

EMPIRICAL TAIL COPULAS FOR FUNCTIONAL DATA

J. Einmahl and J. Segers

LIDAM Discussion Paper ISBA/2021/03

EMPIRICAL TAIL COPULAS FOR FUNCTIONAL DATA

BY JOHN H.J. EINMAHL¹ AND JOHAN SEGERS²¹*Department of Econometrics and OR and CentER, Tilburg University, j.h.j.einmahl@uvt.nl*²*LIDAM/ISBA, UCLouvain, joan.segers@uclouvain.be*

For multivariate distributions in the domain of attraction of a max-stable distribution, the tail copula and the stable tail dependence function are equivalent ways to capture the dependence in the upper tail. The empirical versions of these functions are rank-based estimators whose inflated estimation errors are known to converge weakly to a Gaussian process that is similar in structure to the weak limit of the empirical copula process. We extend this multivariate result to continuous functional data by establishing the asymptotic normality of the estimators of the tail copula, uniformly over all finite subsets of at most D points (D fixed). An application for testing tail copula stationarity is presented. The main tool for deriving the result is the uniform asymptotic normality of all the D -variate tail empirical processes. The proof of the main result is non-standard.

1. Introduction. Consider the statistical theory of extreme values for functional data taking values in $C([0, 1])$, the space of continuous functions on the interval $[0, 1]$. Max-stability in $C([0, 1])$ was characterized in [Giné, Hahn and Vatan \(1990\)](#) and the corresponding domain of attraction conditions have been studied in [de Haan and Lin \(2001\)](#). Consistency of extreme value index estimator functions and of estimators of the exponent measure was established in [de Haan and Lin \(2003\)](#) and asymptotic normality of extreme value index estimator functions was obtained in [Einmahl and Lin \(2006\)](#).

In this paper we extend the setup of the latter two papers by replacing the interval $[0, 1]$ (with absolute value metric) with a general, compact metric space (T, ρ) . Interesting choices for T can be the unit square, the unit cube, or the unit hypercube in any finite dimension as well as a general, compact set in dimension two, three, or higher. Statistical inference is based on independent and identically distributed data that are random functions with values in $C(T)$. Now consider multivariate extreme value theory by restricting the data to a finite subset of d points in T . We assume the existence of the d -variate tail copula R and the stable tail dependence function l , each of which determines the tail dependence structure of the multivariate data, see, e.g., [Drees and Huang \(1998\)](#), [Schmidt and Stadtmüller \(2006\)](#), [Peng and Qi \(2008\)](#), [Nikoloulopoulos, Joe and Li \(2009\)](#) and [Bücher and Dette \(2013\)](#). The asymptotic normality, uniform in the argument, of the usual rank-based estimator of l can be found in, e.g., [Einmahl, Krajina and Segers \(2012\)](#).

It is the main purpose of this paper to extend the latter multivariate result to random functions in $C(T)$ by considering the asymptotic normality of the rank-based estimators of the tail copula R , uniformly over all finite subsets of T of at most D points (D fixed). To the best of our knowledge this is the first paper that establishes asymptotic normality of estimators of the tail dependence structure for functional data in this general, nonparametric setting. Note that as a special case we obtain the uniform asymptotic normality of all estimated upper tail dependence coefficients (corresponding to two points in T). The main tool for the main result

MSC2020 subject classifications: Primary 62M99, 62G05, 62G20, 62G30, 62G32; secondary 60F17, 60G70. JEL codes: C13, C14.

Keywords and phrases: extreme value statistics, functional data, tail empirical process, tail dependence, tail copula estimation, uniform asymptotic normality.

is the uniform asymptotic normality of the multivariate tail empirical processes, defined for all D -tuples in T . This result is of independent interest and generalizes a similar result for $D = 1$ (and $T = [0, 1]$) in [Einmahl and Lin \(2006\)](#).

As in the multivariate case, the limiting process for the tail copula estimators requires the existence and continuity of the partial derivatives of the tail copula which yields an unexpected, challenging problem for functional data. The problem arises since points in T that are close have, because of strong dependence due to continuity, a bivariate tail copula that is close to the non-differentiable “comonotone” tail copula, being the minimum of the two coordinates. This phenomenon requires a novel, non-standard proof-technique for the main result.

The fact that our results are uniform over T sets them apart from those that fix a finite vector of locations in T . Thanks to the uniformity, it becomes possible to treat functionals that depend on all $t \in T$ simultaneously, such as suprema and integrals. By way of motivation, some possible applications of our results are sketched below.

- In order to detect non-stationarity of the extreme-value attractor of a random field over T , a subset of Euclidean space, [Dombry \(2017\)](#) suggests to compute an integral $\hat{I}(t_0)$ of estimated tail dependence coefficients at pairs $(t_0, t) \in T^2$, for fixed t_0 and integrating over t in a neighbourhood of t_0 . Stationarity then yields that the function $t_0 \mapsto \hat{I}(t_0)$ is about constant. Our main result yields the asymptotic behavior of $\hat{I}(t_0)$. We elaborate on this idea in Section 4 where we present as a test statistic the range of a discrete version of these integrals, assuming that we observe the functional data only on a fine, but finite grid on $T = [0, 1]$. A similar integral, but then based on the notion of concurrence probabilities, is used in [Dombry, Ribatet and Stoev \(2018\)](#) to illustrate the spatial structure of extremes.
- [Koch \(2017\)](#) introduces spatial risk measures, some of which can be expressed as double integrals of pairwise extremal coefficients over pairs of locations. Our weak convergence theory yields the asymptotic normality of nonparametric estimates of these risk measures.
- Matching nonparametric tail copula estimators to a parametric form yields minimum-distance estimators of parameters of tail dependence models. For multivariate data, such estimators are proposed and analyzed in [Einmahl, Krajina and Segers \(2012\)](#) and [Einmahl, Kiriliouk and Segers \(2018\)](#). Our uniform central limit theorem implies the asymptotic normality of such estimators in case of functional data.
- The goodness-of-fit of a parametric family of spatial tail dependence models may be assessed via a supremum over $(t_1, \dots, t_D) \in T^D$ of Kolmogorov–Smirnov type test statistics based on the difference between a nonparametric and a parametric estimator of the D -variate tail copula. Our results not only yield the asymptotic distribution of the test statistic under the null hypothesis, they even cover the setting of infill asymptotics, where the process is observed only on a finite, possibly random set $T_n \subset T$ of locations, with the property that T_n becomes dense in T as n grows (formally, for every open subset $G \subset T$, the probability that T_n intersects G increases to 1 as $n \rightarrow \infty$).

The paper is organized as follows. The joint asymptotic behavior of all the multivariate tail empirical processes is studied in Section 2. The main result on the joint asymptotic behavior of all the empirical tail copulas is presented in Section 3. The application to testing stationarity of the tail copula is developed in Section 4. The proofs of the results in Sections 2 and 3 are deferred to Sections 5 and 6, respectively. Some auxiliary results are collected in Appendices A and B.

2. Tail empirical processes. Let (T, ρ) be a compact metric space; in [Einmahl and Lin \(2006\)](#), the space T is $[0, 1]$ with the absolute value metric. Let $C(T)$ be the space of continuous functions $f : T \rightarrow \mathbb{R}$. Actually the metric ρ on T does not play much of a role. It is merely introduced to have a complete, separable subset of $\ell^\infty(T)$, the space of bounded

functions $f : T \rightarrow \mathbb{R}$, facilitating the study of stochastic processes on T . Let $C(T)$ be equipped with the supremum norm and distance and the associated Borel σ -field.

Let $\xi_1, \xi_2, \dots$ be i.i.d. random elements in $C(T)$. For $(t, x) \in T \times \mathbb{R}$, define $F_t(x) = \mathbb{P}\{\xi_i(t) \leq x\}$. Since the trajectories of ξ_i are continuous, the map $t \mapsto F_t$ is continuous with respect to the topology of weak convergence, that is, if $t_n \rightarrow t$ in T then $F_{t_n} \rightarrow F_t$ weakly as $n \rightarrow \infty$. We will consider the (inverse) probability integral transform at each coordinate $t \in T$, and this requires the following condition.

CONDITION 1. *The distribution functions F_t , $t \in T$, are continuous.*

Under Condition 1, the random variables $U_i(t) = 1 - F_t(\xi_i(t))$ for $i = 1, 2, \dots$ and $t \in T$ are uniformly distributed on $(0, 1)$. Moreover, the map $t \mapsto F_t$ is now continuous with respect to the supremum distance: if $t_n \rightarrow t$ in T then $\sup_{x \in \mathbb{R}} |F_{t_n}(x) - F_t(x)| \rightarrow 0$ as $n \rightarrow \infty$. The stochastic processes $U_1, U_2, \dots$ are i.i.d. random elements of $C(T)$ too. Indeed, we have $U_i = 1 - F(\xi_i)$ where $F : C(T) \rightarrow C(T)$ is defined by $(F(x))(t) = F_t(x(t))$ for $x \in C(T)$ and $t \in T$. The map F is continuous and thus Borel measurable.

For $d = 1, 2, \dots$ and $\mathbf{t} = (t_1, \dots, t_d) \in T^d$, define the d -variate empirical distribution function $S_{n,\mathbf{t}}$ by

$$(1) \quad S_{n,\mathbf{t}}(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) < u_j\}$$

for $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$. Note that $S_{n,\mathbf{t}}$ involves the marginal distribution functions F_{t_j} , which are unknown. In Section 3 we will replace these functions by their empirical counterparts. But to study the resulting statistic, we first need to consider the case where the functions F_{t_j} are known.

Fix $D = 1, 2, \dots$ and $M > 0$ and put $\mathcal{V} = \bigcup_{d=1}^D \mathcal{V}_d$ where $\mathcal{V}_d = T^d \times [0, M]^d$ for $d = 1, \dots, D$. Let $k = k_n \in (0, n]$ and consider

$$(2) \quad G_{n,\mathbf{t}}(\mathbf{x}) = \frac{n}{k} S_{n,\mathbf{t}}\left(\frac{k}{n}\mathbf{x}\right) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) < \frac{k}{n}x_j\},$$

$$(3) \quad E_{n,\mathbf{t}}(\mathbf{x}) = \mathbb{E}[G_{n,\mathbf{t}}(\mathbf{x})] = \frac{n}{k} \mathbb{P}\{\forall j = 1, \dots, d : U_i(t_j) < \frac{k}{n}x_j\}$$

for $\mathbf{v} = (\mathbf{t}, \mathbf{x}) \in \mathcal{V}$. The tail empirical process W_n indexed by $\mathcal{V}$ is defined by

$$W_n(\mathbf{v}) = \sqrt{k} \{G_{n,\mathbf{t}}(\mathbf{x}) - E_{n,\mathbf{t}}(\mathbf{x})\}, \quad \mathbf{v} = (\mathbf{t}, \mathbf{x}) \in \mathcal{V}.$$

CONDITION 2. *The sequence $k = k_n \in (0, n]$ is such that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$.*

We will investigate weak convergence of W_n in $\ell^\infty(\mathcal{V})$. To control its finite-dimensional distributions, we will assume the following tail dependence condition. Note that we need to specify it up to dimension $2D$ in order to control the relevant covariances.

CONDITION 3. *For all $m = 2, \dots, 2D$ and all $(\mathbf{t}, \mathbf{x}) \in T^m \times [0, \infty)^m$, the following limit exists:*

$$R_{\mathbf{t}}(\mathbf{x}) = \lim_{s \downarrow 0} s^{-1} \mathbb{P}\{U_1(t_1) \leq sx_1, \dots, U_1(t_m) \leq sx_m\}.$$

Observe that for $m = 1$, we have $R_t(x) = x$ for all $(t, x) \in T \times [0, \infty)$. For $D = 1, 2, \dots$, the function $R_t : ([0, \infty]^D \setminus \{\infty\}) \rightarrow [0, \infty)$ is called the (lower and upper, respectively) tail copula of the random vectors $(U_1(t_1), \dots, U_1(t_D))$ and $(\xi_1(t_1), \dots, \xi_1(t_D))$. It is the main subject of this paper. The tail copula, together with the marginal distributions, determines a multivariate extreme-value distribution G , say. In other words, it determines the dependence structure of G and hence the tail dependence structure of a probability distribution that is in the max-domain of attraction of G . More explanations on and various properties and representations of the tail copula and the closely related stable tail dependence function defined in Remark 3 below are given in, e.g., [Drees and Huang \(1998\)](#), [Nikoloulopoulos, Joe and Li \(2009\)](#), [Ressel \(2013\)](#), [Mercadier and Rostant \(2019\)](#), and [Falk \(2019\)](#).

The process W_n is centered. Let us calculate its covariance function. Write $U_1(\mathbf{t}) = (U_1(t_1), \dots, U_1(t_m))$ for $\mathbf{t} \in T^m$ and let inequalities between vectors be meant coordinatewise. For $\mathbf{v}_j = (\mathbf{t}_j, \mathbf{x}_j) \in \mathcal{V}$, with $j = 1, 2$, we have

$$\begin{aligned} & \mathbb{E}[W_n(\mathbf{v}_1)W_n(\mathbf{v}_2)] \\ &= \frac{n}{k} [\mathbb{P}\{U_1(\mathbf{t}_1, \mathbf{t}_2) \leq \frac{k}{n}(\mathbf{x}_1, \mathbf{x}_2)\} - \mathbb{P}\{U_1(\mathbf{t}_1) \leq \frac{k}{n}\mathbf{x}_1\} \mathbb{P}\{U_1(\mathbf{t}_2) \leq \frac{k}{n}\mathbf{x}_2\}] \\ (4) \quad & \rightarrow R_{\mathbf{t}_1, \mathbf{t}_2}(\mathbf{x}_1, \mathbf{x}_2), \quad n \rightarrow \infty, \end{aligned}$$

in view of Conditions 2 and 3. Note that the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ may have different lengths, say $2d_1$ and $2d_2$, with $d_1, d_2 \in \{1, \dots, D\}$, and that the limit in Eq. (4) involves the tail copula of Condition 3 in dimension $m = d_1 + d_2$.

Let $W = (W(\mathbf{v}))_{\mathbf{v} \in \mathcal{V}}$ be a zero-mean Gaussian process defined on $\mathcal{V}$ with covariance function

$$(5) \quad \mathbb{E}[W(\mathbf{v}_1)W(\mathbf{v}_2)] = R_{\mathbf{t}_1, \mathbf{t}_2}(\mathbf{x}_1, \mathbf{x}_2), \quad \mathbf{v}_j = (\mathbf{t}_j, \mathbf{x}_j) \in \mathcal{V}, \quad j = 1, 2.$$

Because of (4), the function $(\mathbf{v}_1, \mathbf{v}_2) \mapsto R_{\mathbf{t}_1, \mathbf{t}_2}(\mathbf{x}_1, \mathbf{x}_2)$ is the pointwise limit as $n \rightarrow \infty$ of a sequence of covariance functions on $\mathcal{V} \times \mathcal{V}$ and therefore positive semidefinite, so that a Gaussian process W with covariance function in Eq. (5) indeed exists. Further, note that $R_{\mathbf{t}, \mathbf{t}}(\mathbf{x}, \mathbf{x}) = R_{\mathbf{t}}(\mathbf{x}) = \text{var}\{W(\mathbf{v})\}$. The standard deviation semimetric on $\mathcal{V}$ is

$$\begin{aligned} (6) \quad \rho_2(\mathbf{v}_1, \mathbf{v}_2) &= (\mathbb{E}[\{W(\mathbf{v}_1) - W(\mathbf{v}_2)\}^2])^{1/2} \\ &= (R_{\mathbf{t}_1}(\mathbf{x}_1) - 2R_{\mathbf{t}_1, \mathbf{t}_2}(\mathbf{x}_1, \mathbf{x}_2) + R_{\mathbf{t}_2}(\mathbf{x}_2))^{1/2}. \end{aligned}$$

To show the asymptotic tightness of the processes W_n , we need to control the local increments of U_i . The following condition is inspired by [Einmahl and Lin \(2006, Eq. \(16\)\)](#) but is weaker, see Remark 1 below.

CONDITION 4. *There exist positive scalars c_1, c_2, u_0 , and ε_0 , and, for all $\varepsilon \in (0, \varepsilon_0]$, a covering of $T = \bigcup_{j=1}^{N_{T, \varepsilon}} T_{\varepsilon, j}$ by sets $T_{\varepsilon, j}$ such that for all $(u, \varepsilon, j) \in (0, u_0] \times (0, \varepsilon_0] \times \{1, \dots, N_{T, \varepsilon}\}$, we have*

$$(7) \quad \mathbb{P} \left[\sup_{t \in T_{\varepsilon, j}} U_1(t) > u \exp(c_1 \varepsilon^2) \mid \inf_{t \in T_{\varepsilon, j}} U_1(t) \leq u \right] \leq c_2 \varepsilon^2$$

and such that $\int_0^1 \sqrt{\log N_{T, \varepsilon}} d\varepsilon < \infty$.

In (7), we could also replace ε^2 by ε and require $\int_0^1 \sqrt{\log N_{T, \varepsilon^2}} d\varepsilon < \infty$.

