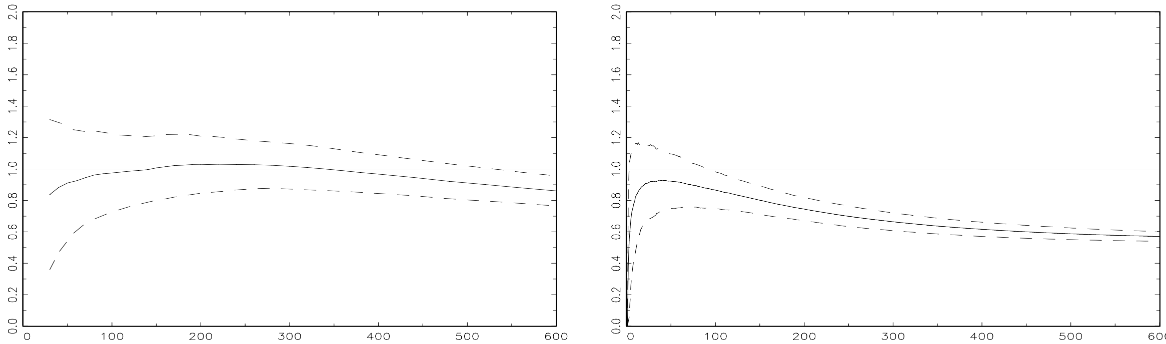


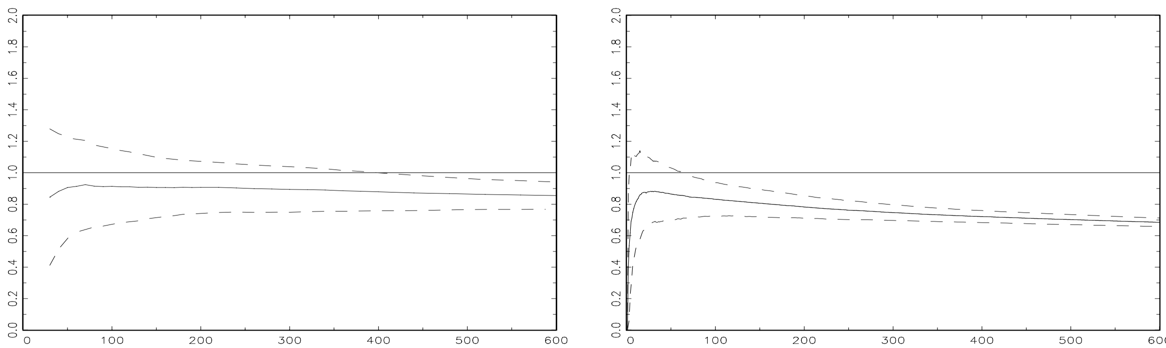
Figure 1. Scatterplot of a sample of size $n = 1000$ of each distribution. The BEV and Morgenstern distributions have Fréchet marginal distributions; for easy comparison a marginal transformation to the same distribution was applied to the Cauchy and normal samples.

Table 1. The ML-estimator, $\hat{\eta}_1$, and the Hill estimator, $\hat{\eta}_2$ (sample size $n = 1000$). Tabulated are mean and observed standard deviation of the estimator, and mean of estimates $\hat{\sigma}_{(i)}$ and $\hat{\sigma}_{(d)}$. The last column indicates the proportion of samples in which asymptotic dependence hypothesis is accepted in size 5% tests, based on $\hat{\sigma}_{(i)}$ resp. $\hat{\sigma}_{(d)}$.

