Annals of the Institute of Statistical Mathematics

Supplementary material to “Testing for a δ-neighborhood of a generalized Pareto copula”

Stefan Aulbach · Michael Falk · Timo Fuller

1 A copula not in any max-domain of attraction

The following result provides a one parametric family of bivariate rv, which are easy to simulate. Each member of this family has the property that its corresponding copula does not satisfy the extreme value condition (2). However, as the parameter tends to zero, the copulas of interest come arbitrarily close to a GPC, which, in general, is in the domain of attraction of an EVD.

Lemma S.1 *Let the rv V have df $H_\lambda(u) := u\big(1+\lambda\sin(\log(u))\big)$, $0 \le u \le 1$, where $\lambda \in \big[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\big]$. Note that $H_\lambda(0) = 0$, $H_\lambda(1) = 1$ and $H'_\lambda(u) \ge 0$ for $0 < u < 1$. Furthermore let the rv U be independent of V and uniformly distributed on $(0,1)$. Put $S_1 := U =: 1 - S_2$. Then the copula C_λ corresponding to the bivariate rv*

$$\boldsymbol{X} := -\frac{V}{2}\left(\frac{1}{S_1}, \frac{1}{S_2}\right)^{\intercal} \in (-\infty, 0]^2 \tag{S.1}$$

is not in the domain of attraction of any multivariate EVD if $\lambda \ne 0$, whereas C_0 is a GPC with corresponding D-norm

$$\|\boldsymbol{x}\|_D = \|\boldsymbol{x}\|_1 - \frac{|x_1||x_2|}{\|\boldsymbol{x}\|_1}$$

for $\boldsymbol{x} = (x_1, x_2)^{\intercal} \ne \boldsymbol{0}$.

Denote by F_λ the df of $-V/S_1 =_D -V/S_2$. Elementary computations yield that it is given by

$$F_\lambda(x) = \begin{cases} |x|^{-1}\big(\frac{1}{2} + \frac{\lambda}{5}\big), & \text{if } x \le -1, \\ 1 - |x|\Big(\frac{1}{2} + \frac{\lambda}{5}\big(2\sin(\log|x|) - \cos(\log|x|)\big)\Big), & \text{if } -1 < x < 0, \end{cases}$$

S. Aulbach · M. Falk · T. Fuller
University of Würzburg, Institute of Mathematics, Emil-Fischer-Str. 30, D-97074 Würzburg
E-mail: stefan.aulbach@uni-wuerzburg.de

and, thus, F_λ is continuous and strictly increasing on $(-\infty, 0]$.

Proof We show that

$$\lim_{s\downarrow 0}\frac{1-C_\lambda(1-s,1-s)}{s}$$

does not exist for $\lambda \in \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]\setminus\{0\}$. Since C_λ coincides with the copula of $2\boldsymbol{X}$ we obtain setting $s = 1-F_\lambda(t)$, $t\uparrow 0$,

$$\begin{aligned}
\frac{1-C_\lambda\big(F_\lambda(t),F_\lambda(t)\big)}{1-F_\lambda(t)} &= \frac{1-P\big(-V/S_1\le t, -V/S_2\le t\big)}{1-P\big(-V/S_1\le t\big)}\\
&= \frac{1-P\big(V\ge |t|\max\{U,1-U\}\big)}{1-P\big(V\ge |t|U\big)}\\
&= \frac{\int_0^1 P\big(V\le |t|\max\{u,1-u\}\big)\,du}{\int_0^1 P\big(V\le |t|u\big)\,du}\\
&= \frac{\int_0^{1/2} H_\lambda\big(|t|(1-u)\big)\,du + \int_{1/2}^1 H_\lambda\big(|t|u\big)\,du}{\int_0^1 H_\lambda\big(|t|u\big)\,du}\\
&= 2\frac{\int_{1/2}^1 H_\lambda\big(|t|u\big)\,du}{\int_0^1 H_\lambda\big(|t|u\big)\,du}.
\end{aligned}$$

The substitution $u\mapsto u/|t|$ yields

$$1-\frac{1}{2}\frac{1-C_\lambda\big(F_\lambda(t),F_\lambda(t)\big)}{1-F_\lambda(t)} = 1-\frac{\int_{|t|/2}^{|t|} H_\lambda(u)\,du}{\int_0^{|t|} H_\lambda(u)\,du} = \frac{\int_0^{|t|/2} H_\lambda(u)\,du}{\int_0^{|t|} H_\lambda(u)\,du}$$

where we have for each $0<c\le 1$

$$\int_0^c H_\lambda(u)\,du = \frac{c^2}{2} + \lambda\int_0^c u\sin(\log(u))\,du.$$

and

$$\int_0^c u^2\cdot\frac{1}{u}\sin(\log(u))\,du = \frac{c^2}{5}\big(2\sin(\log(c))-\cos(\log(c))\big)$$

which can be seen by applying integration by parts twice. Hence we obtain

$$\frac{\int_0^{|t|/2} H_\lambda(u)\,du}{\int_0^{|t|} H_\lambda(u)\,du} = \frac{1}{4}\frac{\frac{1}{2}+\frac{\lambda}{5}\big(2\sin(\log|t|-\log(2))-\cos(\log|t|-\log(2))\big)}{\frac{1}{2}+\frac{\lambda}{5}\big(2\sin(\log|t|)-\cos(\log|t|)\big)},$$

whose limit does not exist for $t\uparrow 0$ if $\lambda\in\left[-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right]\setminus\{0\}$; consider, e.g., the sequences $t_n^{(1)} = -\exp\big((1-2n)\pi\big)$ and $t_n^{(2)} = -\exp\big((1/2-2n)\pi\big)$ as $n\to\infty$.

On the other hand, elementary computations show for $\boldsymbol{x} = (x_1,x_2)^\intercal \in (-\infty,0]^2\setminus\{\boldsymbol{0}\}$

$$\lim_{\varepsilon\downarrow 0}\frac{1-C_0(1+\varepsilon\boldsymbol{x})}{\varepsilon} = 2E(\max\{|x_1|S_1,|x_2|S_2\}) = \|\boldsymbol{x}\|_1 - \frac{|x_1||x_2|}{\|\boldsymbol{x}\|_1}.$$

The remaining assertion is thus implied by Section 2 of Aulbach et al (2012). $\square$

Remark S.1 Similar results as in Lemma S.1 can be obtained for different distributions of the rv (S_1, S_2) in (S.1), which is still assumed to be independent of V. If $\lambda = 0$, then $S_1 = U_1$, $S_2 = U_2$ gives

$$\|\boldsymbol{x}\|_D = \|\boldsymbol{x}\|_\infty + \frac{\left(\|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty\right)^2}{3\|\boldsymbol{x}\|_\infty}, \qquad \boldsymbol{x} \neq \boldsymbol{0},$$

where U_1, U_2 are independent and uniformly distributed on $[0,1]$. However, if $\lambda \neq 0$, then we obtain for $t \in (-1, 0)$

$$\frac{1 - C_\lambda\big(F_\lambda(t), F_\lambda(t)\big)}{1 - F_\lambda(t)} = \frac{4}{3}\left(1 + \lambda\,\frac{\sin(\log|t|) + \cos(\log|t|)}{10 + 4\lambda\big(2\sin(\log|t|) - \cos(\log|t|)\big)}\right),$$

which has no limit for $t \uparrow 0$; consider, e.g., the sequences $\big(t_n^{(1)}\big)_n$ and $\big(t_n^{(2)}\big)_n$ as in the proof of Lemma S.1.