THEOREM 1. *Let $\xi_1, \xi_2, \dots$ be i.i.d. random elements in $C(T)$, with (T, ρ) a compact metric space. Suppose Conditions 1 to 4 hold and let $D = 1, 2, \dots$ and $M > 0$. Then the semimetric space $(\mathcal{V}, \rho_2)$ is totally bounded and there exists a version of W of which almost all sample paths are uniformly ρ_2 -continuous. Moreover, we have the weak convergence*

$$W_n \rightsquigarrow W \text{ as } n \rightarrow \infty \text{ in } \ell^\infty(\mathcal{V}).$$

Note that, moreover, the processes W_n are asymptotically uniformly equicontinuous in probability with respect to ρ_2 , see Addendum 1.5.8 and Example 1.5.10 in [van der Vaart and Wellner \(1996\)](#).

So far, the original metric ρ on T has played hardly a role. The covering sets $T_{\varepsilon,j}$ in Condition 4 do not involve ρ and uniform continuity of the sample paths of the limit process W in Theorem 1 is with respect to the standard deviation metric. We can make the link more explicit by requiring that the sets $T_{\varepsilon,j}$ are balls in the original metric ρ .

CONDITION 5. *There exist positive scalars c_1, c_2, u_0 and ε_0 , and, for all $\varepsilon \in (0, \varepsilon_0]$, a scalar $\delta(\varepsilon) > 0$, such that for all $(u, \varepsilon, s) \in (0, u_0] \times (0, \varepsilon_0] \times T$ we have*

$$\mathbb{P} \left[\sup_{t \in T: \rho(s,t) \leq \delta(\varepsilon)} U_1(t) > u \exp(c_1 \varepsilon^2) \mid \inf_{t \in T: \rho(s,t) \leq \delta(\varepsilon)} U_1(t) \leq u \right] \leq c_2 \varepsilon^2.$$

Moreover, if $N_{T,\delta}$ denotes the minimal number of closed balls with radius δ needed to cover T , then we have $\int_0^1 \sqrt{\log N_{T,\delta(\varepsilon)}} d\varepsilon < \infty$.

Condition 5 implies Condition 4; the sets $T_{\varepsilon,j}$ in Condition 4 can be chosen to be the $N_{T,\delta(\varepsilon)}$ balls of radius $\delta(\varepsilon)$ in Condition 5. It is a rather weak condition. For natural examples it is amply satisfied, see Examples 1 and 2 below.

COROLLARY 1. *The conclusions of Theorem 1 continue to hold if Condition 4 is replaced by Condition 5. In that case, there exists for every $\eta > 0$ a scalar $\delta > 0$ such that for all $(s, t) \in T^2$ with $\rho(s, t) < \delta$, we have $1 - R_{s,t}(1, 1) < \eta$. As a consequence, the function $(t, x) \mapsto R_t(x)$ on $T^d \times [0, \infty)^d$ equipped with the product topology induced by ρ and the absolute value metric is jointly continuous.*

REMARK 1 (Increments). We show that for $T = [0, 1]$ equipped with the absolute value metric, Condition 4 is more general than Eq. (16) in [Einmahl and Lin \(2006\)](#) (with $\beta = 0$). Suppose that there exist constants $K, c > 0$ and $q > 1$ such that for all sufficiently small $u, \delta > 0$, we have, for all $s \in [0, 1]$,

$$(8) \quad \mathbb{P} \left[\sup_{t \in [s, s+\delta]} \frac{|U_1(t) - U_1(s)|}{U_1(t)} > K \{\log(\delta^{-1})\}^{-q} \mid \inf_{t \in [s, s+\delta]} U_1(t) \leq u \right] \leq c \{\log(\delta^{-1})\}^{-q}.$$

Eq. (16) in [Einmahl and Lin \(2006\)](#) implies Eq. (8) with $q = 2$. We show that Eq. (8) implies Condition 4. For $\varepsilon > 0$, put $\delta = \delta(\varepsilon) = \exp(-\varepsilon^{-2/q})$ and consider the covering $T_{\varepsilon,j} = [(j-1)\delta, j\delta] \cap [0, 1]$ for $j = 1, \dots, \lceil 1/\delta \rceil$. The number of covering sets is $N_{T,\varepsilon} = \lceil 1/\delta \rceil = \lceil \exp(\varepsilon^{-2/q}) \rceil$, so that $\sqrt{\log N_{T,\varepsilon}} = O(\varepsilon^{-1/q})$ as $\varepsilon \downarrow 0$, which is integrable near zero since $q > 1$. By definition, we have $\{\log(1/\delta)\}^{-q} = \varepsilon^2$. If there exists $\eta > 0$ such

that $\sup_{t \in [s, s+\delta]} |U_1(t) - U_1(s)| / U_1(t) \leq \eta$, then $U_1(s)/(1+\eta) \leq U_1(t) \leq U_1(s)/(1-\eta)$ for all $t \in [s, s+\delta]$, and thus $\sup_{t \in [s, s+\delta]} U_1(t) \leq \frac{1+\eta}{1-\eta} \inf_{t \in [s, s+\delta]} U_1(t)$. It follows that

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [s, s+\delta]} U_1(t) > \frac{1+\eta}{1-\eta} u \mid \inf_{t \in [s, s+\delta]} U_1(t) \leq u \right] \\ \leq \mathbb{P} \left[\sup_{t \in [s, s+\delta]} \frac{|U_1(t) - U_1(s)|}{U_1(t)} > \eta \mid \inf_{t \in [s, s+\delta]} U_1(t) \leq u \right]. \end{aligned}$$

Setting $\eta = K\varepsilon^2$, we have $\exp(c_1\varepsilon^2) \geq \frac{1+\eta}{1-\eta}$ for $c_1 = 3K$ and all $\varepsilon > 0$ small enough. Condition 4 follows. (Actually the stronger Condition 5 also readily follows, essentially by relabeling $[s, s+\delta]$ by $[s-\delta, s+\delta]$.)

REMARK 2 (Function spaces). Rather than with the space $C(T)$ of continuous functions on a compact metric space (T, ρ) , we could work with the space of uniformly continuous functions on a totally bounded semimetric space. However, by [van der Vaart \(1998, Example 18.7\)](#), this would not really be more general, since we could always consider the completion of the space, which is compact, and extend the functions appropriately. To pass from a semimetric (allowing distinct points to be at distance zero) to a genuine metric can be achieved by considering the appropriate quotient space.

We could even work with stochastic processes on T whose common distribution is a tight Borel measure on $\ell^\infty(T)$, the space of bounded functions $f : T \rightarrow \mathbb{R}$. However, by [van der Vaart and Wellner \(1996, Addendum 1.5.8\)](#), we could then find a semimetric ρ on T such that (T, ρ) is totally bounded and such that the processes have uniformly ρ -continuous trajectories. By the previous paragraph, this could again be reduced to the case $C(T)$ after appropriate identifications.

3. Rank-based empirical tail copulas. Since we do not know the marginal distribution functions F_t , we cannot compute $S_{n,t}$ in Eq. (1) from the data. Replacing F_t by the marginal empirical distribution function of the random variables $\xi_1(t), \dots, \xi_n(t)$ yields a rank-based estimator of the tail copula. Let

$$(9) \quad \text{rank}_{n,i}(t) = \sum_{\alpha=1}^n \mathbb{1}\{\xi_\alpha(t) \leq \xi_i(t)\}$$

denote the rank of $\xi_i(t)$ among $\xi_1(t), \dots, \xi_n(t)$. The empirical tail copula is defined as

$$(10) \quad \hat{R}_{n,t}(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : \text{rank}_{n,i}(t_j) > n - kx_j + 1\}$$

for d -tuples of coordinates $\mathbf{t} \in T^d$ and for $\mathbf{x} \in [0, \infty)^d$. The empirical tail copula is a rank-based version of the random function $(\mathbf{t}, \mathbf{x}) \mapsto G_{n,t}(\mathbf{x})$ in Eq. (2). Using $E_{n,t}$ in Eq. (3) for centering, the penultimate empirical tail copula process is defined as

$$\hat{W}_n(\mathbf{v}) = \sqrt{k}\{\hat{R}_{n,t}(\mathbf{x}) - E_{n,t}(\mathbf{x})\}, \quad \mathbf{v} = (\mathbf{t}, \mathbf{x}) \in \mathcal{V}.$$

For $\mathbf{t} \in T^d$ and $j \in \{1, \dots, d\}$, let $\dot{R}_{\mathbf{t},j}(\mathbf{x})$ denote the partial derivative of $R_{\mathbf{t}}$ at $\mathbf{x}$ with respect to $x_j > 0$, provided the partial derivative exists. The standard deviation semimetric ρ_2 on $\mathcal{V}$ in Eq. (6) induces a semimetric $\rho_{2,T}$ on T via

$$(11) \quad \rho_{2,T}(s, t) = \rho_2((s, 1), (t, 1)) = [2\{1 - R_{s,t}(1, 1)\}]^{1/2}, \quad (s, t) \in T^2.$$

The semimetric space $(T, \rho_{2,T})$ is totally bounded since $(\mathcal{V}, \rho_2)$ is totally bounded by Theorem 1. Note that this means that for every $\varepsilon > 0$, there exists a finite set $T(\varepsilon) = \{t_1, \dots, t_n\} \subset T$ such that for every $t \in T$ there exists $t_j \in T(\varepsilon)$ such that $R_{t,t_j}(1, 1) \geq 1 - \varepsilon$. The semimetric $\rho_{2,T}$ is a metric if and only if $R_{s,t}(1, 1) < 1$ whenever $s \neq t$. The (semi)metric space $(T, \rho_{2,T})$ is complete if for every sequence $(t_n)_n$ in T with the property that $\lim_{n \rightarrow \infty} \inf_{k, \ell \geq n} R_{t_k, t_\ell}(1, 1) = 1$ there exists $t \in T$ such that $\lim_{n \rightarrow \infty} R_{t_k, t}(1, 1) = 1$.

CONDITION 6. *The tail copulas satisfy the following properties:*

- (R1) $R_{s,t}(1, 1) < 1$ for all $(s, t) \in T^2$ such that $s \neq t$;
- (R2) the metric space $(T, \rho_{2,T})$ is complete;
- (R3) for all $d = 1, \dots, D$, for all $j = 1, \dots, d$, and for all $\mathbf{t} \in T^d$ such that $t_\alpha \neq t_\beta$ for $\alpha \neq \beta$, the partial derivative $\dot{R}_{\mathbf{t},j}$ exists and is continuous on the set of $\mathbf{x} \in [0, \infty)^d$ such that $x_j > 0$.

If $x_j = 0$, then $\dot{R}_{\mathbf{t},j}(\mathbf{x})$ is defined as the right-hand partial derivative, which exists since $R_{\mathbf{t}}$ is a concave function on $[0, \infty)^d$.

We will need to control the difference between the penultimate tail copula $E_{n,\mathbf{t}}(\mathbf{x})$ and the tail copula $R_{\mathbf{t}}(\mathbf{x})$ itself. Define $E_{n,\mathbf{t}}(\mathbf{x}) = 0 = R_{\mathbf{t}}(\mathbf{x})$ if one or more coordinates of $\mathbf{x}$ are negative.

CONDITION 7. *For all $d = 1, 2, \dots, D$ and all $K, M > 0$, as $n \rightarrow \infty$,*

$$\sup_{\substack{\mathbf{t} \in T^d, \mathbf{x} \in [0, M]^d \\ \mathbf{h} \in [-K, K]^d}} \left| \sqrt{k} \left\{ E_{n,\mathbf{t}} \left(\mathbf{x} + \frac{1}{\sqrt{k}} \mathbf{h} \right) - E_{n,\mathbf{t}}(\mathbf{x}) \right\} - \sqrt{k} \left\{ R_{\mathbf{t}} \left(\mathbf{x} + \frac{1}{\sqrt{k}} \mathbf{h} \right) - R_{\mathbf{t}}(\mathbf{x}) \right\} \right| \rightarrow 0.$$

It is notationally convenient, natural, and not restrictive to consider only tuples of points in T^d with different coordinates. For given $d = 1, 2, \dots$ and given $M > 0$, let $\mathcal{V}'_d$ be the set of points $(\mathbf{t}, \mathbf{x}) \in T^d \times [0, M]^d$ such that all d coordinates of $\mathbf{t}$ are different. As in Theorem 1 we fix $D = 1, 2, \dots$, and write $\mathcal{V}' = \bigcup_{d=1}^D \mathcal{V}'_d$.

THEOREM 2. *Consider the setting and conditions of Theorem 1. Assume in addition that Conditions 6 and 7 hold. Then for all $d = 1, 2, \dots, D$,*

$$(12) \quad \sup_{\mathbf{v} \in \mathcal{V}'_d} \left| \hat{W}_n(\mathbf{v}) - W_n(\mathbf{v}) + \sum_{j=1}^d \dot{R}_{\mathbf{t},j}(\mathbf{x}) W_n(t_j, x_j) \right| \xrightarrow{p} 0, \quad n \rightarrow \infty.$$

As a consequence, we have the weak convergence

$$\hat{W}_n \rightsquigarrow \hat{W}, \quad n \rightarrow \infty$$

in the space $\ell^\infty(\mathcal{V}')$, where

$$\hat{W}(\mathbf{t}, \mathbf{x}) = W(\mathbf{v}) - \sum_{j=1}^d \dot{R}_{\mathbf{t},j}(\mathbf{x}) W(t_j, x_j)$$

and W is the centered Gaussian process in Theorem 1. Almost all trajectories of $\hat{W}$ are uniformly $\hat{\rho}_2$ -continuous on $\mathcal{V}'$, where

$$\hat{\rho}_2((s, \mathbf{x}), (t, \mathbf{y})) = \left(\mathbb{E}[\{\hat{W}(s, \mathbf{x}) - \hat{W}(t, \mathbf{y})\}^2] \right)^{1/2}$$

is the standard deviation semimetric based on $\hat{W}$.

Condition 7 controls the difference between the local increments of $E_{n,t}$ and those of R_t . The following condition is stronger and controls the global rate of convergence of $E_{n,t}$ to R_t .

CONDITION 8. *For all $M > 0$ and for all $d = 1, 2, \dots, D$, we have*

$$\sup_{v \in \mathcal{V}_d} \sqrt{k} |E_{n,t}(\mathbf{x}) - R_t(\mathbf{x})| \rightarrow 0, \quad n \rightarrow \infty.$$

This stronger condition allows us to center $\hat{R}_{n,t}(\mathbf{x})$ with the target $R_t(\mathbf{x})$ instead of $E_{n,t}(\mathbf{x})$: the empirical tail copula process is defined by

$$(13) \quad \sqrt{k} \{ \hat{R}_{n,t}(\mathbf{x}) - R_t(\mathbf{x}) \}, \quad v = (t, \mathbf{x}) \in \mathcal{V}.$$

The following corollary to Theorem 2 is the main result of the paper.

COROLLARY 2. *If, in Theorem 2, we replace Condition 7 by the stronger Condition 8, then all conclusions in that theorem hold, and they continue to hold if the penultimate empirical tail copula process $\hat{W}_n$ is replaced by the empirical tail copula process in Eq. (13).*

REMARK 3 (Stable tail dependence function). Fix $D = 1, 2, \dots$. For $(t, \mathbf{x}) \in T^D \times [0, \infty)^D$, the stable tail dependence function (stdf) is given by

$$l_t(\mathbf{x}) = \lim_{s \downarrow 0} s^{-1} \mathbb{P}[U_1(t_1) \leq sx_1 \text{ or } \dots \text{ or } U_1(t_D) \leq sx_D].$$

It is well-known that the stdf characterizes multivariate tail dependence, see, e.g., [Einmahl, Krajina and Segers \(2012\)](#). Since the definition of $l_t(\mathbf{x})$ on $T^D \times [0, \infty)^D$ involves the probability of a finite union of events it can be obtained from all the tail copulas R_t on $T^d, d = 1, \dots, D$, through the inclusion-exclusion principle, and the same relation holds between the rank-based estimator

$$\hat{l}_{n,t}(\mathbf{x}) = \frac{1}{k} \sum_{i=1}^n \mathbb{1} \{ \exists j = 1, \dots, D : \text{rank}_{n,i}(t_j) > n - kx_j + 1 \}$$

of the stdf and all the $\hat{R}_{n,t}$ ([Chiapino, Sabourin and Segers, 2019](#)). Hence we obtain immediately the result corresponding to Corollary 2 (in particular the weak convergence) for the empirical stdf process

$$\sqrt{k} \{ \hat{l}_{n,t}(\mathbf{x}) - l_t(\mathbf{x}) \}, \quad v = (t, \mathbf{x}) \in \mathcal{V}'_D.$$

That result would have been another way to present the main result of this paper, but for notational ease we chose to focus on tail copulas.