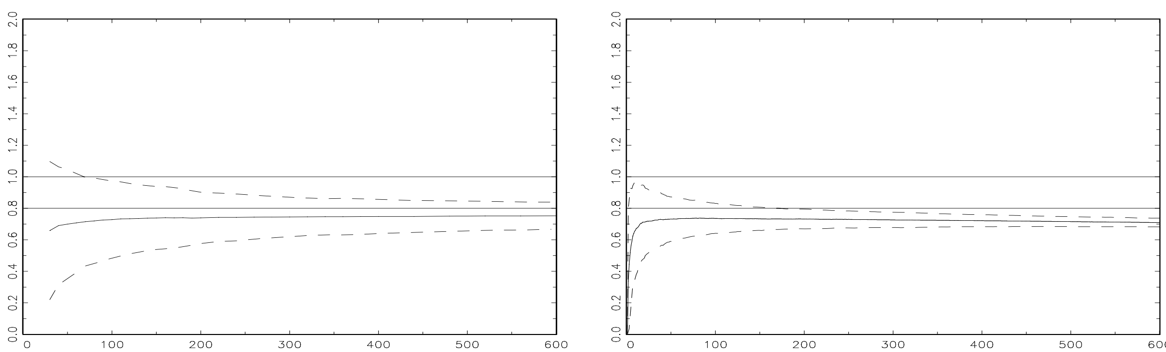
	m	$\hat{\eta}$ avg.	Standard deviation observed	$\hat{\sigma}_{(i)}$	$\hat{\sigma}_{(d)}$	$\eta = 1$ accepted; test with $\hat{\sigma}_{(i)}$	with $\hat{\sigma}_{(d)}$
ML, $\hat{\eta}_1$							
Cauchy	80	0.96	0.171	0.167	0.171	0.89	0.92
	160	1.01	0.125	0.112	0.111	0.93	0.95
	240	1.03	0.099	0.083	0.082	0.96	0.96
BEV	80	0.91	0.159	0.146	0.153	0.81	0.86
	160	0.91	0.112	0.094	0.098	0.68	0.72
	240	0.90	0.093	0.070	0.073	0.55	0.58
Normal	80	0.72	0.166	0.125	0.146	0.36	0.38
	160	0.74	0.120	0.080	0.092	0.16	0.18
	240	0.74	0.090	0.059	0.067	0.04	0.05
Morgenstern	80	0.47	0.156	0.123	0.167	0.05	0.06
	160	0.49	0.105	0.077	0.104	0.00	0.00
	240	0.50	0.082	0.057	0.076	0.00	0.00
Hill, $\hat{\eta}_2$							
Cauchy	40	0.93	0.119	0.115	0.124	0.81	0.88
	80	0.89	0.083	0.076	0.085	0.57	0.63
	120	0.84	0.064	0.056	0.067	0.15	0.22
BEV	40	0.87	0.112	0.100	0.114	0.60	0.71
	80	0.84	0.075	0.064	0.076	0.29	0.34
	120	0.82	0.058	0.049	0.059	0.05	0.08
Normal	40	0.73	0.099	0.0817	0.112	0.12	0.18
	80	0.74	0.067	0.0535	0.073	0.01	0.01
	120	0.73	0.052	0.0409	0.056	0.00	0.00
Morgenstern	40	0.51	0.072	0.0663	0.129	0.00	0.00
	80	0.53	0.050	0.0443	0.084	0.00	0.00
	120	0.54	0.042	0.0344	0.064	0.00	0.00



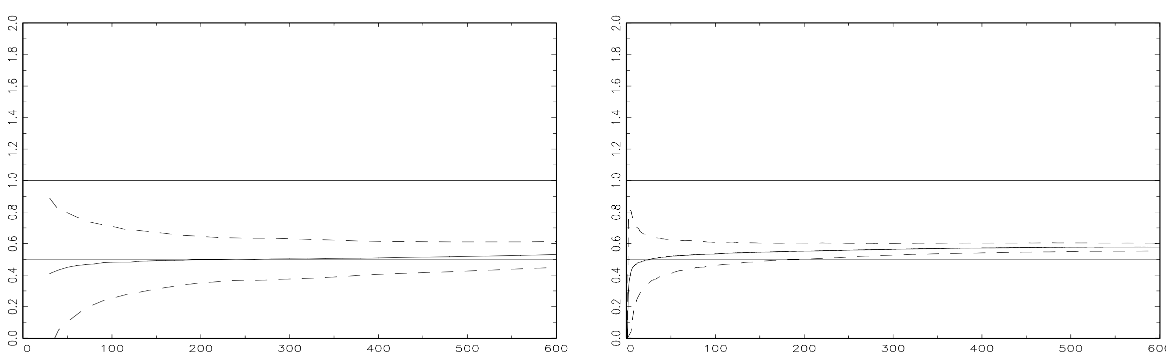
Cauchy distribution



Bivariate extreme value distribution



Normal distribution



Morgenstern distribution

FIGURE 2. The ML-estimator, $\hat{\eta}_1$, on the left and the Hill estimator, $\hat{\eta}_2$, on the right as a function of m (sample size $n = 1000$). The graphs show the average over 250 samples (solid line). Observed standard errors are indicated by the dashed lines ($\pm$ 1.64 st.deviations). The horizontal lines indicate $\eta = 1$ and the true η for each distribution.