However, if U is uniformly distributed on $[0,1]$, then $S_1 = S_2 = U$ implies

$$\begin{aligned} C_\lambda(u_1, u_2) &= P\left(-\frac{V}{S_1} \leq F_\lambda^{-1}(u_1), -\frac{V}{S_2} \leq F_\lambda^{-1}(u_2)\right) \\ &= F_\lambda\big(\min\{F_\lambda^{-1}(u_1), F_\lambda^{-1}(u_2)\}\big) \\ &= \min\{u_1, u_2\} \quad \text{for} \quad u_1, u_2 \in (0,1), \end{aligned}$$

which does not depend on λ. Moreover, C_λ is the copula of an EVD with D-norm $\|\cdot\|_D = \|\cdot\|_\infty$, cf. (3), and satisfies $\varepsilon^{-1}(1 - C_\lambda(\boldsymbol{1} + \varepsilon\boldsymbol{x})) = \|\boldsymbol{x}\|_\infty$ for $\boldsymbol{x} \in (-\infty, 0]^2$ whenever $0 < \varepsilon \leq \|\boldsymbol{x}\|_\infty^{-1}$.

Example S.1 Let η_1, η_2 be two independent and standard negative exponential distributed rv. Put for $t \in [0,1]$

$$X_t := \max\left(\frac{-V}{2\exp\left(\frac{\eta_1}{1-t}\right)}, \frac{-V}{2\exp\left(\frac{\eta_2}{t}\right)}\right) = \frac{-V}{2\exp\left(\max\left(\frac{\eta_1}{1-t}, \frac{\eta_2}{t}\right)\right)},$$

where the rv V is independent of η_1, η_2 and follows the df H_λ defined in Lemma S.1 with $\lambda \in \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$. Note that

$$\max\left(\frac{\eta_1}{1-t}, \frac{\eta_2}{t}\right) =_D \eta_1 =_D \eta_2$$

and, thus, the process $\boldsymbol{X} = (X_t)_{t\in[0,1]}$ has identical continuous marginal df.

For $\lambda = 0$, the process $\boldsymbol{X}$ is a *generalized Pareto process*, whose pertaining copula process is in the max-domain of attraction of a SMSP, see Aulbach et al (2013). For $\lambda \neq 0$ this is not true: Just consider the bivariate rv $(X_0, X_1) =_D -\frac{V}{2}(1/U_1, 1/U_2)$, where $U_1 = \exp(\eta_1)$, $U_2 = \exp(\eta_2)$ and repeat the arguments in Lemma S.1.

2 Simulations

In this section we provide some simulations, which indicate the performance of the test statistics $T_n(c_n)$ and $\hat{T}_n(c_n)$ from Theorem 1 and Theorem 2, respectively. All computations were performed using the R package CompQuadForm written by Pierre Lafaye de Micheaux and Pierre Duchesne. We chose Imhof's (1961) method for computing the p-values of our test statistics; cf. Duchesne and Lafaye de Micheaux (2010) for an overview of simulation techniques of quadratic forms in normal variables.

2.1 Selection of the parameters

Recall that the test statistics under consideration depend on several sequences, which are required to have certain asymptotic properties as the sample size grows to infinity. For a finite sample size, however, we will see later in this section that the test results are highly sensitive to a proper selection of the corresponding elements of these sequences. These elements are referred to as the parameters of the test statistics in what follows. Additionally, we assume that $\delta > 0$ is given by the application; e.g., we could use $\delta = 1$ for testing whether the Gumbel-Hougaard family in Example 1 of the article is a candidate for modeling the copula of the observed data.

For convenience, we assume for a moment that our data consist of independent copies $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$ of a rv $\boldsymbol{U}$ which is distributed according to a copula. If we want to test the hypothesis that this copula is in a δ-neighborhood of a GPC for some fixed $\delta > 0$, the test statistic $T_n(c_n)$ depends on the parameters k and c_n, where $k \geq 2$ is an integer and $c_n \in (0,1)$ has the asymptotic properties $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to 0$ as $n \to \infty$. Note that, if the copula is actually in a δ_0-neighborhood for some $\delta_0 > \delta$, then we get

$$\begin{aligned} &1 - P\left(\boldsymbol{U} \leq \left(1 - \frac{c_n}{j}\right)\mathbf{1}\right) \\ &= \frac{c_n}{j}\, m_D + \left(\frac{c_n}{j}\right)^{1+\delta} \left(\frac{c_n}{j}\right)^{\delta_0 - \delta} \frac{1 - P\big(\boldsymbol{U} \leq \big(1 - \frac{c_n}{j}\big)\mathbf{1}\big) - \frac{c_n}{j}\, m_D}{\big(\frac{c_n}{j}\big)^{1+\delta_0}} \\ &= \frac{c_n}{j}\, m_D + o\left(\left(\frac{c_n}{j}\right)^{1+\delta}\right) \quad \text{for} \quad j = 1, \dots, k. \end{aligned}$$

Thus, the same arguments that proved Theorem 1 also show:

Corollary S.1 *Let $\delta > 0$ and $k \in \mathbb{N}$, $k \geq 2$. If a copula C is in the δ_0-neighborhood of a GPC for some $\delta_0 > \delta$, and $c_n \in (0,1)$ satisfies $c_n \to 0$, $nc_n \to \infty$, and $nc_n^{1+2\delta} \to s \geq 0$, then the conclusions of Theorem 1 remain valid.*

Since this result suggests that the condition $nc_n^{1+2\delta} \to_{n\to\infty} 0$ is rather a mild one, we focus on exploiting $c_n \to 0$ and $nc_n \to \infty$ as $n \to \infty$ in order to derive a reasonable value c_n for a finite sample size n. Given that the df of $\boldsymbol{U}$ is in the δ-neighborhood of a GPC for some $\delta > 0$, we obtain for $j = 1, \dots, k$

$$E(n_j(c_n)) = n\,P\left(S_{\boldsymbol{U}}\left(-\frac{c_n}{j}\right)\right) = nc_n \cdot \frac{1 - P\big(\boldsymbol{U} \le \big(1 - \frac{c_n}{j}\big)\mathbf{1}\big)}{c_n} \sim \frac{nc_n m_D}{j} \tag{S.2}$$

and

$$\max_{1 \le j \le k-1} \left| \frac{1 - P\big(\boldsymbol{U} \le \big(1 - \frac{c_n}{j+1}\big)\mathbf{1}\big)}{1 - P\big(\boldsymbol{U} \le \big(1 - \frac{c_n}{j}\big)\mathbf{1}\big)} - \frac{j}{j+1} \right| = O\big(c_n^\delta\big) \tag{S.3}$$

as $n \to \infty$, where $n_j(c_n)$ is for each $j = 1, \dots, k$ the number of exceedances of the threshold $\big(1 - \frac{c_n}{j}\big)\mathbf{1}$ among $\boldsymbol{U}^{(1)}, \dots, \boldsymbol{U}^{(n)}$. Therefore, the task is, on the one hand, to choose c_n small enough such that the thresholds $\big(1 - \frac{c_n}{j}\big)\mathbf{1}$, $j = 1, \dots, k$, are sufficiently close to $\mathbf{1}$ in order to detect the δ-neighborhood; cf. (δ-n). On the other hand, c_n must be large enough in order to guarantee that there are sufficiently many observations "above" the thresholds, cf. (S.2), such that the asymptotic normality as in the proof of Theorem 1 is justified.