REMARK 4 (Challenges of the proof). The proof of Theorem 2 is particularly challenging because the partial derivatives $R_{t,j}$ are not uniformly equicontinuous since some of the coordinates of t can be arbitrarily close together. This problem arises for functional data only and asks for a novel approach. It is easily illustrated for $d = 2$. When t_1 is fixed and t_2 approaches t_1 , then R_{t_1,t_2} approaches the non-differentiable comonotone tail copula R_{t_1,t_1} , given by $R_{t_1,t_1}(x_1, x_2) = x_1 \wedge x_2$. Now consider an arbitrarily small neighborhood of the point (x, x) . Then if t_2 approaches t_1 , the partial derivative $\hat{R}_{t_1,t_2,1}$ takes, in this neighborhood, values, that range from about 0 to about 1. We solve this problem essentially by taking directional derivatives along the vector $(1, 1)$, that is, along the diagonal. When d is arbitrary, these directional derivatives in the direction of a lower dimensional 1-vector, can be expressed in the “smooth” partial derivatives with respect to the remaining coordinates and the tail copula

itself, which opens the door towards the solution. This program is carried out in the treatment of the term $\Delta_{n,3}$ in the proof, more precisely in Case II starting on page 22. In order to make this approach work various interesting lemmas on tail copulas and their partial derivatives are derived in Appendix A.

REMARK 5 (Empirical copulas). Taking $k = n$ instead of $k/n \rightarrow 0$, one could also consider empirical copulas rather than empirical tail copulas. It is known from the multivariate case that the limiting process will be different then, because all n data are taken into account which leads to a tied-down process, a bridge, rather than a Wiener process, as in Theorem 1. This goes back to Ruymgaart (1973); see Segers (2012) and the references therein. To prove in the functional setting an analogue of Theorem 2, however, is a new research project, since the proof techniques used here cannot be transferred, mainly because the partial derivatives of a copula enjoy fewer regularity properties than those of a tail copula, since a copula is in general not concave and not homogeneous.

We conclude this section with two examples for which we work out the conditions in some detail. From these, more examples can be generated by transforming the trajectories $t \mapsto \xi(t)$ coordinate-wise to $t \mapsto f(t, \xi(t))$, where $f : T \times \mathbb{R} \rightarrow \mathbb{R}$ is continuous and $z \mapsto f(t, z)$ is increasing for every $t \in T$. Such a transformation only changes the marginal distributions F_t but leaves the uniformized process $U_t = 1 - F_t(\xi(t))$ unaffected.

EXAMPLE 1 (Smith model). Let $T = [0, 1]^r$ for some $r = 1, 2, \dots$ and let $\{\mathbf{S}_i, Y_i\}_{i \geq 1}$ be the points of a unit-rate homogeneous Poisson process on $\mathbb{R}^r \times (0, \infty)$. Consider the process

$$\xi(\mathbf{u}) = \max_{i=1,2,\dots} Y_i^{-1} f(\mathbf{S}_i - \mathbf{u}), \quad \mathbf{u} \in [0, 1]^r,$$

where

$$f(\mathbf{s}) = (2\pi)^{-r/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} \mathbf{s}' \Sigma^{-1} \mathbf{s}\}, \quad \mathbf{s} \in \mathbb{R}^r,$$

is a centered normal density, with Σ an invertible $r \times r$ covariance matrix. The process ξ follows the so-called Smith model, after Smith (1990). It has been applied and extended in Coles (1993), Schlather (2002) and Kabluchko, Schlather and de Haan (2009), among others, and is a special case of the Brown–Resnick process. The process ξ is a continuous, stationary, max-stable process with unit-Fréchet marginals $F_{\mathbf{u}}(z) = \exp(-1/z)$, $z > 0$.

The finite-dimensional distributions of ξ can be computed via the probability that the Poisson process with points $(\mathbf{S}_i, Y_i)$ misses a certain subset of $\mathbb{R}^r \times (0, \infty)$. Let $\mathbf{S}$ be a random vector with density f , i.e., a $N_r(\mathbf{0}, \Sigma)$ random vector, and define a stochastic process on $\mathbb{R}^r$ by

$$B(\mathbf{u}) = \exp\left(\mathbf{S}' \Sigma^{-1} \mathbf{u} - \frac{1}{2} \mathbf{u}' \Sigma^{-1} \mathbf{u}\right), \quad \mathbf{u} \in \mathbb{R}^r.$$

Note that $B(\mathbf{u})$ is a log-normal random variable with unit expectation. We find, for $(\mathbf{u}_1, \dots, \mathbf{u}_d) \in T^d$ and $(z_1, \dots, z_d) \in (0, \infty)^d$, that

$$\begin{aligned} -\ln \mathbb{P}\{\forall j = 1, \dots, d : \xi(\mathbf{u}_j) \leq z_j\} &= \int_{\mathbf{s} \in \mathbb{R}^r} \max_{j=1,\dots,d} \{z_j^{-1} f(\mathbf{s} - \mathbf{u}_j)\} d\mathbf{s} \\ &= \mathbb{E} \left[\max_{j=1,\dots,d} z_j^{-1} B(\mathbf{u}_j) \right]. \end{aligned}$$

The second identity follows from computing $f(\mathbf{s} - \mathbf{u}_j)/f(\mathbf{s})$ inside the maximum in the integral. The finite-dimensional distributions of ξ are multivariate Hüsler–Reiss distributions

(Hüsler and Reiss, 1989). The last expression represents the finite-dimensional distributions via a D -norm (Falk, 2019).

The tail copula of ξ can be obtained from the finite-dimensional distributions via the inclusion–exclusion formula and the minimum–maximum identity: for $(\mathbf{u}_1, \dots, \mathbf{u}_d) \in T^d$ and $\mathbf{x} \in [0, \infty)^d$

$$(14) \quad \begin{aligned} R_{\mathbf{u}_1, \dots, \mathbf{u}_d}(\mathbf{x}) &= \int_{\mathbf{x} \in \mathbb{R}^r} \min_{j=1, \dots, d} \{x_j f(\mathbf{s} - \mathbf{u}_j)\} d\mathbf{s} \\ &= \mathbb{E} \left[\min_{j=1, \dots, d} x_j B(\mathbf{u}_j) \right]. \end{aligned}$$

The partial derivative of $R_{\mathbf{u}_1, \dots, \mathbf{u}_d}(\mathbf{x})$ with respect to x_j in a point $\mathbf{x}$ such that x_j is positive is

$$\dot{R}_{\mathbf{u}_1, \dots, \mathbf{u}_d; j}(\mathbf{x}) = \mathbb{E} \left[B(\mathbf{u}_j) \mathbb{1} \left\{ x_j B(\mathbf{u}_j) = \min_{\alpha=1, \dots, d} x_\alpha B(\mathbf{u}_\alpha) \right\} \right]$$

By the dominated convergence theorem, this expression is continuous in $\mathbf{x}$ such that $x_j > 0$.

The bivariate tail copula can be further computed from the integral formula (14) via some elementary algebra: for $(\mathbf{u}, \mathbf{v}, x, y) \in T^2 \times (0, \infty)^2$,

$$R_{\mathbf{u}, \mathbf{v}}(x, y) = x \bar{\Phi} \left(\frac{a(\mathbf{u}, \mathbf{v})}{2} + \frac{1}{a(\mathbf{u}, \mathbf{v})} \ln(x/y) \right) + y \bar{\Phi} \left(\frac{a(\mathbf{u}, \mathbf{v})}{2} + \frac{1}{a(\mathbf{u}, \mathbf{v})} \ln(y/x) \right)$$

where $a(\mathbf{u}, \mathbf{v}) = \{(\mathbf{u} - \mathbf{v})' \Sigma^{-1} (\mathbf{u} - \mathbf{v})\}^{1/2}$, while $\bar{\Phi} = 1 - \Phi$ and Φ is the standard normal distribution function. In particular, $R_{\mathbf{u}, \mathbf{v}}(1, 1) = 2\bar{\Phi}(a(\mathbf{u}, \mathbf{v})/2)$, which is less than 1 as soon as $\mathbf{u} \neq \mathbf{v}$. The standard deviation semimetric $\rho_{2,T}$ in (11) is equivalent to the Euclidean metric on T .

We verify Condition 8. By the inclusion–exclusion formula, $\mathbb{P}\{\forall j = 1, \dots, d : \xi(\mathbf{u}_j) > z_j\}$ is a linear combination of expressions of the form $\exp[-\mathbb{E}\{\max_{j \in A} z_j^{-1} B(\mathbf{u}_j)\}]$, where A ranges over the non-empty subsets of $\{1, \dots, d\}$. It follows that $E_{n, \mathbf{u}_1, \dots, \mathbf{u}_d}(\mathbf{x})$ in (3) is a linear combination of expressions of the form

$$\exp(-\mathbb{E}[\max_{j \in A} \{-\ln(1 - \frac{k}{n} x_j) B(\mathbf{u}_j)\}]).$$

A Taylor expansion of the functions $h \mapsto \exp(-h)$ and $h \mapsto -\ln(1 - h)$ as $h \rightarrow 0$ yields that the convergence rate of $E_{n, \mathbf{u}_1, \dots, \mathbf{u}_d}(\mathbf{x})$ to $R_{\mathbf{u}_1, \dots, \mathbf{u}_d}(\mathbf{x})$ is $O(k/n)$ uniformly in $T^d \times [0, M]^d$, for $M > 0$. It follows that Condition 8 is fulfilled as soon as $k^{3/2} = o(n)$ as $n \rightarrow \infty$.

We have thus verified that the conditions of Corollary 2 are satisfied, except for Condition 5. But the latter can be shown to hold for the uniformized version U of ξ using the arguments on page 476 of Einmahl and Lin (2006).

Since the process ξ in this example is max-stable, the tail copulas of its finite-dimensional distributions are linked to the Pickands dependence functions of these distributions. Using this additional knowledge, the latter functions can be estimated nonparametrically at rate $O_p(1/\sqrt{n})$, see for instance Gudendorf and Segers (2012). Our estimator is constructed to be valid for a much broader class of processes for which the tail copula only arises in the limit, and this explains its convergence rate $O_p(1/\sqrt{k})$.

EXAMPLE 2 (Pareto process). Let $T = [0, 1]$. Let Y be a standard Pareto random variable, i.e., $\mathbb{P}(Y \leq x) = 1 - 1/x$, for $x \geq 1$, and let W' be a standard Wiener process on $[0, 1]$. Consider $B(t) = \exp\{W'(t) - t/2\}$, $t \in [0, 1]$, a geometric Brownian motion. We have $B(t) > 0$, $\mathbb{E}\{B(t)\} = 1$ for all $t \in [0, 1]$, and $\mathbb{E}\{\sup_{t \in [0, 1]} B(t)\} < \infty$. Assume Y and W' are independent. Define

$$\xi(t) = Y B(t), \quad t \in [0, 1],$$

a process introduced in [Gomes, de Haan and Pestana \(2004\)](#); see [Ferreira and de Haan \(2014\)](#) and [Dombry and Ribatet \(2015\)](#) for related processes. The process ξ is continuous and its finite-dimensional distributions can be found via Fubini's theorem, giving

$$\mathbb{P}[\forall j = 1, \dots, d : \xi(t_j) \geq y/x_j] = y^{-1} \mathbb{E}[\min\{y, x_1 B(t_1), \dots, x_d B(t_d)\}]$$

for $y, x_1, \dots, x_d \in [0, \infty)$ and $t_1, \dots, t_d \in T$. It follows that the tails of the margins of F_t are very close to standard Pareto,

$$\forall m > 0, \quad \sup_{t \in T} |1 - F_t(y) - 1/y| = o(y^{-m}), \quad y \rightarrow \infty,$$

and that the tail copula is

$$R_t(\mathbf{x}) = \mathbb{E}[\min\{x_1 B(t_1), \dots, x_d B(t_d)\}]$$

for $d = 1, 2, \dots$ and $(t, \mathbf{x}) \in T^d \times [0, \infty)^d$. Moreover,

$$\forall m > 0, \quad \sup_{v \in \mathcal{V}_d} |E_{n,t}(\mathbf{x}) - R_t(\mathbf{x})| = o((k/n)^m), \quad n \rightarrow \infty.$$

Condition 8 is thus satisfied as soon as $k^{1+\eta} = o(n)$ for some $\eta > 0$. If all coordinates $t_1, \dots, t_d$ are distinct, the partial derivative of $R_t(\mathbf{x})$ with respect to x_j is

$$\dot{R}_{t,j}(\mathbf{x}) = \mathbb{E} \left[B_j(t_j) \mathbb{1} \left\{ x_j B(t_j) = \min_{\alpha=1, \dots, d} x_\alpha B(t_\alpha) \right\} \right],$$

which is continuous in points $\mathbf{x} \in [0, \infty)^d$ such that $x_j > 0$, thanks to the dominated convergence theorem. The bivariate tail copula is

$$R_{s,t}(x, y) = x \bar{\Phi} \left(\frac{\sqrt{|s-t|}}{2} + \frac{1}{\sqrt{|s-t|}} \ln(x/y) \right) + y \bar{\Phi} \left(\frac{\sqrt{|s-t|}}{2} + \frac{1}{\sqrt{|s-t|}} \ln(y/x) \right)$$

for $(s, t, x, y) \in [0, 1]^2 \times (0, \infty)^2$. In particular, $R_{s,t}(1, 1) = 2\bar{\Phi}(\sqrt{|s-t|}/2)$, so that the semimetric $\rho_{2,T}$ on T is equivalent to the absolute value metric.

Again Condition 5 can be shown to hold for the uniformized version U of ξ using arguments in [Einmahl and Lin \(2006, pages 479–480\)](#).

4. Testing stationarity. Under the conditions of Corollary 2, the empirical tail copula process in (13) converges weakly in the space $\ell^\infty(\mathcal{V}')$ to the centred, Gaussian process $\hat{W}$ described in Theorem 2. Here, we apply the result to implement a novel test for spatial stationarity of the tail copula in case $T = [0, 1]$. We evaluate the finite-sample performance of the test for the Smith model and the Pareto process in Examples 1 and 2 respectively by means of Monte Carlo simulations.

The functional data ξ_i are said to be tail copula stationary if

$$(15) \quad R_{t_1+h, \dots, t_d+h}(\mathbf{x}) = R_{t_1, \dots, t_d}(\mathbf{x})$$

for any $d = 2, 3, \dots$, any $t \in [0, 1]^d$, any $\mathbf{x} \in [0, \infty)^d$ and any $h \in [-1, 1]$ such that $t_j + h \in [0, 1]$ for all $j = 1, \dots, d$. Given a random sample $\xi_1, \dots, \xi_n$ in $C([0, 1])$, we will use the empirical tail copula $\hat{R}_{n,t}$ in (10) at threshold parameter $k = k_n$ as in Condition 2 to test whether (15) holds for $d = 2$ and $(x_1, x_2) = (1, 1)$. Further, we assume that we only observe $\xi_i(t)$ for $t \in \{r/N : r = 0, 1, \dots, N\}$ for some integer $N \geq 3$.

For a positive integer $\Delta < N/2$ and for $r_0 \in \{\Delta, \dots, N - \Delta\}$, put

$$(16) \quad \hat{I}_n^{(N)}(r_0/N) = \frac{1}{2\Delta} \sum_{r: 1 \leq |r-r_0| \leq \Delta} \hat{R}_{n;r/N, r_0/N}(1, 1)$$

and define the test statistic

$$(17) \quad D_n^{(N)} = \sqrt{k} \left[\max_{r_0=\Delta, \dots, N-\Delta} \hat{I}_n^{(N)}(r_0/N) - \min_{r_0=\Delta, \dots, N-\Delta} \hat{I}_n^{(N)}(r_0/N) \right].$$

Assume the conditions of Corollary 2. Under the null hypothesis of stationarity and for fixed N and Δ , the asymptotic distribution of the test statistic is

$$(18) \quad D_n^{(N)} \rightsquigarrow D^{(N)} = \max_{r_0=\Delta, \dots, N-\Delta} V^{(N)}(r_0/N) - \min_{r_0=\Delta, \dots, N-\Delta} V^{(N)}(r_0/N), \quad n \rightarrow \infty,$$

where, for $r_0 \in \{\Delta, \dots, N - \Delta\}$,

$$(19) \quad V^{(N)}(r_0/N) = \frac{1}{2\Delta} \sum_{r:1 \leq |r-r_0| \leq \Delta} \hat{W}(r/N, r_0/N; 1, 1).$$

and $\hat{W}$ is the limit process in Theorem 2. Corollary 2 even allows us to find the limit of $D_n^{(N)}$ in case $N = N_n \rightarrow \infty$ and $\Delta = \Delta_n \rightarrow \infty$ in such a way that $\Delta/N \rightarrow \delta \in (0, 1/2)$. This is the more relevant and more interesting case when considering functional data. If the function $(s, t) \mapsto R_{s,t}(1, 1)$ is continuous with respect to the Euclidean metric on $[0, 1]^2$, then

$$D_n^{(N)} \rightsquigarrow D = \sup_{t \in [\delta, 1-\delta]} V(t) - \inf_{t \in [\delta, 1-\delta]} V(t)$$

where $V(t_0) = \frac{1}{2\delta} \int_{t_0-\delta}^{t_0+\delta} \hat{W}(t, t_0; 1, 1) dt$ for $t_0 \in [\delta, 1-\delta]$. A sketch of the proofs of these statements is given in Appendix B.