Table 2. Pengs estimator, $\hat{\eta}_3$, and our estimator, $\hat{\eta}_4$ (sample size $n = 1000$). Tabulated are mean and observed standard deviation of the estimator, and mean of estimates $\hat{\sigma}_{(i)}$ and $\hat{\sigma}_{(d)}$; the proportion of samples in which asymptotic dependence hypothesis is accepted in size 5% tests, based on $\hat{\sigma}_{(i)}$ resp. $\hat{\sigma}_{(d)}$; the last column gives the number of samples (out of 250) in which either $\hat{\eta}$ or $\hat{\sigma}^2$ could not be calculated.

	k	$\hat{\eta}$ avg.	Standard deviation observed	$\hat{\sigma}_{(i)}$	$\hat{\sigma}_{(d)}$	$\eta = 1$ accepted $\hat{\sigma}_{(i)}$	$\hat{\sigma}_{(d)}$	Missing
Peng, $\hat{\eta}_3$								
Cauchy	40	1.05	0.361	0.164	0.176	0.84	1.00	6
	80	0.97	0.178	0.112	0.124	0.76	0.95	1
	120	0.88	0.120	0.081	0.102	0.52	0.65	1
BEV	40	0.96	0.228	0.138	0.164	0.76	0.98	5
	80	0.85	0.124	0.087	0.120	0.42	0.65	2
	120	0.80	0.086	0.065	0.097	0.15	0.24	0
Normal	40	0.78	0.194	0.128	0.215	0.41	0.96	2
	80	0.75	0.093	0.082	0.135	0.14	0.30	0
	120	0.74	0.072	0.061	0.102	0.02	0.03	0
Morgenstern	40	0.55	0.221	0.169	0.524	0.21	1.00	10
	80	0.54	0.108	0.086	0.263	0.03	0.19	0
	120	0.55	0.070	0.062	0.179	0.00	0.00	0
This paper, $\hat{\eta}_4$								
Cauchy	80	1.03	0.230	0.202	0.191	0.89	0.94	1
	160	0.97	0.144	0.133	0.136	0.83	0.92	0
	240	0.89	0.098	0.095	0.109	0.61	0.72	0
BEV	80	0.96	0.174	0.171	0.176	0.83	0.91	2
	160	0.86	0.099	0.098	0.115	0.55	0.68	0
	240	0.82	0.067	0.072	0.090	0.23	0.30	0
Normal	80	0.79	0.146	0.144	0.187	0.50	0.70	0
	160	0.76	0.080	0.083	0.113	0.12	0.26	0
	240	0.75	0.058	0.061	0.084	0.04	0.06	0
Morgenstern	80	0.55	0.187	0.181	0.343	0.23	0.66	0
	160	0.54	0.085	0.087	0.172	0.02	0.04	0
	240	0.55	0.055	0.060	0.117	0.00	0.00	0