Recall that $T_n(c_n)$ is based on the estimators $\frac{j}{nc_n} n_j(c_n)$, $j = 1, \dots, k$, of the unknown value m_D. Thus, if (δ-n) holds, we expect the graph of the function $\gamma_* : (0, 1] \to [0, \infty)$, $c \mapsto \gamma_*(c) := \frac{1}{nc} n_1(c)$ to be almost constant on some interval, since – for proper arguments – this function estimates the (unknown) constant m_D; cf. (S.3). Moreover, we obtain pointwise approximate confidence intervals for m_D from the asymptotic normality

$$\left(\frac{nc_n}{\gamma_*(c_n)}\right)^{1/2} (\gamma_*(c_n) - m_D) \to_D N(0, 1), \tag{S.4}$$

which can be derived from (10) and (11). Therefore, we expect that there is some interval $I \subset (0, 1]$ with the both properties that the restriction of γ_* to I is constant, apart from random fluctuations, and that the pointwise approximate confidence intervals based on (S.4) are not too wide for $c \in I$. Due to the identity $n_j(c) = n_1(c/j)$ for $j = 1, \dots, k$, it is reasonable to choose c_n and k such that c_n and c_n/k are points in I *and* such that $n_k(c_n)$ is not too small. Altogether, the selection of c_n for finite n is a typical tradeoff situation, similar to the problem of choosing a threshold for the adaption of a generalized Pareto distribution to univariate data, see e.g. Embrechts et al (1997, Section 6.5).

In the more general case where the margins of the data are unknown, cf. Theorem 2, we have the additional parameter $m_n \in \mathbb{N}$ with $m_n \le n$, which has, among others, the asymptotical properties $m_n \to \infty$ and $m_n/n \to 0$ as $n \to \infty$. Since m_n is the size of the subsample that is used to compute the test statistic $\hat{T}_n(c_n)$ from – a restriction that is due to the need to estimate the margins from the data –, it is reasonable to require that the rate of convergence of $m_n/n \to_{n \to \infty} 0$ is rather low. Indeed, our practical experience from simulation studies suggests $m_{200} = 86$ and $m_{10\,000} = 2037$. For different values of n, the sequence $m_n = n/(\log(n))^{5/7}$ might be helpful, which roughly reproduces our empirical findings. Based on an earlier version of this paper, where the conditions on m_n were more restrictive, we originally aimed at a graphical approach that would allow to choose c_n and m_n simultaneously. It turned out, however, that the results drastically improved once that we decided to specify m_n first and c_n afterwards. The aforementioned values of m_n resulted from this investigation and turned out to perform well also in the current setting.

Once the value m_n has been chosen, the approach for copula data from above can be modified in order to select c_n and k. As remarked directly after Theorem 2, we now have the asymptotic properties $c_n \to 0$, $m_n c_n \to \infty$, and $m_n c_n^{1+2\delta} \to 0$ as $n \to \infty$, where we will focus on the former two; cf. Corollary S.1. Thus, the function γ_* is replaced with $\hat{\gamma}_*^{(m_n)}(c) := \frac{1}{m_n c}\hat{n}_{1,m_n}(c)$, where $\hat{n}_{1,m_n}(c)$ is given as in (16), and we obtain pointwise asymptotic confidence intervals from

$$\left(\frac{m_n c_n}{\hat{\gamma}_*^{(m_n)}(c_n)}\right)^{1/2}\left(\hat{\gamma}_*^{(m_n)}(c_n) - m_D\right) \to_D N(0,1) \quad \text{as} \quad n \to \infty. \tag{S.5}$$

Again, if the copula that underlies the data is in a δ-neighborhood of a GPC, we expect that there is an interval $\hat{I} \subset (0,1]$ such that $\hat{\gamma}_*^{(m_n)}$ is almost constant on $\hat{I}$ and that the confidence intervals motivated by (S.5) are not too wide. Thus, we choose c_n and k such that $c_n, c_n/k \in \hat{I}$ and such that $\hat{n}_{k,m_n}(c_n)$ is not too small.

2.2 Simulation results

In order to check the performance of the test statistics $T_n(c_n)$ and $\hat{T}_n(c_n)$ from Theorem 1 and Theorem 2, respectively, we now apply our tests to simulated data, i.e. we *know* whether the underlying copula is in a δ-neighborhood. Precisely, we consider the parametric family of copulas introduced in Lemma S.1, which is indexed with a parameter $\lambda \in \left[-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$, and apply the strategies derived in Section 2.1. Recall that for $\lambda = 0$ we obtain a GPC with $m_D = \frac{3}{2}$, whereas the case $\lambda \neq 0$ leads to a copula that is not in any δ-neighborhood of any GPC.

2.2.1 Copula data

First we apply the approach for copula data: For $\lambda = 0$ and $n = 10\,000$, Figure S.1 shows that the estimation of m_D by means of the function γ_* performs quite well for a wide range of arguments. According to Section 2.1, it appears reasonable to choose $c_{10\,000} = 0.2$ and $k = 2$, which means that there are $n_2(0.2) = 1\,482$ exceedances of the threshold $(1 - \frac{0.2}{2})\mathbf{1}$. Table S.1 summarizes the results derived from the test in Theorem 1, where the p-values are computed as $p(c_n) = 1 - F_k(T_n(c_n))$ and F_k denotes the df of $\sum_{i=1}^{k-1} \lambda_i \xi_i^2$ in Theorem 1. It can be seen that the hypothesis that the underlying copula belongs to a δ-neighborhood is not rejected for $\lambda = 0$ at a 5 % type I error level, as expected. Moreover, for $k = 2$, the test detects that the hypothesis

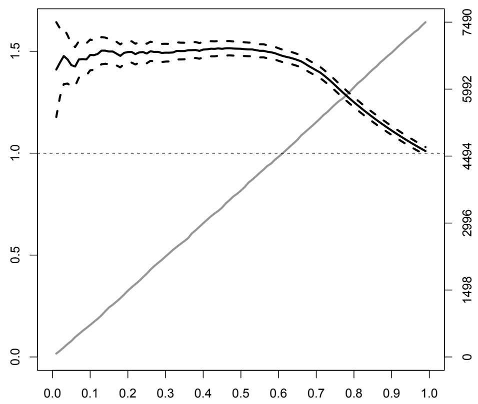

Fig. S.1 Plot of the function γ_* (black, left scale) for $\lambda = 0$ and $n = 10\,000$, together with pointwise approximate 95 % confidence intervals (dashed), cf. (S.4). The dashed horizontal line marks the lower bound of the constant m_D, cf. Section 3. The gray line (right scale) displays the the function $c \mapsto n_2(c)$, i.e. $k = 2$.