If the tail copula fails to be stationary, we can expect $D_n^{(N)} \rightarrow \infty$ in distribution. Therefore we reject the null hypothesis of stationarity for large values of $D_n^{(N)}$. To find the p -value of the observed test statistic, we rely on the asymptotic distribution under the null hypothesis $D^{(N)}$, which is based on a centered Gaussian random vector of dimension $N - 2\Delta + 1$. In view of the definition of $\hat{W}$ in Theorem 2, the covariance matrix of $V^{(N)}$ can be expressed in terms of the tail dependence coefficients $R_t(\mathbf{1})$ for $t \in \{0, 1/N, \dots, 1\}^d$ for $d \in \{2, 3, 4\}$ and the partial derivatives $\dot{R}_{t_1, t_2; j}(1, 1)$ for $(t_1, t_2) \in \{0, 1/N, \dots, 1\}^2$ and $j \in \{1, 2\}$. These can be estimated consistently, again by Corollary 2; for the partial derivatives use finite differencing at a bandwidth sequence that goes to zero more slowly than $k^{-1/2}$ and the fact that by stationarity, $\dot{R}_{t_1, t_2; j}(1, 1)$ only depends on $t_2 - t_1$. Under the null hypothesis, the covariance matrix of $V^{(N)}$ is Toeplitz, a fact which can be used in its estimation as well. The distribution function of $D^{(N)}$ at the estimated covariance matrix of $V^{(N)}$ can then be calculated in terms of the multivariate normal cumulative distribution function (cdf). For details, we refer to Appendix B.

We have implemented the test in R (R Core Team, 2018), relying on the package `mnormt` (Azzalini and Genz, 2020) for calculating the multivariate normal cdf as detailed in Appendix B. We then evaluated the test's finite-sample performance for the Smith model with $\Sigma = 1$ and the Pareto process in Examples 1 and 2, respectively, both of which are tail copula stationary. For drawing samples from the Smith model, we used the package `spatialExtremes` (Ribatet, 2020). To assess the power of the test against alternatives, we simulated from the same processes but on distorted grids of the form $\{f_\theta(r/N) : r = 0, 1, \dots, N\} \in [0, 1]$ where $f_\theta(t) = [(2 - \theta)t]/[1 - \theta + \sqrt{(2 - \theta)t\theta + (1 - \theta)^2}]$ for $\theta \in [0, 1]$. Note that $f_0(t) = t$ while $f_1(t) = \sqrt{(2 - t)t}$, the graph of a quarter circle with center $(1, 0)$ and radius 1.

The results for $N = 20$, $\Delta = 2$, $n = 500$ and $k = 50$ are shown in Figure 1, with the Smith model on the left and the Pareto process on the right:

- The top row shows the probability mass function (pmf) of the integer-valued random variable $2\Delta\sqrt{k}D_n^{(N)}$ under the null hypothesis together with the probability density function (pdf)

of the random variable $2\Delta\sqrt{k}D^{(N)}$ with $D^{(N)}$ in (18); please see Appendix B for how we calculated the latter. Note that, despite the bell shape of the pdf, the limit variable is not Gaussian. The pmf is estimated on the basis of 5 000 random samples.

- The middle row shows PP-plots of the p -values based on the estimated covariance matrix of $V^{(N)}$ based on 1 000 samples under the null hypothesis.
- The bottom row shows the rejection probabilities at significance level $\alpha = 0.05$ at the null hypothesis ($\theta = 0$) and at alternatives ($\theta \in \{0.1, 0.2, \dots, 1\}$) and this for $\Delta \in \{1, \dots, 4\}$, based on 1 000 samples for each value of θ . The null hypothesis is rejected if the p -value calculated using the estimated covariance matrix of $V^{(N)}$ does not exceed α .

The asymptotic theory works well in the sense that the continuous limit distribution of the test statistic matches its discrete finite-sample distribution quite closely. The p -values look reasonably uniform, especially for the Pareto process, despite the complicated setup and the fact that the covariance matrix of $V^{(N)}$ is re-estimated for each sample. Also for the Smith process the fit is very good for the relevant small values of the significance level. From the power plots, we see that the test picks up the alternative quite well. Similar patterns appeared for other choices n, k, N, Δ . Overall, the plots show that distributional approximations based on Corollary 2 are very useful already at moderate sample sizes.

5. Proofs for the results of Section 2. We first collect some results for the proof of Theorem 1.

LEMMA 1. *On some non-empty set T , let $U = (U(t))_{t \in T}$ be a stochastic process with uniform-(0, 1) margins. Assume that there exists a finite covering $T = \bigcup_{j=1}^N T_j$ of T and a scalar $\lambda > 1$ such that*

$$(20) \quad \forall j = 1, \dots, N, \quad \liminf_{x \downarrow 0} \mathbb{P} \left[\sup_{t \in T_j} U(t) \leq \lambda x \mid \inf_{t \in T_j} U(t) \leq x \right] > 0.$$

Then $\mathbb{P}[\inf_{t \in T} U(t) \leq x] = O(x)$ as $x \downarrow 0$.

PROOF. Write $A_j = \inf_{t \in T_j} U(t)$. Since

$$\mathbb{P} \left[\inf_{t \in T} U(t) \leq x \right] = \mathbb{P} \left[\min_{j=1, \dots, N} A_j \leq x \right] \leq \sum_{j=1}^N \mathbb{P} [A_j \leq x],$$

it is sufficient to show that $\mathbb{P}[A_j \leq x] = O(x)$ as $x \downarrow 0$ for every $j = 1, \dots, N$. Write $B_j = \sup_{t \in T_j} U(t)$ and note that $\mathbb{P}[B_j \leq y] \leq \mathbb{P}[U(t) \leq y] \leq y$ for $y \geq 0$, where $t \in T_j$ is arbitrary. Then

$$\begin{aligned} \mathbb{P} [A_j \leq x] &= \mathbb{P} [A_j \leq x, B_j \leq \lambda x] + \mathbb{P} [A_j \leq x, B_j > \lambda x] \\ &\leq \lambda x + \mathbb{P} [A_j \leq x] \mathbb{P} [B_j > \lambda x \mid A_j \leq x]. \end{aligned}$$

Solving for $\mathbb{P} [A_j \leq x]$ yields

$$\mathbb{P} [A_j \leq x] \leq \frac{\lambda x}{\mathbb{P} [B_j \leq \lambda x \mid A_j \leq x]}.$$

For sufficiently small x , the denominator on the right-hand side is bounded away from zero by assumption. $\square$

The following theorem is a corollary to Theorem 2.11.9 in [van der Vaart and Wellner \(1996\)](#).

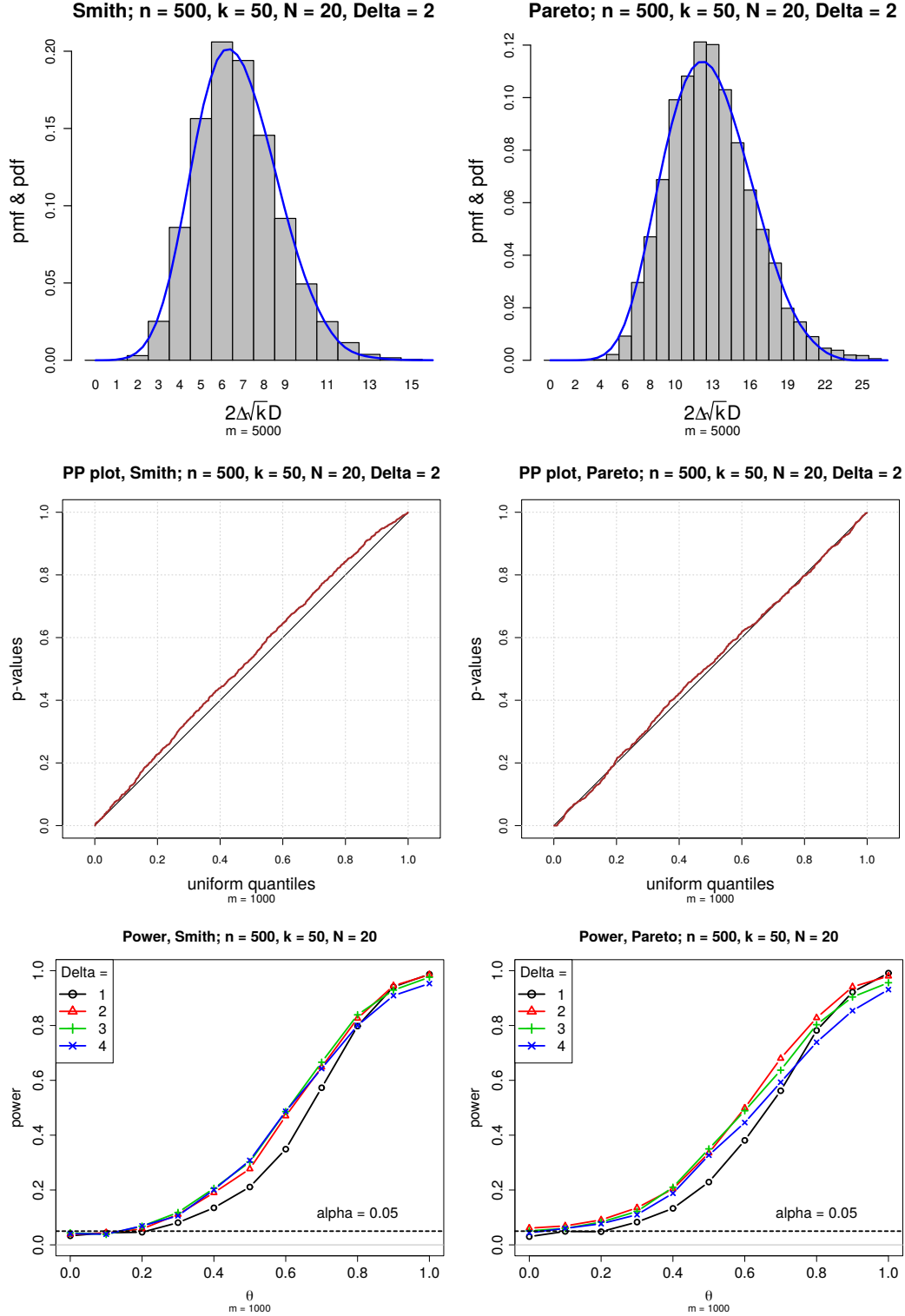


FIG 1. Test of Section 4 for tail copula stationarity for stochastic processes on $[0, 1]$ observed on $\{r/N : r = 0, 1, \dots, N\}$ with $N = 20$. Sample size $n = 500$, threshold parameter $k = 50$. Left: Smith model (Example 1). Right: Pareto process (Example 2). Top: sampling distribution of the discrete test statistic under the null hypothesis together with the density of the limiting variable. Middle: PP-plot of the p -values of the test under the null hypothesis. Bottom: Powers at significance level $\alpha = 0.05$ against alternatives parametrized by $\theta \in [0, 1]$. Plots based on 5000 samples at the top, 1000 samples in the middle, and 1000 samples per value of θ at the bottom.

THEOREM 3. *For each n , let $Z_{n,1}, \dots, Z_{n,n}$ be independent stochastic processes with finite second moments indexed by a set $\mathcal{F}$. Write $\|Z_{n,i}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |Z_{n,i}(f)|$. Suppose*

$$(21) \quad \forall \lambda > 0, \quad \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}^* [\|Z_{n,i}\|_{\mathcal{F}} \mathbb{1}\{\|Z_{n,i}\|_{\mathcal{F}} > \lambda\}] = 0,$$

and suppose that there exists $c > 0$ and, for all sufficiently small $\varepsilon > 0$, a covering $\mathcal{F} = \bigcup_{j=1}^{N_\varepsilon} \mathcal{F}_{\varepsilon,j}$ of $\mathcal{F}$ by sets $\mathcal{F}_{\varepsilon,j}$ such that, for every set $\mathcal{F}_{\varepsilon,j}$ and every n ,

$$(22) \quad \sum_{i=1}^n \mathbb{E}^* \left[\sup_{f,g \in \mathcal{F}_{\varepsilon,j}} |Z_{n,i}(f) - Z_{n,i}(g)|^2 \right] \leq c\varepsilon^2$$

and, for some $\delta > 0$, we have

$$(23) \quad \int_0^\delta \sqrt{\log N_\varepsilon} d\varepsilon < \infty.$$

Then the sequence $\sum_{i=1}^n \{Z_{n,i} - \mathbb{E}(Z_{n,i})\}$ is asymptotically tight in $\ell^\infty(\mathcal{F})$ and converges weakly provided the finite-dimensional distributions converge weakly.

PROOF. We show that the conditions of Theorem 2.11.9 in [van der Vaart and Wellner \(1996\)](#) are fulfilled. First, from a given covering, we can easily extract a partition meeting the same requirements by considering successive differences of the covering sets, and hence the sets in the partition are subsets of the corresponding sets in the covering.

In comparison to the cited theorem, our partitions do not depend on n . As a consequence, the middle one of the three displayed conditions in that theorem is redundant. But then the semimetric ρ on $\mathcal{F}$ does not play any role either and it can be any semimetric that makes $\mathcal{F}$ totally bounded; take for instance the trivial zero semimetric.

Eq. (23) obviously implies that $\int_0^{\delta_n} \sqrt{\log N_\varepsilon} d\varepsilon \rightarrow 0$ for every $\delta_n \downarrow 0$.

Finally, the upper bound in (22) has $c\varepsilon^2$ rather than just ε^2 as in the displayed equation just above the statement of Theorem 2.11.9 in [van der Vaart and Wellner \(1996\)](#). This makes no difference since we can just take the cover or partition associated to $\varepsilon/\sqrt{c}$. Indeed, the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_2^n)$ in that theorem is then bounded by $N_{\varepsilon/\sqrt{c}}$, and for sufficiently small $\delta > 0$, the integral $\int_0^{\sqrt{c}\delta} \sqrt{\log N_{\varepsilon/\sqrt{c}}} d\varepsilon = \sqrt{c} \int_0^\delta \sqrt{\log N_\varepsilon} d\varepsilon$ is finite. $\square$

PROOF OF THEOREM 1. We will apply Theorem 3 with $\mathcal{F}$ equal to $\mathcal{V}$. Define stochastic processes $Z_{n,i}$ on $\mathcal{V}$ by

$$Z_{n,i}(\mathbf{v}) = k^{-1/2} \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) \leq \frac{k}{n} x_j\}, \quad \mathbf{v} = (\mathbf{t}, \mathbf{x}) \in \mathcal{V}_d,$$

so that $W_n(\mathbf{v}) = \sum_{i=1}^n [Z_{n,i}(\mathbf{v}) - \mathbb{E}\{Z_{n,i}(\mathbf{v})\}]$.

Step 1: Finite-dimensional distributions. — The finite-dimensional distributions of W_n converge due to the Lindeberg central limit theorem. Indeed, W_n is centered and the covariance function of W_n converges pointwise to the one of W , see Eq. (4). On the one hand, if $\text{var}\{W(\mathbf{v})\} = R_{\mathbf{t}}(\mathbf{x}) = 0$, then $\lim_{n \rightarrow \infty} \text{var}(W_n(\mathbf{v})) = 0$ too, and thus $W_n(\mathbf{v})$ converges in probability to $W(\mathbf{v}) = 0$. On the other hand, if the limit variance is positive, then the Lindeberg condition is trivially fulfilled, since $0 \leq Z_{n,i} \leq k^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$, so that for every $\varepsilon > 0$, the indicators $\mathbb{1}\{|Z_{n,i}(\mathbf{v}) - \mathbb{E}[Z_{n,i}(\mathbf{v})]| \geq \varepsilon \sqrt{\text{var } W_n(\mathbf{v})}\}$ in the Lindeberg condition are all equal to zero for sufficiently large n .

Step 2: Checking Eq. (21). — Fix $\lambda > 0$. Since $\sup_{\mathbf{v} \in \mathcal{V}} |Z_{n,i}(\mathbf{v})| \leq k^{-1/2} \rightarrow 0$ as $n \rightarrow \infty$ by Condition 2, the indicator variable inside the outer expectation is zero for n sufficiently large, such that $k^{-1/2} \leq \lambda$.

Step 3: Constructing the covering. — Let $\varepsilon > 0$ be small. For each $d = 1, \dots, D$, we construct a covering of $\mathcal{V}_d = \bigcup_{m=1}^{N_{d,\varepsilon}} \mathcal{V}_d(\varepsilon, m)$ such that, for every covering set $\mathcal{V}_d(\varepsilon, m)$, we have

$$\sum_{i=1}^n \mathbb{E}^* \left[\sup_{\mathbf{v}, \mathbf{w} \in \mathcal{V}_d(\varepsilon, m)} |Z_{n,i}(\mathbf{v}) - Z_{n,i}(\mathbf{w})|^2 \right] \leq c\varepsilon^2$$

for some $c > 0$ not depending on ε and such that $\int_0^\delta \sqrt{\log N_{d,\varepsilon}} d\varepsilon$ is finite for some $\delta > 0$. Uniting the coverings over $d = 1, \dots, D$ yields a covering $\mathcal{V} = \bigcup_{d=1}^D \bigcup_{m=1}^{N_{d,\varepsilon}} \mathcal{V}_d(\varepsilon, m)$ by $N_\varepsilon = \sum_{d=1}^D N_{d,\varepsilon}$ sets.