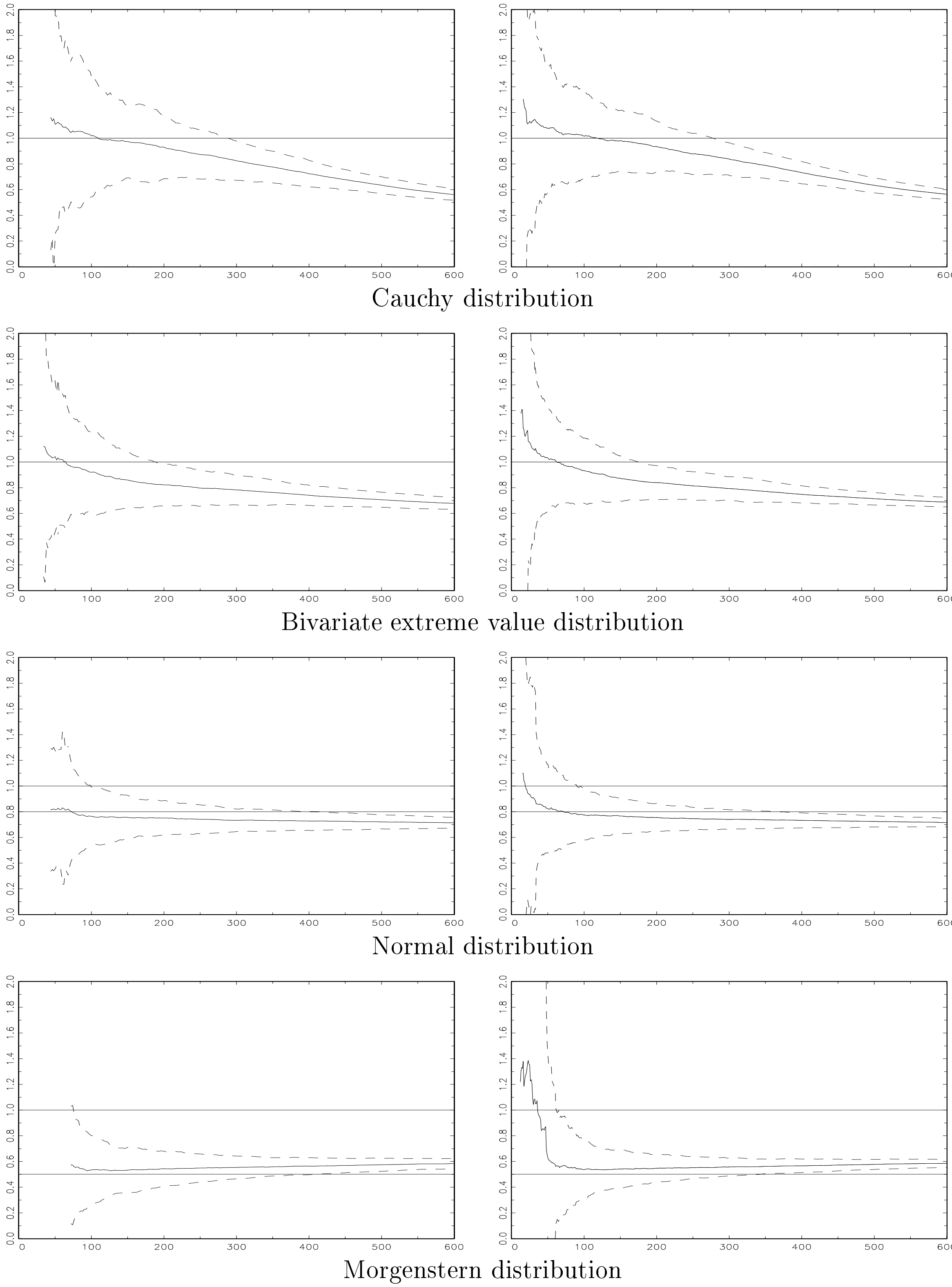


FIGURE 3. Peng's estimator, $\hat{\eta}_3$, as a function of $2k$ on the left, and our estimator, $\hat{\eta}_4$, as a function of k on the right (sample size $n = 1000$). The graphs show the average over 250 samples (solid line). Observed standard errors are indicated by the dashed lines ($\pm$ 1.64 st.deviations). The horizontal lines indicate $\eta = 1$ and the true η for each distribution.

TABLE 3. Estimating failure probabilities. The table lists the median values of the estimates. The probability estimates are the estimate for the general case ($\hat{p}_{\hat{\eta}}$), for the asymptotic dependent case ($\hat{p}_1$), and $\hat{p} = \hat{p}_{\hat{\eta}}$ or $= \hat{p}_1$, depending on whether asymptotic dependence is rejected resp. accepted.

k	γ_1	γ_2	η	$\hat{p}_{\hat{\eta}}$	$\hat{p}_1$	$\hat{p}$
Cauchy	1	1	1		$\times 10^{-5}$	
80	0.95	0.98	1.01	0.52	0.51	0.47
160	1.02	0.96	0.94	0.63	0.80	0.75
240	1.00	1.04	0.90	0.42	1.14	0.99
Normal	0	0	0.8		$\times 10^{-5}$	
80	-0.15	-0.17	0.77	0.00003	0.00240	0.00005
160	-0.20	-0.13	0.75	0.00000	0.00000	0.00000
240	-0.17	-0.20	0.74	0.00000	0.00000	0.00000
Exponential/Normal	0	0	0.8		$\times 10^{-5}$	
80	0.016	0.040	0.77	0.20	2.2	0.61
160	0.062	0.036	0.75	0.36	3.7	0.44
240	0.045	0.058	0.74	0.39	6.1	0.43
Morgenstern	1	1	0.5		$\times 10^{-5}$	
80	0.99	1.01	0.54	1.1	27	19.7
160	1.03	0.97	0.53	1.3	58	1.3
240	1.00	1.02	0.55	2.0	84	2.0