Table S.1 Test results for different values of λ and k with $n = 10\,000$ and $c_n = 0.2$.

k	λ	p-value	$n_k(0.2)$
2	0.00000	0.60789	1 482
	0.70711	0.00063	1 548
	0.20000	0.03504	1 434
	−0.20000	0.40295	1 524
	−0.70711	0.00004	1 703
3	0.00000	0.44411	972
	0.70711	0.00000	908
	0.20000	0.08966	952
	−0.20000	0.36178	1 030
	−0.70711	0.00002	1 167

is not true in the cases $\lambda \in \left\{-\frac{\sqrt{2}}{2}, 0.2, \frac{\sqrt{2}}{2}\right\}$. For $\lambda = -0.2$, however, the test does not reject the hypothesis although it is not satisfied, i.e., the test appears to be more sensitive to positive values of λ than to negative ones. Note that we obtain similar results if we choose $k = 3$ instead of $k = 2$. One exception is that the p-value for $\lambda = 0.2$ increases slightly, which leads not to a rejection of the hypothesis.

These results can be analyzed further if we consider the p-value as a function of c_n, i.e., we plot the function $c \mapsto p(c)$. We observe that the shape of this graph depends on the cases $\lambda = 0$, $\lambda > 0$, and $\lambda < 0$. For a GPC ($\lambda = 0$), cf. Figure S.2, the p-value is typically above the 5 % line for $c \in (0, 0.5]$. Even if the curve falls below this line on this range, it normally returns to greater values almost instantly.[1] Opposed to that, the p-value curve has for $\lambda > 0$, cf. Figure S.3, typically some high peaks for small values of c and then falls

[1] For the sake of completeness, we remark that a copula which is not a GPC itself but which is in a δ-neigborhood of a GPC would have a similar p-value curve. However, the point where the graph falls below the 5 % line would be notably smaller than $c = 0.5$.

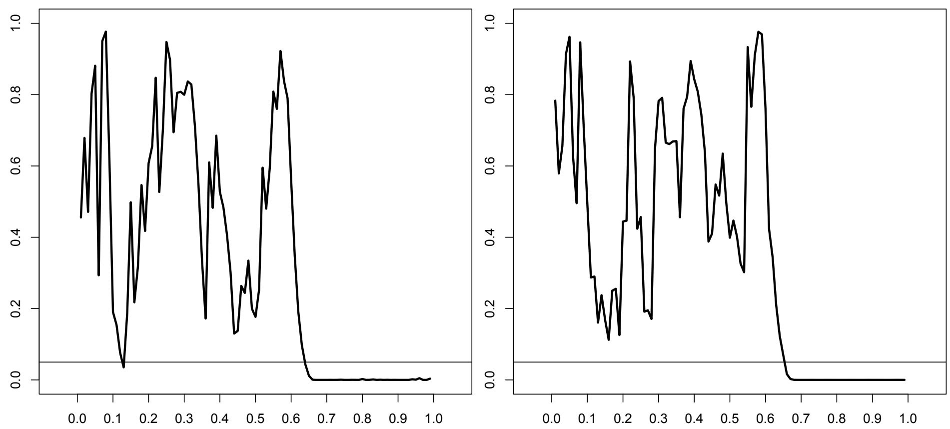

Fig. S.2 p-values as a function of $c \in (0,1)$ for $\lambda = 0$ with $k = 2$ (left) and $k = 3$ (right). The horizontal solid line marks the 5 % type I error level.

below the 5 % line. After another set of peaks for intermediate values of c, the graph normally attains values smaller than 0.05. Contrary to Table S.1, where the case $\lambda = 0.2$ performed better for $k = 2$ than for $k = 3$, Figure S.3 reveals that the curve is somewhat coarse for $k = 2$, whereas it has its expected shape if we put $k = 3$. Moreover, the case $\lambda < 0$, cf. Figure S.4, appears to yield a curve that is, roughly, above the 5 % line on some interval with left endpoint zero and then falls and stays below this line. Although the top line of this figure suggests that the right endpoint of this interval is relatively close to zero as well, the plots for $\lambda = -0.2$ indicate that the interval may also include intermediate values of c, but with some kind of a downward trend. Opposed to that, the curves for $\lambda = 0$ tend to attain large values for c close to 0.5 and then fall below the 5 % line abruptly, cf. Figure S.2.

In order to complement the above results, we also divided for each $\lambda \in \left\{-\frac{\sqrt{2}}{2}, -0.2, 0, 0.2, \frac{\sqrt{2}}{2}\right\}$ the corresponding sample of 10 000 copula observations into 50 subsamples. The resulting sample size of $n = 200$ appears to be too small to obtain stable results since the corresponding p-value curves of the cases $\lambda = 0$, $\lambda > 0$, and $\lambda < 0$ were hardly distinguishable. Moreover, a plot like Figure S.1 showed quite wide 95 % confidence intervals that cover almost the whole range $[1, 2]$; recall that $1 \leq m_D \leq 2$ for bivariate data. However, if we consider the rate of rejection – i.e. the number of subsamples where the δ-neighborhood hypothesis is rejected divided by the total number of subsamples – and plot it as a function of $c \in (0,1)$, cf. Figure S.5, we observe that the test seems to satisfy the type I error level of 5 % for $\lambda = 0$, whereas there is a peak at about $c = 0.11$ for $\lambda = \frac{\sqrt{2}}{2}$. Note that, among the 50 subsamples, the mean number of exceedances above the thresholds $\left(1 - \frac{0.11}{2}\right)\mathbf{1}$ and $\left(1 - \frac{0.11}{3}\right)\mathbf{1}$ were 29.64 and 19.44, respectively, for $\lambda = 0$. This also indicates that there