Consider the covering sets $T_{\varepsilon,j}$ for $j = 1, \dots, N_{T,\varepsilon}$ in Condition 4. Further, let $[0, M] = \bigcup_{r=1}^{\lceil M/\varepsilon^2 \rceil} [l_{\varepsilon,r}, u_{\varepsilon,r}]$ be a covering of $[0, M]$ by intervals with lengths $u_{\varepsilon,r} - l_{\varepsilon,r}$ at most ε^2 . For $m \in \{1, \dots, N_{T,\varepsilon}\}^d \times \{1, \dots, \lceil M/\varepsilon^2 \rceil\}^d = \mathcal{M}_\varepsilon^d$, put

$$(24) \quad \mathcal{V}_d(\varepsilon, m) = \prod_{j=1}^d T_{\varepsilon, m_j} \times \prod_{j=1}^d [l_{\varepsilon, m_{j+d}}, u_{\varepsilon, m_{j+d}}].$$

For fixed $\varepsilon > 0$, we have $\mathcal{V}_d = \bigcup_{m \in \mathcal{M}_\varepsilon^d} \mathcal{V}_d(\varepsilon, m)$.

Step 4: Checking Eq. (23). — The number of covering sets is $N_\varepsilon = \sum_{d=1}^D N_{d,\varepsilon}$ where $N_{d,\varepsilon} = |\mathcal{M}_\varepsilon^d| = (N_{T,\varepsilon} \lceil M/\varepsilon^2 \rceil)^d$. Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for nonnegative a and b , we have

$$\begin{aligned} \sqrt{\log N_\varepsilon} &\leq \sqrt{\log(DN_{D,\varepsilon})} \\ &\leq \sqrt{\log D} + \sqrt{D} \left(\sqrt{\log N_{T,\varepsilon}} + \sqrt{\log(2M)} + \sqrt{2\log(1/\varepsilon)} \right). \end{aligned}$$

The condition $\int_0^1 \sqrt{\log N_{T,\varepsilon}} d\varepsilon < \infty$ implies the same inequality for $N_{T,\varepsilon}$ replaced by N_ε .

Step 5: Checking Ineq. (22). — We check that each covering set $\mathcal{V}_d(\varepsilon, m)$, for $d = 1, \dots, D$ and $m \in \mathcal{M}_\varepsilon^d$, satisfies Ineq. (22), with $c > 0$ to be determined.

Step 5.1: Bounding the supremum. — For $i = 1, \dots, n$ and $j = 1, \dots, d$, write

$$\begin{aligned} A_{i,j} &= \inf_{t \in T_{\varepsilon, m_j}} U_i(t), & a_j &= \frac{k}{n} u_{\varepsilon, m_{j+d}}, \\ B_{i,j} &= \sup_{t \in T_{\varepsilon, m_j}} U_i(t), & b_j &= \frac{k}{n} l_{\varepsilon, m_{j+d}}. \end{aligned}$$

Put $\lambda = \exp(c_1 \varepsilon^2)$ with c_1 as in Condition 4. In view of the definition of $\mathcal{V}_d(\varepsilon, m)$ in Eq. (24), we have

$$\begin{aligned} \sup_{\mathbf{v} \in \mathcal{V}_d(\varepsilon, m)} Z_{n,i}(\mathbf{v}) &= \sup_{\mathbf{v} \in \mathcal{V}_d(\varepsilon, m)} k^{-1/2} \mathbb{1}\{\forall j : U_i(t_j) \leq \frac{k}{n} x_j\} \\ &\leq k^{-1/2} \mathbb{1}\{\forall j : A_{i,j} \leq a_j\} \\ &\leq k^{-1/2} \mathbb{1}\{\forall j : A_{i,j} \leq a_j, B_{i,j} \leq \lambda a_j\} + k^{-1/2} \mathbb{1}\{\exists j : A_{i,j} \leq a_j, B_{i,j} > \lambda a_j\}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \inf_{\mathbf{v} \in \mathcal{V}_d(\varepsilon, m)} Z_{n,i}(\mathbf{v}) &= \inf_{\mathbf{v} \in \mathcal{V}_d(\varepsilon, m)} k^{-1/2} \mathbb{1}\{\forall j : U_i(t_j) \leq \frac{k}{n} x_j\} \\ &\geq k^{-1/2} \mathbb{1}\{\forall j : B_{i,j} \leq b_j\} \\ &\geq k^{-1/2} \mathbb{1}\{\forall j : A_{i,j} \leq \lambda^{-1} b_j, B_{i,j} \leq b_j\}. \end{aligned}$$

Since $\lambda^{-1}b_j \leq b_j \leq a_j \leq \lambda a_j$, we have

$$\begin{aligned} 0 &\leq \mathbb{1}\{\forall j : A_{i,j} \leq a_j, B_{i,j} \leq \lambda a_j\} - \mathbb{1}\{\forall j : A_{i,j} \leq \lambda^{-1}b_j, B_{i,j} \leq b_j\} \\ &\leq \mathbb{1}\{\exists j : \lambda^{-1}b_j < A_{i,j} \leq B_{i,j} \leq \lambda a_j\}. \end{aligned}$$

It follows that

$$\begin{aligned} &\sup_{\mathbf{v}, \mathbf{w} \in \mathcal{Y}_d(\varepsilon, m)} |Z_{n,i}(\mathbf{v}) - Z_{n,i}(\mathbf{w})|^2 \\ &\leq \left| \sup_{\mathbf{v} \in \mathcal{Y}_d(\varepsilon, m)} Z_{n,i}(\mathbf{v}) - \inf_{\mathbf{v} \in \mathcal{Y}_d(\varepsilon, m)} Z_{n,i}(\mathbf{v}) \right|^2 \\ &\leq k^{-1} \left(\mathbb{1}\{\exists j : \lambda^{-1}b_j < A_{i,j} \leq B_{i,j} \leq \lambda a_j\} + \mathbb{1}\{\exists j : A_{i,j} \leq a_j, B_{i,j} > \lambda a_j\} \right). \end{aligned}$$

Step 5.2: Bounding the expectation. — By the previous step, we have

$$\begin{aligned} (25) \quad &\sum_{i=1}^n \mathbb{E}^* \left[\sup_{\mathbf{v}, \mathbf{w} \in \mathcal{Y}_d(\varepsilon, m)} |Z_{n,i}(\mathbf{v}) - Z_{n,i}(\mathbf{w})|^2 \right] \\ &\leq \frac{n}{k} \sum_{j=1}^d \left[\mathbb{P}\{\lambda^{-1}b_j < A_{1,j} \leq B_{1,j} \leq \lambda a_j\} + \mathbb{P}\{A_{i,j} \leq a_j, B_{i,j} > \lambda a_j\} \right], \end{aligned}$$

since the stochastic processes U_i are identically distributed. We treat each of the two probabilities in turn.

First, picking any $t \in T_{\varepsilon, m_j}$, we have $A_{1,j} \leq U_1(t) \leq B_{1,j}$ and thus, since $U_1(t)$ is uniformly distributed on $(0, 1)$, we have

$$\begin{aligned} \mathbb{P}\{\lambda^{-1}b_j < A_{1,j} \leq B_{1,j} \leq \lambda a_j\} &\leq \mathbb{P}\{\lambda^{-1}b_j < U_1(t) \leq \lambda a_j\} \\ &\leq \lambda a_j - \lambda^{-1}b_j \\ &\leq \frac{k}{n} \left((\lambda - \lambda^{-1})M + \varepsilon^2 \right), \end{aligned}$$

where we used that $u_{\varepsilon, m_j+d} \in [0, M]$ and $l_{\varepsilon, m_j+d} \geq u_{\varepsilon, m_j+d} - \varepsilon^2$.

Second, by Condition 4, we have, for n sufficiently large such that $(k/n)M$ and thus $a_j = (k/n)u_{\varepsilon, m_j+d}$ is sufficiently small (Condition 2),

$$\begin{aligned} \mathbb{P}\{A_{i,j} \leq a_j, B_{i,j} > \lambda a_j\} &= \mathbb{P}\{B_{i,j} > \lambda a_j \mid A_{i,j} \leq a_j\} \mathbb{P}\{A_{i,j} \leq a_j\} \\ &\leq c_2 \varepsilon^2 \mathbb{P}\left\{ \inf_{t \in T} U_1(t) \leq \frac{k}{n} M \right\}. \end{aligned}$$

In view of Lemma 1, the probability on the right-hand side is bounded by $c_3(k/n)M$ for some $c_3 > 0$; the condition in Eq. (20) in that lemma is easily fulfilled thanks to Condition 4.

Putting things together, we obtain that the upper bound in Eq. (25) is, for sufficiently large n , dominated by

$$\frac{n}{k} D \left[\frac{k}{n} \{ (\lambda - \lambda^{-1})M + \varepsilon^2 \} + c_2 \varepsilon^2 \cdot c_3 \frac{k}{n} M \right].$$

Now $\lambda = \exp(c_1 \varepsilon^2)$, so that, for $\varepsilon > 0$ sufficiently close to zero, we have $0 \leq \lambda - \lambda^{-1} \leq 3c_1 \varepsilon^2$. We find the upper bound in Eq. (22) with $c = D(3c_1 M + 1 + c_2 c_3 M)$.

Step 6: Weak convergence and limit process. — By Steps 2 to 5, the criteria of Theorem 3 are met, so that the sequence of processes W_n is asymptotically tight in $\ell^\infty(\mathcal{V})$. Moreover,

the finite-dimensional distributions converge in view of Step 1. By [van der Vaart and Wellner \(1996, Theorem 1.5.4\)](#), W_n converges weakly in $\ell^\infty(\mathcal{V})$ to a tight version of the process W , denoted by W as well. This process W is Gaussian, and by Example 1.5.10 in [van der Vaart and Wellner \(1996\)](#), its tightness implies that the semimetric space $(\mathcal{V}, \rho_2)$ is totally bounded and that almost all sample paths $v \mapsto W(v)$ are uniformly ρ_2 -continuous, where ρ_2 is the standard deviation semimetric in Eq. (6). $\square$

PROOF OF COROLLARY 1. Let $(s, t) \in T^2$ and $\lambda > 1$. We have

$$\begin{aligned} u^{-1} \mathbb{P}[U_1(s) \leq u < \lambda u < U_1(t)] &= u^{-1} \mathbb{P}[U_1(s) \leq u] - u^{-1} \mathbb{P}[U_1(s) \leq u, U_1(t) \leq \lambda u] \\ &\rightarrow 1 - R_{s,t}(1, \lambda), \quad u \downarrow 0. \end{aligned}$$

Suppose in addition that $\rho(s, t) \leq \delta$ for some $\delta > 0$. Put $B = \{r \in T : \rho(r, s) \leq \delta\}$. Then

$$\begin{aligned} u^{-1} \mathbb{P}[U_1(s) \leq u < \lambda u < U_1(t)] &\leq u^{-1} \mathbb{P}\left[\inf_{r \in B} U_1(r) \leq u < \lambda u < \sup_{r \in B} U_1(r)\right] \\ &\leq u^{-1} \mathbb{P}\left[\inf_{r \in T} U_1(r) \leq u\right] \mathbb{P}\left[\sup_{r \in B} U_1(r) > \lambda u \mid \inf_{r \in B} U_1(r) \leq u\right] \end{aligned}$$

Recall Condition 5. Let $\varepsilon \in (0, \varepsilon_0]$ and let $\delta(\varepsilon)$ be as in that condition. Put $\lambda = \exp(c_1 \varepsilon^2)$. By Lemma 1, we have $S = \sup_{u > 0} u^{-1} \mathbb{P}[\inf_{r \in T} U_1(r) \leq u] < \infty$. As a consequence, we find

$$u^{-1} \mathbb{P}[U_1(s) \leq u < u \exp(c_1 \varepsilon^2) < U_1(t)] \leq S c_2 \varepsilon^2$$

for all $u \in (0, u_0]$ and thus

$$1 - R_{s,t}(1, \exp(c_1 \varepsilon^2)) \leq S c_2 \varepsilon^2.$$

By the 1-Lipschitz property of tail copulas, we find

$$\begin{aligned} 1 - R_{s,t}(1, 1) &= [1 - R_{s,t}(1, \exp(c_1 \varepsilon^2))] + [R_{s,t}(1, \exp(c_1 \varepsilon^2)) - R_{s,t}(1, 1)] \\ &\leq S c_2 \varepsilon^2 + \exp(c_1 \varepsilon^2) - 1. \end{aligned}$$

For a given $\eta > 0$, we can find $\varepsilon = \varepsilon(\eta)$ such that the previous upper bound is less than η . The choice $\delta = \delta(\varepsilon(\eta))$ then fulfills the requirement.

The last statement follows from the property just proved in combination with the increment bound in Lemma 3. $\square$

6. Proofs for the results of Section 3. We now prepare for the proof of Theorem 2. Recall $U_i(t) = 1 - F_t(\xi_i(t))$ for $(i, t) \in \{1, \dots, n\} \times T$ and let $U_{1:n}(t) \leq \dots \leq U_{n:n}(t)$ denote the order statistics of $U_1(t), \dots, U_n(t)$; set $U_{0:n}(t) = 0$. Put

$$\begin{aligned} K_{n,t}(x) &= \frac{n}{k} U_{[kx]:n}(t), \\ \mathbf{K}_{n,t}(\mathbf{x}) &= (K_{n,t_1}(x_1), \dots, K_{n,t_d}(x_d)), \end{aligned}$$

for $(t, x) \in T \times [0, n/k]$ and $(\mathbf{t}, \mathbf{x}) \in T^d \times [0, n/k]^d$. We will approximate the penultimate empirical tail copula process $\hat{W}_n$ by the process

$$\begin{aligned} (26) \quad \check{W}_n(\mathbf{v}) &= \sqrt{k} \{G_{n,\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - E_{n,\mathbf{t}}(\mathbf{x})\} \\ &= W_n(\mathbf{t}, \mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) + \sqrt{k} \{E_{n,\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - E_{n,\mathbf{t}}(\mathbf{x})\} \end{aligned}$$

for $\mathbf{v} = (\mathbf{t}, \mathbf{x}) \in \mathcal{V}$.

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

$$\sup_{\mathbf{v} \in \mathcal{V}} \left| \hat{W}_n(\mathbf{v}) - \check{W}_n(\mathbf{v}) \right| \xrightarrow{P} 0, \quad n \rightarrow \infty.$$

PROOF. For $t \in T$, let $\xi_{1:n}(t) \leq \dots \leq \xi_{n:n}(t)$ denote the ascending order statistics of $\xi_1(t), \dots, \xi_n(t)$. For $i, m \in \{1, \dots, n\}$, we have

$$\begin{aligned} \mathbb{1}\{U_i(t) < U_{m:n}(t)\} &\leq \mathbb{1}\{\xi_i(t) > \xi_{n-m+1:n}(t)\} = \mathbb{1}\{\text{rank}_{n,i}(t) > n - m + 1\} \\ &\leq \mathbb{1}\{\xi_i(t) \geq \xi_{n-m+1:n}(t)\} \\ &\leq \mathbb{1}\{U_i(t) \leq U_{m:n}(t)\}. \end{aligned}$$

Let $\mathbf{v} = (t, \mathbf{x}) \in \mathcal{V}$ and consider the right-hand limit

$$G_{n,t}(\mathbf{x}+) = \lim_{h \downarrow 0} G_{n,t}(\mathbf{x} + h\mathbf{1}) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) \leq \frac{k}{n}x_j\},$$

where $\mathbf{1}$ denotes a vector of ones. It follows that

$$\begin{aligned} G_{n,t}(\mathbf{K}_{n,t}(\mathbf{x})) &= \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) < U_{\lceil kx_j \rceil : n}(t_j)\} \\ &\leq \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : \text{rank}_{n,i}(t_j) > n - \lceil kx_j \rceil + 1\} = \hat{R}_{n,t}(\mathbf{x}) \\ &\leq \frac{1}{k} \sum_{i=1}^n \mathbb{1}\{\forall j = 1, \dots, d : U_i(t_j) \leq U_{\lceil kx_j \rceil : n}(t_j)\} = G_{n,t}(\mathbf{K}_{n,t}(\mathbf{x})+). \end{aligned}$$

To see the identity on the third line involving $\hat{R}_{n,t}(\mathbf{x})$, note that, for integer q and real y , we have $q < y \iff q < \lceil y \rceil$; as a consequence, the ceiling function can be omitted and we get the definition in Eq. (10). On the one hand, we have the lower bound

$$\begin{aligned} \hat{W}_n(\mathbf{v}) &= \sqrt{k} \{\hat{R}_{n,t}(\mathbf{x}) - E_{n,t}(\mathbf{x})\} \\ &\geq \sqrt{k} \{G_{n,t}(\mathbf{K}_{n,t}(\mathbf{x})) - E_{n,t}(\mathbf{x})\} = \check{W}_n(\mathbf{v}). \end{aligned}$$

On the other hand, since the map $\mathbf{x} \mapsto E_{n,t}(\mathbf{x})$ is continuous, using the same notation as above for right-hand limits, we have the upper bound

$$\begin{aligned} \hat{W}_n(\mathbf{v}) &\leq \sqrt{k} \{G_{n,t}(\mathbf{K}_{n,t}(\mathbf{x})+) - E_{n,t}(\mathbf{x})\} \\ &= W_n(t, \mathbf{K}_{n,t}(\mathbf{x})+) + \sqrt{k} \{E_{n,t}(\mathbf{K}_{n,t}(\mathbf{x})) - E_{n,t}(\mathbf{x})\} \\ &= \check{W}_n(\mathbf{v}) + \eta_{n,t}(\mathbf{x}), \end{aligned}$$

where the remainder term is

$$\eta_{n,t}(\mathbf{x}) = W_n(t, \mathbf{K}_{n,t}(\mathbf{x})+) - W_n(t, \mathbf{K}_{n,t}(\mathbf{x})) \geq 0.$$

As noted right after Theorem 1, the sequence W_n is asymptotically uniformly ρ_2 -equicontinuous in probability. By the definition (6) of ρ_2 and the uniform continuity of the map $\mathbf{x} \mapsto R_t(\mathbf{x})$, a property that follows from the Lipschitz property of a tail copula, we find that $\sup_{\mathbf{v} \in \mathcal{V}} |\eta_{n,t}(\mathbf{x})| \xrightarrow{P} 0$ as $n \rightarrow \infty$. $\square$

We need to know the asymptotic behavior of the empirical tail quantile function $K_{n,t}$ uniformly in $t \in T$. Since $K_{n,t}$ is a generalized inverse of the tail empirical distribution function $G_{n,t}$ in (2), we can obtain this through a Vervaat-type lemma as given in Lemma 4 of [Einmahl, Gantner and Sawitzki \(2010\)](#), in combination with Theorem 1 which gives the asymptotic behavior of $G_{n,t}(x)$, uniformly in $(t, x) \in T \times [0, M]$. Note that the functions in there are defined on $[0, 1]$ but for our purposes should be defined on $[0, M + 1]$ so that $G_{n,t}$ can be extended on $(M, M + 1]$ in such a way that $G_{n,t}(M + 1) = M + 1$. Thus we obtain the following corollary to Theorem 1.