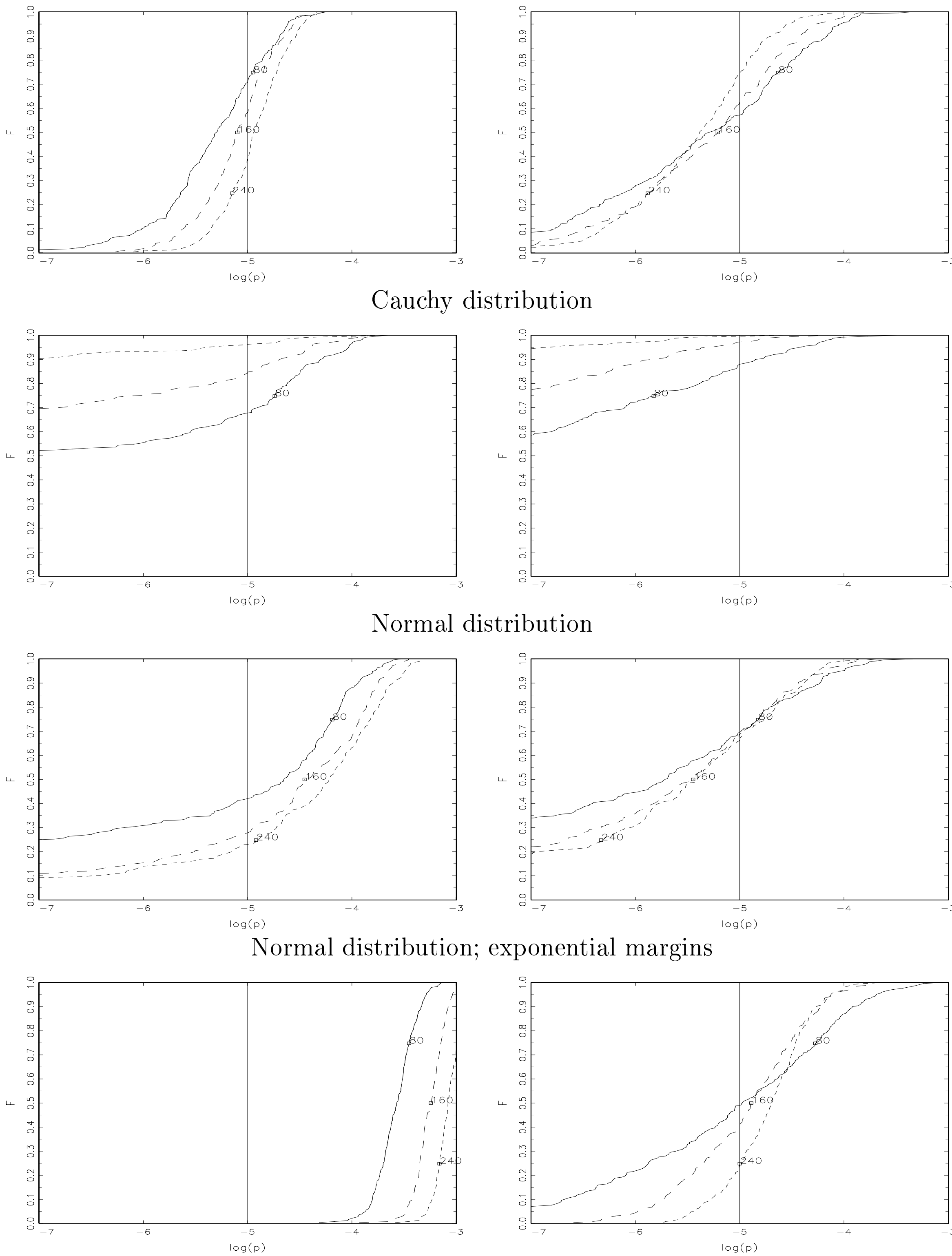


FIGURE 4. Failure probabilities. The graphs show the empirical distributions of the estimates for different k. The graphs on the left refer to $\hat{p}_1$ (the estimate assuming dependence) and the graphs on the right to $\hat{p}_{\hat{\eta}}$ (the general estimate). The true value $p = 10^{-5}$ is indicated by the vertical line ($n = 1000$).