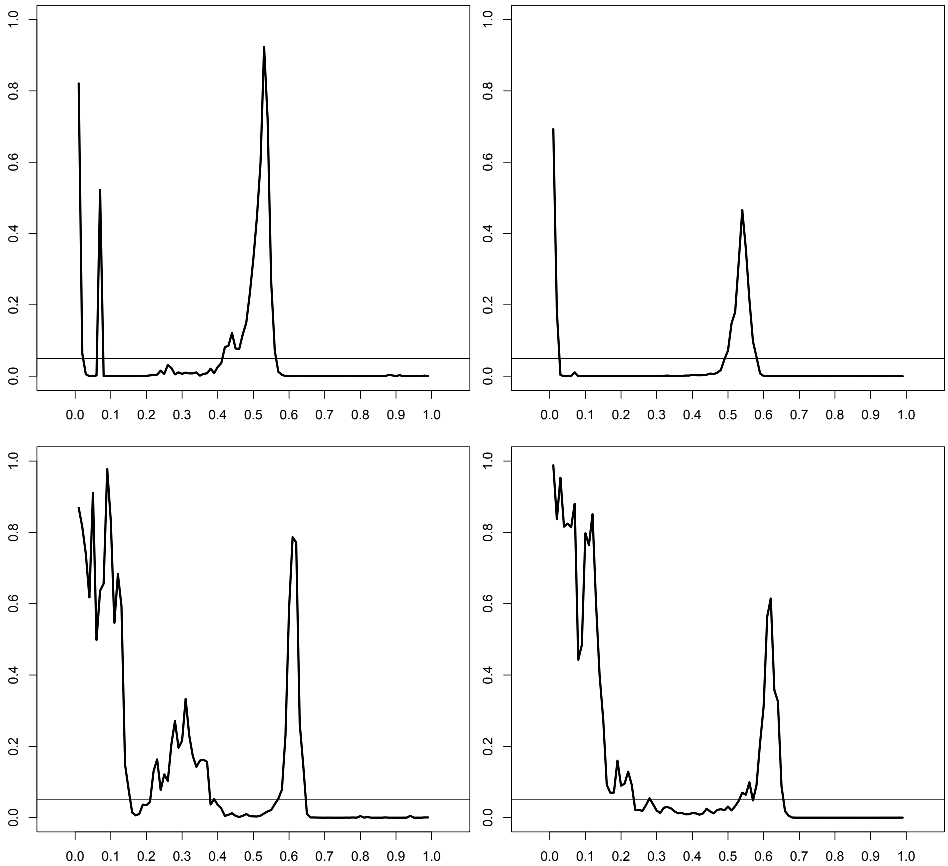

Fig. S.3 p-values as a function of $c \in (0, 1)$ for $\lambda > 0$. Top: $\lambda = \sqrt{2}/2$ with $k = 2$ (left) and $k = 3$ (right). Bottom: $\lambda = 0.2$ with $k = 2$ (left) and $k = 3$ (right).

may be too few observations exceeding the thresholds in order to justify the required approximate normal distribution.

2.2.2 More general data

We close our empirical analysis with some results for data with unknown margins. In order to compare these results with the ones for copula data, we apply the test of Theorem 2 to the same samples as in Section 2.2.1, but we claim that we would *not* know the margins of the data.

As outlined previously, we choose $m_n = 2\,037$ for $n = 10\,000$ and apply essentially the same strategy as in Section 2.2.1. Since we now have to estimate the margins – or merely certain quantiles of the margins – from the data, the test itself is applied to the first m_n observations only, cf. Section 3.2 of the

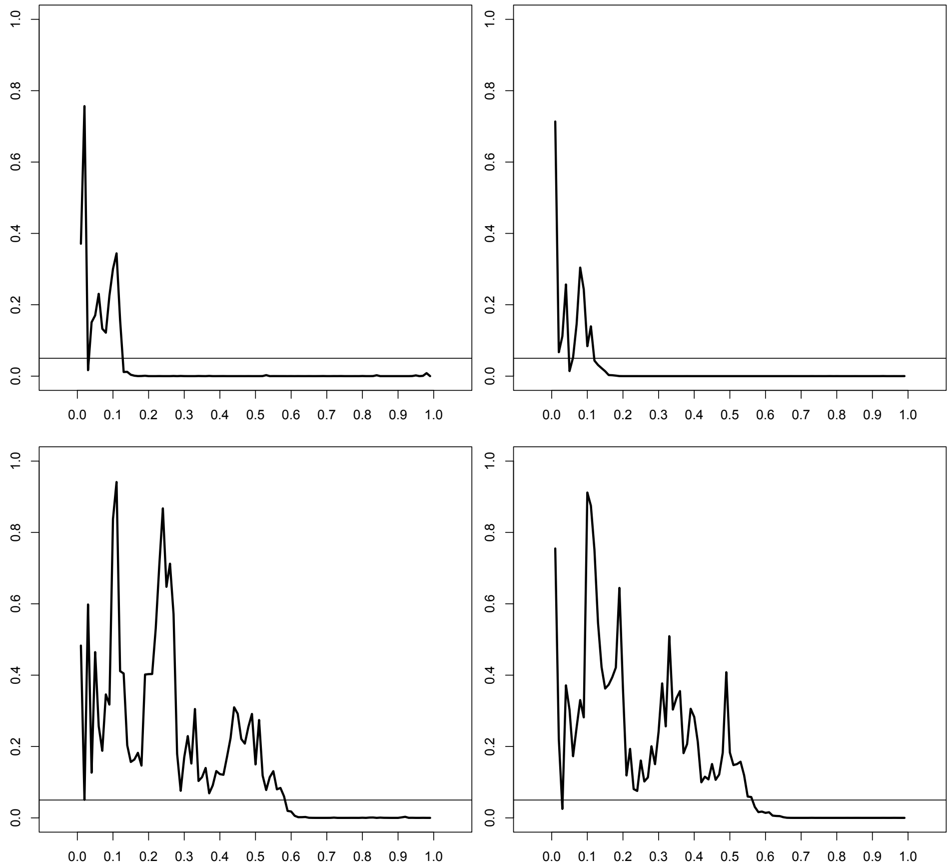

Fig. S.4 p-values as a function of $c \in (0, 1)$ for $\lambda < 0$. Top: $\lambda = -\sqrt{2}/2$ with $k = 2$ (left) and $k = 3$ (right). Bottom: $\lambda = -0.2$ with $k = 2$ (left) and $k = 3$ (right).

main work. Therefore, we expect the following results to be coarser than those discussed above.

By considering Figure S.6, it appears reasonable to choose again $c_{10\,000} = 0.2$ and $k = 2$ or $k = 3$. We observe that the results in Table S.2 are similar to those in Table S.1. The main differences are that the numbers of exceedances of the corresponding thresholds are, of course, smaller and that in the case $\lambda = 0.2$ the alternative is not detected anymore; the p-value is now larger than 0.05. Note that considering the p-value curves for $\lambda = \pm 0.2$ reveals that increasing k may lead to weaker results, cf. Figure S.7, whereas in Section 2.2.1 one could get the impression that greater values of k cause slightly better results. Since the remaining p-value plots had their typical shapes as discussed in Section 2.2.1, we skip them here.