COROLLARY 3. *Under the conditions of Theorem 1, we have*

$$\sup_{(t,x) \in T \times [0,M]} \left| \sqrt{k} \{K_{n,t}(x) - x\} + W_n(t, x) \right| \xrightarrow{p} 0, \quad n \rightarrow \infty.$$

PROOF OF THEOREM 2. Suppose that the asymptotic expansion of $\hat{W}_n$ in the convergence statement (12) has been established. Define a mapping Φ from $\ell^\infty(\mathcal{V})$ to $\ell^\infty(\mathcal{V}')$ by

$$(\Phi(f))(v) = f(v) - \sum_{j=1}^d \dot{R}_{t,j}(x) f(t_j, x_j), \quad v = (t, x) \in \mathcal{V}'.$$

The mapping is linear and bounded, since the partial derivatives of a tail copula are all between 0 and 1. By the continuous mapping theorem ([van der Vaart and Wellner, 1996](#), Theorem 1.3.6), we have $\Phi(W_n) \rightsquigarrow \Phi(W) = \tilde{W}$ as $n \rightarrow \infty$. Equation (12) further says that the distance in $\ell^\infty(\mathcal{V}')$ between $\hat{W}_n$ and $\Phi(W_n)$ converges to zero in probability as $n \rightarrow \infty$. By Slutsky's lemma ([van der Vaart and Wellner \(1996](#), Example 1.4.7) or [van der Vaart \(1998](#), Theorem 18.10(iv))), then also $\hat{W}_n \rightsquigarrow \tilde{W}$ as $n \rightarrow \infty$.

The process $\tilde{W}$ is Gaussian. It is also tight, since it arises as the image of a tight process by a continuous mapping, and the image of compact set by a continuous function is compact. By [van der Vaart and Wellner \(1996](#), Example 1.5.10), almost all trajectories of $\tilde{W}$ are uniformly $\hat{\rho}_2$ -continuous on $\mathcal{V}'$.

It remains to show the expansion in (12). By Lemma 2 and Slutsky's Lemma ([van der Vaart and Wellner, 1996](#), Example 1.4.7), we can replace the penultimate tail empirical copula process $\hat{W}_n$ by its approximation $\tilde{W}_n$ defined in Eq. (26). We analyze the two terms in the decomposition on the second line of (26) separately.

The first term in (26). We will show that

$$(27) \quad \sup_{(t,x) \in \mathcal{V}} |W_n(t, K_{n,t}(x)) - W_n(t, x)| = o_p(1), \quad n \rightarrow \infty.$$

Recall the standard deviation metric ρ_2 in (6). Since $R_{t,t}(x_1, x_2) = R_t(x_1 \wedge x_2)$, the minimum being coordinate-wise, we have

$$\rho_2((t, x_1), (t, x_2)) = \{R_t(x_1) - 2R_t(x_1 \wedge x_2) + R_t(x_2)\}^{1/2}.$$

The map $x \mapsto R_t(x)$ is 1-Lipschitz w.r.t. the L_1 -norm $|\cdot|_1$ on Euclidean space, whence

$$(28) \quad \begin{aligned} \rho_2((t, x_1), (t, x_2)) &\leq (|x_1 - x_1 \wedge x_2|_1 + |x_2 - x_1 \wedge x_2|_1)^{1/2} \\ &\leq 2^{1/2} |x_1 - x_2|_1^{1/2}. \end{aligned}$$

Corollary 3 implies $\sup_{(t,x) \in T \times [0,M]} |K_{n,t}(x) - x| = o_p(1)$ as $n \rightarrow \infty$, and thus also $\sup_{(t,x) \in \mathcal{V}} |K_{n,t}(x) - x|_1 = o_p(1)$ as $n \rightarrow \infty$. By (28), we get

$$\sup_{(t,x) \in \mathcal{V}} \rho_2((t, K_{n,t}(x)), (t, x)) = o_p(1), \quad n \rightarrow \infty.$$

In view of Theorem 1, the processes W_n are asymptotically uniformly ρ_2 -equicontinuous in probability, see Example 1.5.10 in [van der Vaart and Wellner \(1996\)](#). Eq. (27) follows.

The second term in (26). We aim to show that

$$(29) \quad \sup_{(\mathbf{t}, \mathbf{x}) \in \mathcal{V}'} \left| \sqrt{k} \{E_{n,\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - E_{n,\mathbf{t}}(\mathbf{x})\} + \sum_{j=1}^d \dot{R}_{\mathbf{t},j}(\mathbf{x}) W_n(t_j, x_j) \right| = o_p(1)$$

as $n \rightarrow \infty$. To this end, we bound the absolute value in (29) by the sum of three terms:

$$\begin{aligned} \Delta_{n,1}(\mathbf{t}, \mathbf{x}) &= \left| \sqrt{k} \{E_{n,\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - E_{n,\mathbf{t}}(\mathbf{x})\} - \sqrt{k} \{R_{\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - R_{\mathbf{t}}(\mathbf{x})\} \right|, \\ \Delta_{n,2}(\mathbf{t}, \mathbf{x}) &= \sqrt{k} \left| R_{\mathbf{t}}(\mathbf{K}_{n,\mathbf{t}}(\mathbf{x})) - R_{\mathbf{t}}\left(\mathbf{x} - \frac{1}{\sqrt{k}} \mathbf{W}_n(\mathbf{t}, \mathbf{x})\right) \right|, \end{aligned}$$

and

$$(30) \quad \Delta_{n,3}(\mathbf{t}, \mathbf{x}) = \left| \sqrt{k} \left\{ R_{\mathbf{t}}\left(\mathbf{x} - \frac{1}{\sqrt{k}} \mathbf{W}_n(\mathbf{t}, \mathbf{x})\right) - R_{\mathbf{t}}(\mathbf{x}) \right\} + \sum_{j=1}^d \dot{R}_{\mathbf{t},j}(\mathbf{x}) W_n(t_j, x_j) \right|,$$

where $\mathbf{W}_n(\mathbf{t}, \mathbf{x}) = (W_n(t_j, x_j) : j = 1, \dots, d)$. For each of the three terms, we will show that the supremum over $(\mathbf{t}, \mathbf{x})$ converges to zero in probability. Especially for the third term, the proof is challenging, as foreshadowed already in Remark 4.

The term $\Delta_{n,1}$. — Recall Theorem 1 and Corollary 3. For $\varepsilon > 0$, we can find $K > 0$ such that for sufficiently large n and with probability at least $1 - \varepsilon$, we have $\sup_{t \in T, x \in [0, M]} \sqrt{k} |K_{n,t}(x) - x| \leq K$. On this event, the supremum of $\Delta_{n,1}(\mathbf{t}, \mathbf{x})$ over all $(\mathbf{t}, \mathbf{x}) \in \mathcal{V}_d$ is bounded by the supremum in Condition 7 and is therefore bounded by $\varepsilon > 0$ for sufficiently large n . Since $\varepsilon > 0$ was arbitrary, we find that the supremum of $\Delta_{n,1}(\mathbf{t}, \mathbf{x})$ over $(\mathbf{t}, \mathbf{x}) \in \mathcal{V}$ converges to zero in probability.

The term $\Delta_{n,2}$. — The function $R_{\mathbf{t}}$ is 1-Lipschitz with respect to the L_1 -norm, so $\Delta_{n,2}(\mathbf{t}, \mathbf{x})$ is bounded by $\sum_{j=1}^d |\sqrt{k} \{K_{n,t_j}(x_j) - x_j\} + W_n(t_j, x_j)|$. The supremum of $\Delta_{n,2}(\mathbf{t}, \mathbf{x})$ over all $(\mathbf{t}, \mathbf{x}) \in \mathcal{V}$ is thus bounded by D times the supremum in Corollary 3, and therefore converges to zero in probability.

The term $\Delta_{n,3}$. — In what follows, fix $d = 1, 2, \dots, D$. We consider the supremum of $\Delta_{n,3}(\mathbf{t}, \mathbf{x})$ over $(\mathbf{t}, \mathbf{x}) \in \mathcal{V}_d'$. Fix $\varepsilon > 0$ and let $\delta = \delta(\varepsilon) > 0$ be chosen later on as a function of ε in such a way that a number of criteria are met. We bound the supremum as the maximum of the suprema over the d sets $T^d \times \{\mathbf{x} \in [0, M]^d : x_j \leq \delta\}$ for $j = 1, \dots, d$ on the one hand (Case I) and over $(T^d)' \times [\delta, M]^d$ (Case II) on the other hand, where $(T^d)'$ comprises the d -tuples with all coordinates different and in T . We show that, for some constants $c_1, c_2 > 0$ not depending on ε or n , for all sufficiently large n , each of these suprema is less than $c_1 \varepsilon$ with probability at least $1 - c_2 \varepsilon$.

Case I. — We first consider the supremum of $\Delta_{n,3}(\mathbf{t}, \mathbf{x})$ over $\mathbf{t} \in T^d$ and $\mathbf{x} \in [0, M]^d$ with $x_j \leq \delta$, for $j \in \{1, \dots, d\}$ fixed and for $\delta = \delta(\varepsilon) > 0$ to be chosen in function of a fixed $\varepsilon > 0$. For all $\eta > 0$ and for all $(t, x) \in T \times [0, \eta]$, we have

$$\rho_2((t, x), (t, 0)) = \sqrt{x} \leq \sqrt{\eta},$$

whereas $\rho_2((s, 0), (t, 0)) = 0$ for $(s, t) \in T^2$. Furthermore, $W_n(t, 0) = 0$ for all $t \in T$. Recall that the processes W_n are asymptotically uniformly ρ_2 -equicontinuous in probability. We obtain that, for sufficiently small δ , setting $\eta = \sqrt{\delta}$, we have $\rho_2((t, x), (t, 0)) \leq \delta^{1/4}$ for all $(t, x) \in T \times [0, \sqrt{\delta}]$, so that the supremum of $|W_n(t, x)|$ over $(t, x) \in T \times [0, \sqrt{\delta}]$ is bounded by ε with probability at least $1 - \varepsilon$, for large n . On this event and for $\mathbf{t} \in T^d$ and $\mathbf{x} \in [0, M]^d$

such that $x_j \leq \delta$, the 1-Lipschitz property of R_t implies that the term $\Delta_{n,3}(\mathbf{t}, \mathbf{x})$ is bounded by

$$(31) \quad \left| \sqrt{k} \left\{ R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \widetilde{\mathbf{W}}_{n,j}(\mathbf{t}, \mathbf{x}) \right) - R_t(\mathbf{x}) \right\} \right| + \sum_{i \in \{1, \dots, d\} \setminus \{j\}} \dot{R}_{t,i}(\mathbf{x}) |W_n(t_i, x_i)| + 2\varepsilon,$$

where the random d -vector $\widetilde{\mathbf{W}}_{n,j}(\mathbf{t}, \mathbf{x})$ has the same d coordinates $W_n(t_i, x_i)$ as $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$, except for the j th one, which is zero.

For the term $\dot{R}_{t,i}(\mathbf{x}) |W_n(t_i, x_i)|$, we consider again two cases.

- If, on the one hand, $x_i \leq \sqrt{\delta}$, then $|W_n(t_i, x_i)| \leq \varepsilon$ on the event already considered, while $0 \leq \dot{R}_{t,i}(\mathbf{x}) \leq 1$ by the 1-Lipschitz property of R_t .
- If, on the other hand, $x_i > \sqrt{\delta}$, then, by inequalities (37) and (38) in Lemma 5, we have

$$\dot{R}_{t,i}(\mathbf{x}) \leq R_t(\mathbf{x})/x_i \leq x_j/x_i \leq \delta/\sqrt{\delta} = \sqrt{\delta}.$$

Moreover, given ε , we can find $K = K(\varepsilon) > 0$ such that the supremum of $|W_n(t, x)|$ over $(t, x) \in T \times [0, M]$ is bounded by K with probability at least $1 - \varepsilon$.

Hence, with probability at least $1 - 2\varepsilon$, we have $0 \leq \dot{R}_{t,i}(\mathbf{x}) |W_n(t_i, x_i)| \leq \max(\varepsilon, K\sqrt{\delta})$ for all $\mathbf{t} \in T^d$ and all $\mathbf{x} \in [0, M]^d$ with $x_j \leq \delta$.

For the first term in (31), a similar argument works on the same event (which has probability $1 - 2\varepsilon$). For those indices $i \in \{1, \dots, d\} \setminus \{j\}$ for which $x_i \leq \sqrt{\delta}$, we can replace $W_n(t_i, x_i)$ by 0 at the cost of an additional error of at most ε . This amounts to an error of at most $(d-1)\varepsilon$. For the remaining indices i , we write the increment as a telescoping sum of increments involving changes in one coordinate at a time. For each increment in the sum, we apply (38), with upper bound $x_j \leq \delta$ and we proceed as in the second bullet point above. This yields an error of at most $(d-1)K\sqrt{\delta}$.

Given $\varepsilon > 0$ and $j \in \{1, \dots, d\}$, we have thus constructed an event with probability at least $1 - 2\varepsilon$ such that the supremum of $\Delta_{n,t}(\mathbf{x})$ over $\mathbf{t} \in T^d$ and $\mathbf{x} \in [0, M]^d$ with $x_j \leq \delta$ is bounded by a constant multiple (the constant depending on d only) of $\varepsilon + K\sqrt{\delta}$, where K depends on ε and where δ remains to be chosen. By choosing δ small enough, we can ensure that $K\sqrt{\delta} \leq \varepsilon$. We get that the supremum is less than a constant multiple of ε on an event with probability at least $1 - 2\varepsilon$.

Case II. – Finally we consider the supremum of $\Delta_{n,3}(\mathbf{t}, \mathbf{x})$ in (30) over $(\mathbf{t}, \mathbf{x}) \in (T^d)' \times [\delta, M]^d$. We write the supremum as the maximum over all partitions Π of $\{1, \dots, d\}$ of the suprema over those $(\mathbf{t}, \mathbf{x})$ that satisfy the following property related to a given partition Π : for indices $i, j \in \{1, \dots, d\}$ within the same component of the partition, we have $\rho_2((t_i, x_i), (t_j, x_j)) \leq \delta$ while for indices i and j in different components, we have $\rho_2((t_i, x_i), (t_j, x_j)) > \delta/(d-1)$. In this way, every point $(\mathbf{t}, \mathbf{x})$ is taken into account at least once in view of Lemma 4.

The number of partitions being finite, we can fix a single partition $\Pi = \{D_\alpha : \alpha \in A\}$ of $\{1, \dots, d\}$ and study the associated supremum. Consider the component $D_{\alpha(1)}$ of the partition that contains the index 1, let $d_1 = |D_{\alpha(1)}| \in \{1, \dots, d\}$ be the number of elements of this component, and assume for convenience of notation that this component is equal to $D_{\alpha(1)} = \{1, \dots, d_1\}$.