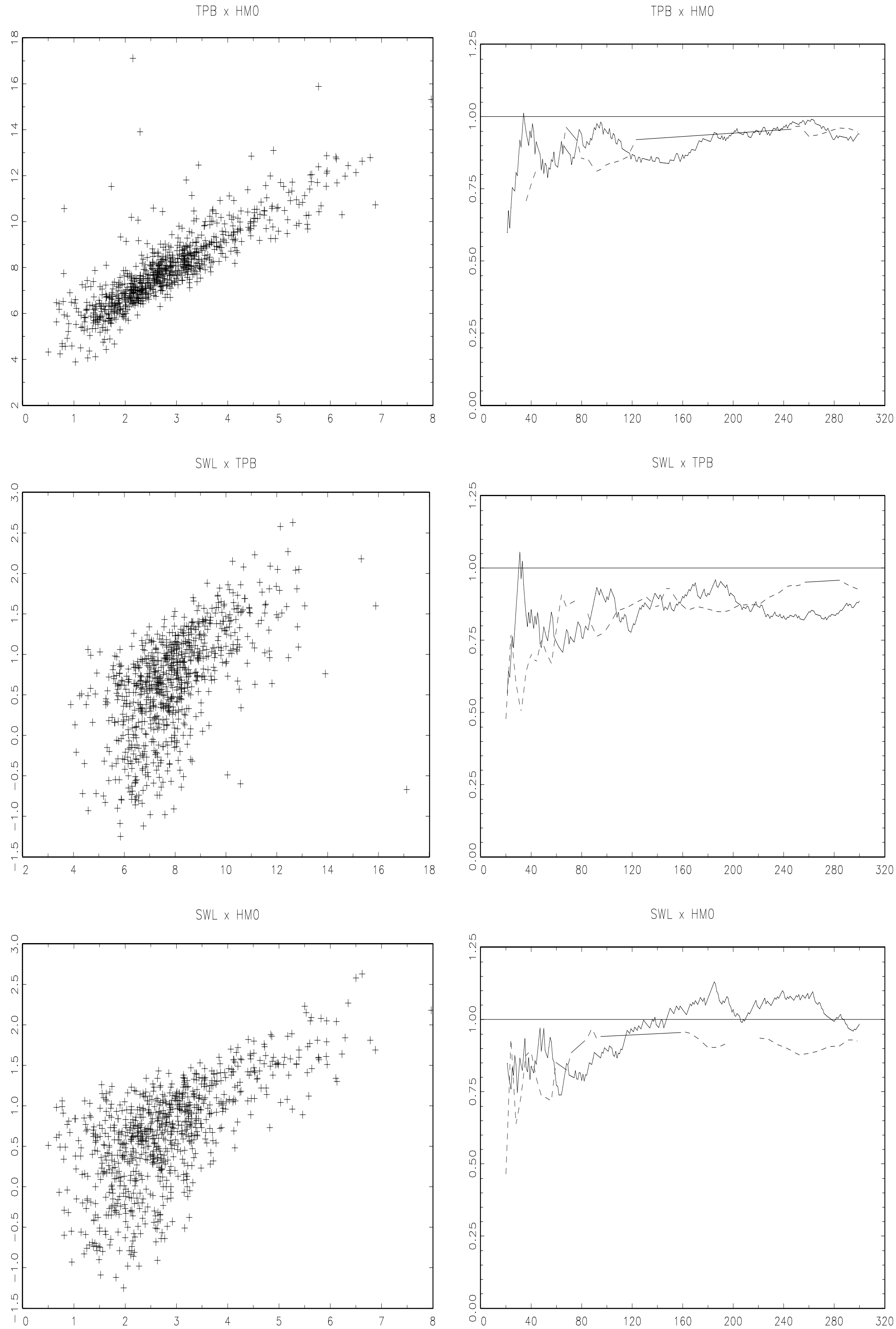


FIGURE 5. Estimating dependence between waveheight Hm0, waveperiod Tpb and still water level SWL. On the left a scatterplot and on the right $\hat{\eta}_4$ as a function of k. The solid lines display the estimates for various k. The horizontal lines indicate the $\eta = 1$ level. The area below the dotted line is the critical area of a one sided, size 5% test for asymptotic dependence.