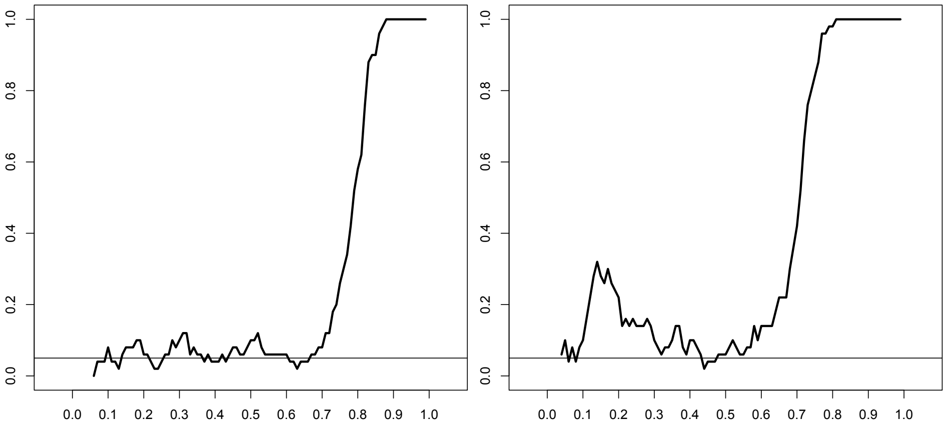

Fig. S.5 Rates of rejection as a function of $c \in (0,1)$ among the test results of the 50 subsamples for $k = 3$. Left: $\lambda = 0$. Right: $\lambda = \sqrt{2}/2$. The plots for $k = 2$ were very similar and are, therefore, omitted.

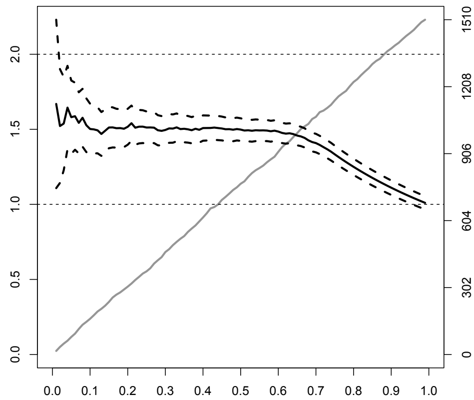

Fig. S.6 Plot of the function $\hat{\gamma}_*^{(m_n)}$ (black, left scale) for $\lambda = 0$, $n = 10\,000$, and $m_n = 2\,037$, together with pointwise approximate 95 % confidence intervals (dashed), cf. (S.5). The dashed horizontal lines mark the lower and the upper bound of the constant m_D, cf. Section 3. The gray line (right scale) displays the the function $c \mapsto \hat{n}_{2,m_n}(c)$, i.e. $k = 2$, cf. (16).

Finally, we again split the sample up into 50 subsamples, each of size 200. As before, we see a slight peak in Figure S.8 for $\lambda = \frac{\sqrt{2}}{2}$ at about $c = 0.11$, which indicates that the test has the tendency to detect the alternative even in relatively small samples. However, this peak is rather small, which is not surprising since we already noted in Section 2.2.1 that this sample size may not be large enough in order to obtain stable results.

Altogether, the tests proposed in Theorem 1 and Theorem 2 perform quite well when they are used to detect a GPC itself. If the copula of the data is not a GPC, the test results are sensitive to a proper selection of c_n, where

Table S.2 Test results for different values of λ and k with $n = 10\,000$, $m_n = 2\,037$, and $c_n = 0.2$.

k	λ	p-value	$\hat{n}_{k,m_n}(0.2)$
2	0.00000	0.80883	306
	0.70711	0.00020	295
	0.20000	0.77657	306
	−0.20000	0.90475	315
	−0.70711	0.00737	364
3	0.00000	0.85423	209
	0.70711	0.00011	183
	0.20000	0.93852	201
	−0.20000	0.84153	205
	−0.70711	0.03447	244

we considered quite a number of graphical tools, whose shapes appear to be a reliable indicator whether the hypothesis is true or not. Where applicable, the choice of m_n, however, appears to be sufficiently solved by the representation in Section 2.1. A great advantage of the p-value curve approach is that a practitioner does not need to specify a suitable value of c_n explicitly, but can make the decision based on a highly intelligible graphical tool. Given that we test for a tail property of a distribution – i.e. a large sample size is needed to detect this property –, the above results suggest that even for relatively small samples, there is at least a tendency of detecting the alternative.

3 Proof of Example 2

Put $\boldsymbol{y} := (\Phi(x_i))_{i=1}^d$. Then we have

$$\begin{aligned} C(\boldsymbol{y}) &= P(\Phi(X_i) \le y_i,\, 1 \le i \le d) \\ &= P(X_i \le x_i,\, 1 \le i \le d) \\ &= 1 - P\left(\bigcup_{i=1}^d \{X_i > x_i\}\right) \\ &= 1 - \sum_{i=1}^d P(X_i > x_i) + \sum_{T \subset \{1,\dots,d\},|T|\ge 2} (-1)^{|T|} P(X_i > x_i,\, i \in T) \\ &= 1 - \|\mathbf{1} - \boldsymbol{y}\|_1 + \sum_{T \subset \{1,\dots,d\},|T|\ge 2} (-1)^{|T|} P(X_i > x_i,\, i \in T) \end{aligned}$$

by the inclusion-exclusion theorem.

By c we denote in what follows a positive generic constant. We have

$$\left| \sum_{T \subset \{1,\dots,d\},|T|\ge 2} (-1)^{|T|} P(X_i > x_i,\, i \in T) \right| \le c \sum_{1 \le i \ne j \le d} P(X_i > x_i, X_j > x_j).$$