Let $(\mathbf{t}, \mathbf{x})$ be a point satisfying the property associated to the given partition as described above. The difference $\sqrt{k} \{ R_t(\mathbf{x} - \frac{1}{\sqrt{k}} \mathbf{W}_n(\mathbf{t}, \mathbf{x})) - R_t(\mathbf{x}) \}$ can be written as a telescoping sum with as many terms as the number $|A|$ of components D_α in the partition Π : we replace $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$ by the zero vector, one component D_α at a time. We will only consider the first term of the telescoping sum, i.e., the one in which the first d_1 elements of $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$ are replaced by zeroes; the other terms can be treated similarly.

Let the random d -vector $\overline{\mathbf{W}}_n(\mathbf{t}, \mathbf{x})$ have the last $d - d_1$ coordinates equal to those of $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$, whereas its first d_1 coordinates are all equal and are the average of the first d_1 coordinates of $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$. This average is denoted below by $\overline{w}$. Similarly, $\widetilde{\mathbf{W}}_n(\mathbf{t}, \mathbf{x})$ has the last $d - d_1$ coordinates equal to those of $\mathbf{W}_n(\mathbf{t}, \mathbf{x})$, whereas its first d_1 coordinates are all equal to zero. We will consider the first term in the aforementioned telescoping sum: $\sqrt{k} \left\{ R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \mathbf{W}_n(\mathbf{t}, \mathbf{x}) \right) - R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \widetilde{\mathbf{W}}_n(\mathbf{t}, \mathbf{x}) \right) \right\}$.

Applying Lemma 8 we obtain

$$(32) \quad \left| \sqrt{k} \left\{ R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \mathbf{W}_n(\mathbf{t}, \mathbf{x}) \right) - R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \overline{\mathbf{W}}_n(\mathbf{t}, \mathbf{x}) \right) \right\} \right| \leq \sum_{j=1}^{d_1} |W_n(t_j, x_j) - \overline{w}|.$$

For the second term on the right-hand side in (32), we invoke again the asymptotic uniform ρ_2 -equicontinuity in probability of the processes W_n . Since all points (t_j, x_j) for $j \in \{1, \dots, d_1\}$ are within ρ_2 -distance δ , the second term is uniformly bounded by $d_1 \varepsilon$ with probability at least $1 - \varepsilon$, for δ small enough and for sufficiently large n .

By the mean-value theorem,

$$(33) \quad \sqrt{k} \left\{ R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \overline{\mathbf{W}}_n(\mathbf{t}, \mathbf{x}) \right) - R_t \left(\mathbf{x} - \frac{1}{\sqrt{k}} \widetilde{\mathbf{W}}_n(\mathbf{t}, \mathbf{x}) \right) \right\} = -\overline{w} \sum_{j=1}^{d_1} \dot{R}_{t,j}(\Theta_n),$$

where $\Theta_n = \Theta_n(\mathbf{t}, \mathbf{x})$ is a random point on the line segment connecting $\mathbf{x} - \frac{1}{\sqrt{k}} \widetilde{\mathbf{W}}_n(\mathbf{t}, \mathbf{x})$ and $\mathbf{x} - \frac{1}{\sqrt{k}} \overline{\mathbf{W}}_n(\mathbf{t}, \mathbf{x})$. Clearly the first d_1 coordinates of $\Theta_n - \mathbf{x}$ are all equal to $\bar{h} = \bar{h}_n(\mathbf{t}, \mathbf{x})$, say. We would like to show that $\sum_{j=1}^{d_1} \dot{R}_{t,j}(\Theta_n)$ is uniformly close to $\sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{x})$, with high probability.

Define the random d -vector $\mathbf{y}_n = \mathbf{y}_n(\mathbf{t}, \mathbf{x})$ to have the last $d - d_1$ coordinates equal to those of $\mathbf{x}$ and the first d_1 coordinates equal to those of Θ_n . For every $j \in \{1, \dots, d_1\}$ we have from $d - d_1$ applications of (41) and Lemma 12 that $|\dot{R}_{t,j}(\Theta_n) - \dot{R}_{t,j}(\mathbf{y}_n)| \leq \varepsilon$ uniformly over the considered $(\mathbf{t}, \mathbf{x})$, with probability at least $1 - \varepsilon$, provided δ is small enough and n is large.

Now Eq. (36) yields

$$\begin{aligned} \sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{x}) &= \frac{1}{x_1} \sum_{j=1}^{d_1} x_j \dot{R}_{t,j}(\mathbf{x}) + \frac{1}{x_1} \sum_{j=1}^{d_1} (x_1 - x_j) \dot{R}_{t,j}(\mathbf{x}) \\ &= -\frac{1}{x_1} \sum_{j=d_1+1}^d x_j \dot{R}_{t,j}(\mathbf{x}) + \frac{1}{x_1} R_t(\mathbf{x}) + \frac{1}{x_1} \sum_{j=1}^{d_1} (x_1 - x_j) \dot{R}_{t,j}(\mathbf{x}), \end{aligned}$$

and we can rewrite $\sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{y}_n)$ similarly. Hence by (39) it follows after some algebra that

$$\begin{aligned} \left| \sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{y}_n) - \sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{x}) \right| &\leq \frac{M}{\delta} \sum_{j=d_1+1}^d \left| \dot{R}_{t,j}(\mathbf{y}_n) - \dot{R}_{t,j}(\mathbf{x}) \right| \\ &\quad + \frac{d_1 |\bar{h}|}{\delta} + \frac{1}{\delta(\delta + \bar{h})} dM |\bar{h}| + 2d_1 \frac{\delta^2}{\delta \wedge (\delta + \bar{h})}. \end{aligned}$$

Take δ such that $3d_1\delta \leq \varepsilon$. Then since $\bar{h} = O_p(1/\sqrt{k})$ uniformly in $(\mathbf{t}, \mathbf{x})$, we have, with probability at least $1 - \varepsilon$, that the sum of the three terms on the second line is uniformly bounded by ε when n is large.

So it remains to consider $|\dot{R}_{t,j}(\mathbf{y}_n) - \dot{R}_{t,j}(\mathbf{x})|$ for $j \in \{d_1 + 1, \dots, d\}$, that is, the first d_1 coordinates of $\mathbf{y}_n$ have to be replaced consecutively by those of $\mathbf{x}$ by writing the difference as

a telescoping sum. We will consider for convenience only the first term in that sum. Let $\mathbf{y}'_n$ have the same coordinates as $\mathbf{y}_n$, except for the first one, which is x_1 . By Ineq. (41) we have

$$|\dot{R}_{t,j}(\mathbf{y}_n) - \dot{R}_{t,j}(\mathbf{y}'_n)| \leq |\dot{R}_{(t_1, t_j), 2}(x_1 + \bar{h}, x_j) - \dot{R}_{(t_1, t_j), 2}(x_1, x_j)|.$$

By Lemma 12, the latter difference is uniformly small since $\bar{h}$ tends to zero in probability, uniformly.

We have thus shown that in (33), with large probability and for large n , the sum $\sum_{j=1}^{d_1} \dot{R}_{t,j}(\Theta_n)$ may be replaced by $\sum_{j=1}^{d_1} \dot{R}_{t,j}(\mathbf{x})$ at the cost of an error bounded by a constant multiple of ε . It remains to replace $\bar{w} = \frac{1}{d_1} \sum_{l=1}^{d_1} W_n(t_l, x_l)$ in (33) by $W_n(t_j, x_j)$ for $j \in \{1, \dots, d_1\}$, written inside the sum over $j = 1, \dots, d_1$. The property that the resulting difference is small follows from the asymptotic ρ_2 -equicontinuity in probability of W_n .

We have thereby found a suitable upper bound on the supremum of $\Delta_{n,3}(\mathbf{t}, \mathbf{x})$ over points $(\mathbf{t}, \mathbf{x}) \in (T^d)' \times [\delta, M]^d$ that satisfy the property associated to a fixed partition Π of $\{1, \dots, d\}$ as described in the opening paragraph of Case II. This completes the analysis of the term $\Delta_{n,3}$ in Eq. (30) and thus of the convergence statement (12). $\square$

PROOF OF COROLLARY 2. The supremum in Condition 7 is bounded by twice the supremum in Condition 8, if we replace M in the latter condition by $M + 1$ and if k is sufficiently large. Hence, Condition 8 implies Condition 7.

The difference between $\dot{W}_n(\mathbf{v})$ and $\sqrt{k}\{\dot{R}_{n,\mathbf{t}}(\mathbf{x}) - R_{\mathbf{t}}(\mathbf{x})\}$ is bounded by the maximum over $d = 1, \dots, D$ of the suprema in Condition 8. Under that condition, the two processes must thus share the same first-order asymptotic expansion and the same weak convergence properties in $\ell^\infty(\mathcal{V})$ and thus also in $\ell^\infty(\mathcal{V}')$. $\square$

APPENDIX A: AUXILIARY RESULTS

LEMMA 3. *Let R be as in Condition 3. For $\mathbf{s}, \mathbf{t} \in T^d$ and $\mathbf{x}, \mathbf{y} \in [0, \infty)^d$, we have*

$$|R_{\mathbf{s}}(\mathbf{x}) - R_{\mathbf{t}}(\mathbf{y})| \leq \sum_{j=1}^d [|x_j - y_j| + (x_j \wedge y_j)\{1 - R_{s_j, t_j}(1, 1)\}].$$

PROOF. Let $\bar{s}, s \in T$ and $\bar{\mathbf{t}} \in T^m$ for some m . Then

$$\begin{aligned} R_{\bar{s}, \bar{\mathbf{t}}}(x, \bar{\mathbf{y}}) - R_{s, \bar{\mathbf{t}}}(x, \bar{\mathbf{y}}) &= R_{\bar{s}, s, \bar{\mathbf{t}}}(x, \infty, \bar{\mathbf{y}}) - R_{\bar{s}, s, \bar{\mathbf{t}}}(\infty, x, \bar{\mathbf{y}}) \\ &\leq R_{\bar{s}, s, \bar{\mathbf{t}}}(x, \infty, \bar{\mathbf{y}}) - R_{\bar{s}, s, \bar{\mathbf{t}}}(x, x, \bar{\mathbf{y}}) \\ &\leq R_{\bar{s}, s}(x, \infty) - R_{\bar{s}, s}(\infty, x) \\ &= x - R_{\bar{s}, s}(x, x) \\ &= x(1 - R_{\bar{s}, s}(1, 1)). \end{aligned}$$

The same inequality holds with the roles of $\bar{s}$ and s interchanged. We obtain

$$(34) \quad |R_{\bar{s}, \bar{\mathbf{t}}}(x, \bar{\mathbf{y}}) - R_{s, \bar{\mathbf{t}}}(x, \bar{\mathbf{y}})| \leq x(1 - R_{\bar{s}, s}(1, 1)).$$

Consider the statement of the lemma. Let $z_j = x_j \wedge y_j$ for $j = 1, \dots, d$. Then $|R_{\mathbf{s}}(\mathbf{x}) - R_{\mathbf{t}}(\mathbf{y})|$ is bounded by

$$(35) \quad |R_{\mathbf{s}}(\mathbf{x}) - R_{\mathbf{s}}(\mathbf{z})| + |R_{\mathbf{s}}(\mathbf{z}) - R_{\mathbf{t}}(\mathbf{z})| + |R_{\mathbf{t}}(\mathbf{z}) - R_{\mathbf{t}}(\mathbf{y})|.$$

The sum of the first and third terms in (35) is bounded by

$$\sum_{j=1}^d (|z_j - x_j| + |z_j - y_j|) = \sum_{j=1}^d |x_j - y_j|.$$

The middle term in (35) can be written as a telescoping sum of d terms, in which the coordinate s_j is replaced by t_j one coordinate at a time. Apply the triangle inequality to bound the absolute value of the sum by the sum of the absolute values. The j th term in the resulting sum is bounded by $z_j(1 - R_{s_j, t_j}(1, 1))$ in view of (34). $\square$

Note that (34) shows that $R_t, t \in T^d$, is continuous in each t_j with respect to the ρ_2 -semimetric on T . By Theorem 25.7 in Rockafellar (1970), the partial derivatives of $R_t(x)$ with respect to x_j , for $j \in \{1, \dots, d\}$, are then continuous in t as well.

LEMMA 4. *Let $x_1, \dots, x_d$ be d points in a semimetric space $(\mathcal{X}, \tilde{\rho})$ and let $\delta > 0$. Then there exists a partition Π of $\{1, \dots, d\}$ such that for all indices i, j that belong to the same component of the partition we have $\tilde{\rho}(x_i, x_j) \leq \delta$ while for all indices i, j that belong to different components of the partition we have $\tilde{\rho}(x_i, x_j) > \delta/(d-1)$.*

PROOF. Construct a graph on $\{1, \dots, d\}$ by connecting different indices i and j as soon as $\tilde{\rho}(x_i, x_j) \leq \delta/(d-1)$. Let Π be the partition formed by the connectivity components of the graph. Indices i and j that belong to different components are not connected by an edge and thus $\tilde{\rho}(x_i, x_j) > \delta/(d-1)$. Indices i and j that belong to the same component are linked up through a chain of at most $d-1$ edges. By the triangle inequality, $\tilde{\rho}(x_i, x_j) \leq \delta$. $\square$

LEMMA 5. *If a d -variate tail copula R is differentiable on $(0, \infty)^d$, then, for all $x \in [0, \infty)^d$, we have*

$$(36) \quad \sum_{j=1}^d x_j \dot{R}_j(x) = R(x)$$

(both sides being zero if $x_j = 0$ for some $j = 1, \dots, d$) and hence for all $j \in \{1, \dots, d\}$,

$$(37) \quad x_j \dot{R}_j(x) \leq R(x).$$

In addition, for all $h \geq -x_1$ such that $h \neq 0$, we have

$$(38) \quad x_1 \frac{R(x_1 + h, x_2, \dots, x_d) - R(x)}{h} \leq R(x) \leq \min(x_1, \dots, x_d).$$

PROOF. Evaluating the derivative of the function $\lambda \mapsto R(\lambda x) = \lambda R(x)$ in $\lambda = 1$ yields the identity in (36) on $(0, \infty)^d$. The identity extends to $x \in [0, \infty)^d$ since for vectors x of which one or more coordinates are zero, both sides of the above equation are zero: if $x_j = 0$, then the definition of $\dot{R}_j(x)$ is immaterial, while for $i \neq j$, we have $\dot{R}_j(x) = 0$.

Inequality (37) is a consequence of (36) and the fact that R is non-decreasing in each variable.

For the first inequality in (38): for $h = -x_1$, the inequality becomes an equality, and by concavity of R , the left-hand side is non-increasing in h . The second inequality in (38) is immediate from the definition (and well-known). $\square$

Let ρ_2 be the standard deviation semimetric associated to a spatial tail copula R as introduced in Eq. (6) in the main paper.

LEMMA 6. *For $s, t \in T$ and $x, y \in [0, \infty)$, we have*

$$(39) \quad [\rho_2((s, x), (t, y))]^2 = |x - y| + 2((x \wedge y) - R_{s,t}(x, y))$$

$$(40) \quad \leq |x - y| + 2(x \wedge y)(1 - R_{s,t}(1, 1)).$$

PROOF. Since $x - 2(x \wedge y) + y = |x - y|$, we have

$$\begin{aligned} \rho_2((s, x), (t, y))^2 &= \mathbb{E}[\{W(s, x) - W(t, y)\}^2] \\ &= x - 2R_{s,t}(x, y) + y \\ &= x - 2(x \wedge y) + y + 2((x \wedge y) - R_{s,t}(x, y)) \\ &= |x - y| + 2((x \wedge y) - R_{s,t}(x, y)), \end{aligned}$$

which is (39). The inequality in (40) then follows from $R_{s,t}(x, y) \geq R_{s,t}(x \wedge y, x \wedge y) = (x \wedge y)R_{s,t}(1, 1)$. $\square$

Recall the semimetric $\rho_{2,T}$ on T induced by ρ_2 as defined in Eq. (11) in the main paper. Then (40) states that

$$[\rho_2((s, x), (t, y))]^2 \leq |x - y| + (x \wedge y)[\rho_{2,T}(s, t)]^2,$$

splitting the semimetric on $T \times [0, \infty)$ into a component on T and a component on $[0, \infty)$.

LEMMA 7. For $(s_n, x_n), (s, x) \in T \times [0, \infty)$, we have, as $n \rightarrow \infty$,

$$\rho_2((s_n, x_n), (s, x)) \rightarrow 0 \iff [x_n \rightarrow x \text{ and } x(R_{s_n,s}(1, 1) - 1) \rightarrow 0].$$

PROOF. By (39), we have

$$\rho_2((s_n, x_n), (s, x))^2 = |x_n - x| + 2((x_n \wedge x) - R_{s_n,s}(x_n, x)).$$

If $x = 0$, then $\rho_2((s_n, x_n), (s, x))^2 = x_n$, which converges to zero if and only if $x_n \rightarrow 0 = x$, as required. Henceforth, suppose $x > 0$.