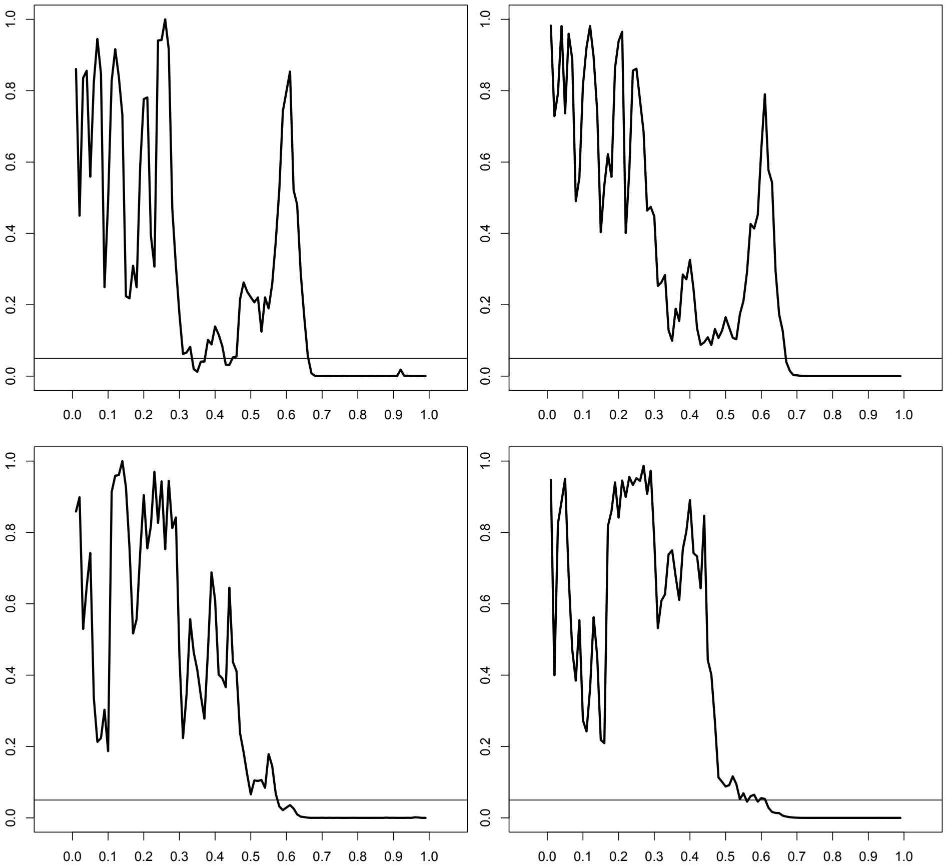

Fig. S.7 p-values as a function of $c \in (0,1)$. Top: $\lambda = 0.2$ with $k = 2$ (left) and $k = 3$ (right). Bottom: $\lambda = -0.2$ with $k = 2$ (left) and $k = 3$ (right).

We will show that for all $i \neq j$

$$\frac{P(X_i > x_i, X_j > x_j)}{\left(\sum_{m=1}^{d}(1 - \Phi(x_m))\right)^{1+\delta}} \le c, \qquad 1 \le i \neq j \le d, \tag{S.6}$$

for $\boldsymbol{x} \geq \boldsymbol{x}_0$, where $\boldsymbol{x}_0 \in \mathbb{R}^d$ is specified later. This, obviously, implies the assertion.

Equation (S.6) is implied by the inequality

$$\frac{P(X_i > x_i, X_j > x_j)^{\frac{1}{1+\delta}}}{1 - \Phi(x_i) + 1 - \Phi(x_j)} \le c, \qquad 1 \le i \neq j \le d, \tag{S.7}$$

for $\boldsymbol{x} \geq \boldsymbol{x}_0$, which we will establish in the sequel.

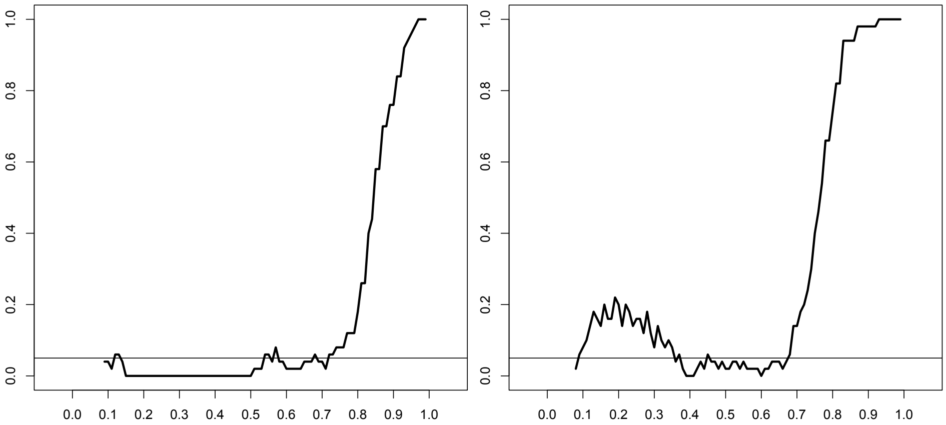

Fig. S.8 Rates of rejection as a function of $c \in (0,1)$ among the test results of the 50 subsamples for $k = 3$, $n = 200$, and $m_n = 86$. Left: $\lambda = 0$. Right: $\lambda = \sqrt{2}/2$. The plots for $k = 2$ were very similar and are, therefore, omitted.

Fix $i \neq j$. To ease the notation we put $X := X_i$, $Y := X_j$, $x := x_i$, $y := x_i$, $\rho := \rho_{ij}$. The covariance matrix of $(X,Y)^\intercal$ is $\boldsymbol{\Sigma}_{X,Y} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, its inverse is $\boldsymbol{\Sigma}_{X,Y}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ and, hence,

$$\boldsymbol{\Sigma}_{X,Y}^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{1-\rho^2}\begin{pmatrix} x - \rho y \\ y - \rho x \end{pmatrix} > \mathbf{0}$$

if $x, y > 0$; recall that $\rho < 0$. From Savage (1962) (see also Tong (1990) and Hashorva and Hüsler (2003)) we obtain the bound

$$P(X > x, Y > y) \leq c\frac{1}{(x-\rho y)(y-\rho x)}\exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right), \quad x, y > 0. \tag{S.8}$$

By the obvious inequality

$$\delta = \min_{1 \leq k \neq m \leq d} \frac{\rho_{km}^2}{1-\rho_{km}^2} \leq \frac{\rho^2}{1-\rho^2}$$

we obtain

$$\frac{1}{1+\delta} \geq 1 - \rho^2$$

and, thus, equation (S.8) implies

$$P(X > x, Y > y)^{\frac{1}{1+\delta}} \leq c\frac{1}{((x-\rho y)(y-\rho x))^{1-\rho^2}}\exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right).$$

From the fact that $1-\Phi(x)\sim\varphi(x)/x$ as $x\to\infty$, where $\varphi=\Phi'$ denotes the standard normal density, we obtain for $x,y\ge x_0$

$$\begin{aligned}\frac{P(X>x,Y>y)^{\frac{1}{1+\delta}}}{1-\Phi(x)+1-\Phi(y)}&\le c\frac{x\exp\left(\frac{x^2}{2}\right)+y\exp\left(\frac{y^2}{2}\right)}{((x-\rho y)(y-\rho x))^{1-\rho^2}\exp\left(\frac{x^2-2\rho xy+y^2}{2}\right)}\\&=c\frac{x\exp\left(-\frac{y^2}{2}\right)+y\exp\left(-\frac{x^2}{2}\right)}{((x-\rho y)(y-\rho x))^{1-\rho^2}\exp(-\rho xy)}\\&\le c\frac{x\exp\left(-\frac{y^2}{2}\right)+y\exp\left(-\frac{x^2}{2}\right)}{(xy)^{1-\rho^2}\exp(-\rho xy)}\\&\le c\frac{(xy)^{\rho^2}}{\exp(-\rho xy)}\\&\le c;\end{aligned}$$

recall that $\rho<0$. This implies equation (S.7) and, thus, the assertion. □

4 Proof of Lemma 2

Analogously to Lemma 1, we assume $j=1$ without loss of generality and obtain for arbitrary $\eta>0$ that

$$\begin{aligned}&P\left(\left|\frac{n_{1,m_n}(c_n)-\hat n_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}\right|>\eta\right)\\&\le P\left(\frac{n_{1,m_n}(c_n)-\hat n_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}>\eta,\ \left|U_{\langle n(1-c_n)\rangle:n}-\mu_n\right|\le\varepsilon_n\right)\\&\quad+P\left(\frac{n_{1,m_n}(c_n)-\hat n_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}<-\eta,\ \left|U_{\langle n(1-c_n)\rangle:n}-\mu_n\right|\le\varepsilon_n\right)\\&\quad+2\,P\left(\left|U_{\langle n(1-c_n)\rangle:n}-\mu_n\right|>\varepsilon_n\right)\end{aligned}$$

where $\mu_n:=E\big(U_{\langle n(1-c_n)\rangle:n}\big)=\langle n(1-c_n)\rangle/(n+1)$ and ε_n is given as in the proof of Lemma 1. Again, Reiss (1989, Lemma 3.1.1) shows that the last term has limit 0 as $n\to\infty$. Moreover, the first term satisfies for large n

$$\begin{aligned}&P\left(\frac{n_{1,m_n}(c_n)-\hat n_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}>\eta,\ \left|U_{\langle n(1-c_n)\rangle:n}-\mu_n\right|\le\varepsilon_n\right)\\&\le P\left(\sum_{i=1}^{m_n}1_{\left\{\boldsymbol{U}^{(i)}\not\le(1-c_n)1_{[0,1]}\right\}}-\sum_{i=1}^{m_n}1_{\left\{\boldsymbol{U}^{(i)}\not\le(\mu_n+\varepsilon_n)1_{[0,1]}\right\}}>(m_nc_n)^{1/2}\eta\right)\\&=1-P\left(\sum_{i=1}^{m_n}1_{\left\{\boldsymbol{U}^{(i)}\le(\mu_n+\varepsilon_n)1_{[0,1]}\right\}\setminus\left\{\boldsymbol{U}^{(i)}\le(1-c_n)1_{[0,1]}\right\}}\le(m_nc_n)^{1/2}\eta\right)\\&\to0\quad\text{as}\quad n\to\infty\end{aligned}$$

by (33), (34), the functional δ-neighborhood condition, and the same arguments as in Lemma 1. Analogously, one shows that the second term vanishes asymptotically. □

5 Proof of Lemma 3

Analogously to Lemma 1, we assume $j = 1$ without loss of generality. As before, Reiss (1989, Lemma 3.1.1) yields

$$\begin{aligned}
&P\left(\max_{1\le r\le d_n}\left|U_{\langle n(1-c_n)\rangle:n,r}-\mu_n\right|>\varepsilon_n\right)\\
&\le d_n\exp\left(-\frac{n}{c_n}\varepsilon_n^2\cdot\frac{1}{3}\big(1+o(1)\big)\right)\\
&=\exp\left(\log(d_n)\left(1-\frac{n\varepsilon_n^2}{c_n\log(d_n)}\cdot\frac{1}{3}\big(1+o(1)\big)\right)\right) \qquad \text{(S.9)}
\end{aligned}$$

as $n \to \infty$, where $\mu_n := E\big(U_{\langle n(1-c_n)\rangle:n,1}\big) = \langle n(1-c_n)\rangle/(n+1)$ and $\varepsilon_n := \big(\frac{c_n}{m_n}\delta_n^{1/2}\big)^{1/2}$ with δ_n as in the proof of Lemma 1. Since (18) shows

$$\frac{n\varepsilon_n^2}{c_n\log(d_n)} \ge \frac{\delta_n^{-1/2}}{\log(d_n)} = \left(\max\left\{\frac{m_n^{1/2}}{n^{1/2}}, \frac{1}{(m_nc_n)^{1/2}}\right\}\log(d_n)\right)^{-1},$$

we obtain from (39) that (S.9) has limit 0 as $n \to \infty$.

Due to (37) and (38), we have for large n and arbitrary $\eta > 0$

$$\begin{aligned}
&P\left(\frac{n_{1,m_n}(c_n)-\hat{n}_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}>\eta,\ \max_{1\le r\le d_n}\left|U_{\langle n(1-c_n)\rangle:n,r}-\mu_n\right|\le\varepsilon_n\right)\\
&\le 1-P\left(\sum_{i=1}^{m_n}1_{[\mathbf{0},(\mu_n+\varepsilon_n)\mathbf{1}]\setminus[\mathbf{0},(1-c_n)\mathbf{1}]}\big(\boldsymbol{U}_{d_n}^{(i)}\big)\le(m_nc_n)^{1/2}\eta\right) \qquad \text{(S.10)}
\end{aligned}$$

where (35) and (36) show

$$\begin{aligned}
&P\Big(\boldsymbol{U}_{d_n}^{(i)}\in[\mathbf{0},(\mu_n+\varepsilon_n)\mathbf{1}]\setminus[\mathbf{0},(1-c_n)\mathbf{1}]\Big)\\
&=\varepsilon_n\left[\left(1+O\left(\frac{1}{n\varepsilon_n}\right)\right)m_{D,d_n}+O\left(\frac{c_n^{1+\delta}}{\varepsilon_n}\right)\right]\\
&\sim\varepsilon_n m_D \quad \text{as} \quad n\to\infty.
\end{aligned}$$

By the arguments in the proof of Lemma 1, (S.10) has limit 0 as $n \to \infty$. Similarly, one also shows

$$P\left(\frac{n_{1,m_n}(c_n)-\hat{n}_{1,m_n}(c_n)}{(m_nc_n)^{1/2}}<-\eta,\ \max_{1\le r\le d_n}\left|U_{\langle n(1-c_n)\rangle:n,r}-\mu_n\right|\le\varepsilon_n\right)\to 0$$

for arbitrary $\eta > 0$ as $n \to \infty$, which implies the assertion, cf. (19). □

References

Aulbach S, Bayer V, Falk M (2012) A multivariate piecing-together approach with an application to operational loss data. *Bernoulli* 18(2):455–475, DOI 10.3150/10-BEJ343

Aulbach S, Falk M, Hofmann M (2013) On max-stable processes and the functional D-norm. *Extremes* 16(3):255–283, DOI 10.1007/s10687-012-0160-3

Duchesne P, Lafaye de Micheaux P (2010) Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics & Data Analysis* 54:858–862, DOI 10.1016/j.csda.2009.11.025

Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling Extremal Events for Insurance and Finance*, Applications of Mathematics - Stochastic Modelling and Applied Probability, vol 33. Springer, Berlin, DOI 10.1007/978-3-642-33483-2

Hashorva E, Hüsler J (2003) On multivariate Gaussian tails. *Annals of the Institute of Statistical Mathematics* 55(3):507–522, DOI 10.1007/BF02517804

Imhof JP (1961) Computing the distribution of quadratic forms in normal variables. *Biometrika* 48(3-4):419–426, DOI 10.1093/biomet/48.3-4.419

Reiss RD (1989) *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer Series in Statistics, Springer, New York, DOI 10.1007/978-1-4613-9620-8

Savage IR (1962) Mills' ratio for multivariate normal distributions. *Journal of Research of the National Bureau of Standards, Section B: Mathematics and Mathematical Physics* 66B(3):93–96, DOI 10.6028/jres.066B.011

Tong YL (1990) *The Multivariate Normal Distribution*. Springer Series in Statistics, Springer, New York, DOI 10.1007/978-1-4613-9655-0