On the one hand, if $x_n \rightarrow x$ and $R_{s_n,s}(1, 1) \rightarrow 1$, then by (40) we also have $\rho_2((s_n, x_n), (s, x)) \rightarrow 0$.

On the other hand, if $\rho_2((s_n, x_n), (s, x)) \rightarrow 0$, then by (39), both $x_n \rightarrow x$ and $(x_n \wedge x) - R_{s_n,s}(x_n, x) \rightarrow 0$. Since $x_n \vee x - x_n \wedge x \rightarrow 0$, also $x_n \vee x - R_{s_n,s}(x_n, x) \rightarrow 0$. But $R_{s_n,s}(x_n, x) \leq R_{s_n,s}(x_n \vee x, x_n \vee x) = (x_n \vee x)R_{s_n,s}(1, 1)$, and thus also $(x_n \vee x)(1 - R_{s_n,s}(1, 1)) \rightarrow 0$. $\square$

LEMMA 8. Let R be a d -variate tail copula, where $d = d_1 + d_2$. For $(\mathbf{x}, \mathbf{y}) \in [0, \infty)^{d_1} \times [0, \infty)^{d_2}$ and for $\mathbf{h} \in \mathbb{R}^{d_1}$ such that $x_j + h_j \geq 0$ for all $j = 1, \dots, d_1$, we have

$$|R(\mathbf{x} + \mathbf{h}, \mathbf{y}) - R(\mathbf{x} + \bar{h}\mathbf{1}, \mathbf{y})| \leq \sum_{j=1}^{d_1} |h_j - \bar{h}|$$

with $\bar{h} = (h_1 + \dots + h_{d_1})/d_1$ and with $\mathbf{1} = (1, \dots, 1)$ of dimension d_1 .

PROOF. This follows immediately from the 1-Lipschitz property of a tail copula with respect to the L_1 -norm. $\square$

LEMMA 9. Let R be a d -variate ($d \geq 2$) tail copula. Assume its partial derivatives $\dot{R}_j$ exist. Write $R(x, y)$ for $R(x, y, \infty)$. For $g > 0$ and $(x, y, z) \in [0, \infty) \times [0, \infty) \times [0, \infty)^{d-2}$, we have

$$(41) \quad 0 \leq \dot{R}_1(x, y + g, z) - \dot{R}_1(x, y, z) \leq \dot{R}_1(x, y + g) - \dot{R}_1(x, y).$$

PROOF. Since R is the distribution function of a nonnegative Borel measure on $[0, \infty)^d$, we have

$$\begin{aligned} 0 &\leq \frac{R(x+h, y+g, z) - R(x, y+g, z)}{h} - \frac{R(x+h, y, z) - R(x, y, z)}{h} \\ &\leq \frac{R(x+h, y+g) - R(x, y+g)}{h} - \frac{R(x+h, y) - R(x, y)}{h}. \end{aligned}$$

Taking the limit as $h \downarrow 0$, we find (41) (for the right-hand partial derivatives). $\square$

LEMMA 10. *Assume Condition 6. Let $s_n, t_n, s, t \in T$ and $x_n, y_n, x, y \in [0, \infty)$. If $x > 0$ and $\rho_2((s, x), (t, y)) > 0$ and if $\rho_2((s_n, x_n), (s, x)) \rightarrow 0$ and $\rho_2((t_n, y_n), (t, y)) \rightarrow 0$, then $\dot{R}_{s_n, t_n; 1}(x_n, y_n) \rightarrow \dot{R}_{s, t; 1}(x, y)$ as $n \rightarrow \infty$.*

PROOF. Since $x > 0$, the assumption $\rho_2((s_n, x_n), (s, x)) \rightarrow 0$ is equivalent to $x_n \rightarrow x$ and $R_{s_n, s}(1, 1) \rightarrow 1$ by Lemma 7. We consider two cases: $y = 0$ and $y > 0$.

First suppose $y = 0$, so that $\dot{R}_{s, t; 1}(x, y) = 0$. Since $\rho_2((t_n, y_n), (t, y)) \rightarrow 0$, Lemma 7 implies that $y_n \rightarrow 0$ as well. By (37), we have $\dot{R}_{s_n, t_n; 1}(x_n, y_n) \leq x_n^{-1} R_{s_n, t_n}(x_n, y_n) \leq x_n^{-1} y_n \rightarrow 0$ too, since $x_n \rightarrow x > 0$.

Second suppose $y > 0$. By Lemma 7, the assumption $\rho_2((t_n, y_n), (t, y)) \rightarrow 0$ implies $y_n \rightarrow y$ and $R_{t_n, t}(1, 1) \rightarrow 1$. Lemma 3 ensures that the functions $f_n = R_{s_n, t_n}$ converge to the function $f = R_{s, t}$ uniformly on bounded subsets of $[0, \infty)^2$. Since $\rho_2((s, x), (t, y)) > 0$, we have $x \neq y$ or $R_{s, t}(1, 1) < 1$ (or both) in view of Eq. (39). In any case, the functions f and f_n (for sufficiently large n) are continuously differentiable in an open neighborhood O of (x, y) : if $R_{s, t}(1, 1) < 1$ this is true by Condition 6, while if $R_{s, t}(1, 1) = 1$, then this is true by the fact that $f = \min$ and $x \neq y$ and for f_n we use either Condition 6 or that $f_n = \min$. Let N be a compact neighbourhood of (x, y) contained in O . By Rockafellar (1970, Theorem 25.7), the gradients of f_n converge to the one of f uniformly on N . Convergence of the partial derivatives as stated in the lemma follows. $\square$

LEMMA 11. *If the semimetric space $(T, \rho_{2, T})$ is complete, then so is the semimetric space $(T \times [0, \infty), \rho_2)$*

PROOF. Let $(s_n, x_n)_n$ be a ρ_2 -Cauchy sequence in $T \times [0, \infty)$: for every $\varepsilon > 0$, there exists an integer $n(\varepsilon) \geq 1$ such that for all integers $k, \ell \geq n(\varepsilon)$ we have $[\rho_2((s_k, x_k), (s_\ell, x_\ell))]^2 \leq \varepsilon$. We need to find $(s, x) \in T \times [0, \infty)$ with the property that $\rho_2((s_n, x_n), (s, x)) \rightarrow 0$ as $n \rightarrow \infty$.

By (39), we have $|x_k - x_\ell| \leq [\rho_2((s_k, x_k), (s_\ell, x_\ell))]^2 \leq \varepsilon$ whenever $k, \ell \geq n(\varepsilon)$. Hence $(x_n)_n$ is a Cauchy sequence in $[0, \infty)$. Since $[0, \infty)$ is complete, there exists $x \in [0, \infty)$ so that $x_n \rightarrow x$ as $n \rightarrow \infty$.

Suppose first that $x = 0$. Then for every $s \in T$, we have $\rho_2((s_n, x_n), (s, 0)) = x_n \rightarrow 0$ as $n \rightarrow \infty$, so that (s_n, x_n) converges to $(s, 0)$. [Note that ρ_2 is a semimetric on $T \times [0, \infty)$ since $\rho((s, 0), (t, 0)) = 0$ for any $s, t \in T$, even if $\rho_{2, T}$ is a metric on T .]

Suppose next that $x > 0$. Then by (39) we have $(x_k \wedge x_\ell) - R_{s_k, s_\ell}(x_k, x_\ell) \leq \varepsilon/2$ whenever $k, \ell \geq n(\varepsilon)$. At the same time, we can increase $n(\varepsilon)$ if necessary to ensure that also $|x_k \wedge x_\ell - x| \leq \varepsilon/4$ and $|R_{s_k, s_\ell}(x_k, x_\ell) - R_{s_k, s_\ell}(x, x)| \leq |x_k - x| + |x_\ell - x| \leq \varepsilon/4$ for all integer $k, \ell \geq n(\varepsilon)$. But then also $x(1 - R_{s_k, s_\ell}(1, 1)) = x - R_{s_k, s_\ell}(x, x) \leq \varepsilon$ for all $k, \ell \geq n(\varepsilon)$. Hence, $(s_n)_n$ is a Cauchy sequence in $(T, \rho_{2, T})$. By assumption, we can find $s \in T$ such that $1 - R_{s_n, s}(1, 1) \rightarrow 1$ as $n \rightarrow \infty$. But then (s_n, x_n) converges to (s, x) with respect to ρ_2 , as required. $\square$

LEMMA 12. Assume Condition 6 and assume that $(T, \rho_{2,T})$ is totally bounded. Let $0 < \delta < M < \infty$ and consider the set

$$K = \{(s, x; t, y) \in (T \times [\delta, M]) \times (T \times [0, M]) : \rho_2((s, x), (t, y)) \geq \delta\}.$$

For every $\varepsilon > 0$, there exists $\eta > 0$ such that for all $(s_i, t_i; x_i, y_i) \in K$, $i \in \{1, 2\}$, we have

$$\begin{aligned} & \rho_2((s_1, x_1), (s_2, x_2)) + \rho_2((t_1, y_1), (t_2, y_2)) \leq \eta \\ \implies & \left| \dot{R}_{s_1, t_1; 1}(x_1, y_1) - \dot{R}_{s_2, t_2; 1}(x_2, y_2) \right| \leq \varepsilon. \end{aligned}$$

PROOF. Consider the product space $\mathbb{D} = (T \times [0, \infty)) \times (T \times [0, \infty))$ equipped with the sum semimetric induced by the semimetric ρ_2 on $T \times [0, \infty)$. By Lemma 11, the product space $\mathbb{D}$ is complete. The set K in the statement is a closed subset of $\mathbb{D}$ and is thus complete as well. Moreover, K is totally bounded, since $(T, \rho_{2,T})$ is totally bounded and since ρ_2 can be bounded as in (40). As a consequence, K is compact with respect to the sum semimetric induced by ρ_2 .

By Lemma 10, the function f with domain $\mathbb{D}_1 = \{(s, x; t, y) \in \mathbb{D} : \rho_2((s, x), (t, y)) > 0, x > 0\}$ and defined by $f(s, x; t, y) = \dot{R}_{s, t; 1}(x, y)$ is continuous. Since K is a compact subset of $\mathbb{D}_1$, the function f is uniformly continuous on K . But this is exactly the statement of the lemma. $\square$

APPENDIX B: TESTING STATIONARITY: FURTHER DETAILS

We provide additional details to the results and calculations concerning the stationarity test in Section 4 in the paper. Notation not defined here is as in that section.

B.1. The limit distribution of the test statistic (fixed N). The population equivalent of the statistic $\hat{I}_n^{(N)}$ in Eq. (16) is

$$I^{(N)}(r_0/N) = \frac{1}{2\Delta} \sum_{r: 1 \leq |r-r_0| \leq \Delta} R_{r/N, r_0/N}(1, 1), \quad r_0 \in \{\Delta, \dots, N - \Delta\}.$$

Under the null hypothesis of stationarity, $I^{(N)}(r_0/N)$ is constant in r_0 . As a consequence, the test statistic in Eq. (17) is equal to

$$D_n^{(N)} = \max_{r_0=\Delta, \dots, N-\Delta} V_n^{(N)}(r_0/N) - \min_{r_0=\Delta, \dots, N-\Delta} V_n^{(N)}(r_0/N)$$

where

$$\begin{aligned} V_n^{(N)}(r_0/N) &= \sqrt{k} \left\{ \hat{I}^{(N)}(r_0/N) - I^{(N)}(r_0/N) \right\} \\ (42) \quad &= \frac{1}{2\Delta} \sum_{r: 1 \leq |r-r_0| \leq \Delta} \sqrt{k} \left\{ \hat{R}_{n; r/N, r_0/N}(1, 1) - R_{r/N, r_0/N}(1, 1) \right\} \end{aligned}$$

for $r_0 \in \{\Delta, \dots, N - \Delta\}$. By the continuous mapping theorem, Corollary 2 implies that $V_n^{(N)} \rightsquigarrow V^{(N)}$ and thus $D_n^{(N)} \rightsquigarrow D^{(N)}$ as $n \rightarrow \infty$, with $V^{(N)}$ and $D^{(N)}$ as in Eq. (19) and (18), respectively. Under the alternative hypothesis, if $I^{(N)}(r_0/N)$ is not constant in r_0 , we have

$$\frac{D_n^{(N)}}{\sqrt{k}} \rightsquigarrow \max_{r_0=\Delta, \dots, N-\Delta} I^{(N)}(r_0/N) - \min_{r_0=\Delta, \dots, N-\Delta} I^{(N)}(r_0/N) > 0,$$

and thus $D_n^{(N)} \rightsquigarrow \infty$ as $n \rightarrow \infty$.

B.2. The limit distribution of the test statistic ($N \rightarrow \infty$). Suppose that $N = N_n \rightarrow \infty$ and $\Delta = \Delta_n \rightarrow \infty$ in such a way that $\delta_n = \Delta_n/N_n \rightarrow \delta \in (0, 1/2)$. For simplicity, assume that $\delta_n = \delta$ for all sufficiently large n . To handle the more general case where $\delta_n \neq \delta$ but $\delta_n \rightarrow \delta$, proceed by an extension of the argument below, considering the asymptotic distributions of $\hat{I}$ below jointly in δ' in a neighborhood around δ . Introduce

$$\hat{I}(t_0) = \frac{1}{2\delta} \int_{t_0-\delta}^{t_0+\delta} \hat{R}_{n;t,t_0}(1,1) dt$$

together with

$$D_n = \sqrt{k} \left[\sup_{t \in [\delta, 1-\delta]} \hat{I}(t_0) - \inf_{t \in [\delta, 1-\delta]} \hat{I}(t_0) \right].$$

Also write

$$I(t_0) = \frac{1}{2\delta} \int_{t_0-\delta}^{t_0+\delta} R_{t,t_0}(1,1) dt.$$

Since

$$(43) \quad V_n(t_0) := \sqrt{k} \{ \hat{I}(t_0) - I(t_0) \} = \frac{1}{2\delta} \int_{t_0-\delta}^{t_0+\delta} \sqrt{k} \{ \hat{R}_{n;t,t_0}(1,1) - R_{t,t_0}(1,1) \} dt,$$

the continuous mapping theorem implies that, in the space $\ell^\infty([\delta, 1-\delta])$, we have weak convergence

$$V_n = \sqrt{k}(\hat{I} - I) \rightsquigarrow V, \quad n \rightarrow \infty$$

where V is defined by

$$(44) \quad V(t_0) = \frac{1}{2\delta} \int_{t_0-\delta}^{t_0+\delta} \hat{W}(t, t_0; 1, 1) dt, \quad t_0 \in [\delta, 1-\delta].$$

Under the null hypothesis of stationarity, $I(t_0)$ does not depend on $t_0 \in [\delta, 1-\delta]$, and thus

$$\begin{aligned} D_n &= \sup_{t \in [\delta, 1-\delta]} V_n(t_0) - \inf_{t \in [\delta, 1-\delta]} V_n(t_0) \\ &\rightsquigarrow \sup_{t \in [\delta, 1-\delta]} V(t_0) - \inf_{t \in [\delta, 1-\delta]} V(t_0) = D, \quad n \rightarrow \infty. \end{aligned}$$

To show weak convergence $D_n^{(N)} \rightsquigarrow D$ as $n \rightarrow \infty$ as well, it is then sufficient to show that $D_n^{(N)} = D_n + o_p(1)$ as $n \rightarrow \infty$. By the assumption that $R_{s,t}(1,1)$ is continuous in (s, t) , the empirical tail copula process is asymptotically uniformly equicontinuous in probability; this follows from Theorem 2 and Lemma 12. The same then holds for the process V_n , whence

$$D_n = \max_{r_0=\Delta, \dots, N-\Delta} V_n(r_0/N) - \min_{r_0=\Delta, \dots, N-\Delta} V_n(r_0/N) + o_p(1), \quad n \rightarrow \infty.$$

Further, the sum over r in $V_n^{(N)}$ in (42) can be written as a Riemann approximation to the integral over t in V_n in Eq. (43) by means of intervals of length $1/N$. Again, by asymptotic uniform equicontinuity in probability of the empirical tail copula process, we find that

$$\max_{r_0=\Delta, \dots, N-\Delta} \left| V_n^{(N)}(r_0/N) - V_n(r_0/N) \right| = o_p(1), \quad n \rightarrow \infty.$$

The relation $D_n^{(N)} = D_n + o_p(1)$ as $n \rightarrow \infty$ follows.

B.3. Calculating the covariance matrix of $V^{(N)}$ for the Smith model and the Pareto process. The Gaussian random vector $V^{(N)}$ in Eq. (19) is a linear combination of the random variables $\hat{W}(r/N, r_0/N; 1, 1)$ for $r_0 \in \{\Delta, N - \Delta\}$ and $r \in \{0, 1, \dots, N\}$ such that $1 \leq |r - r_0| \leq \Delta$. The covariance matrix of $V^{(N)}$ can thus be calculated from the one of $\hat{W}(s, t; 1, 1)$ for $s, t \in \{0, 1/N, \dots, N\}$. In turn, the Gaussian process $\hat{W}$ in Theorem 2 is the result of a linear transformation applied to the Gaussian process W with covariance function in Eq. (5). For brevity of notation, we will temporarily omit the arguments $x_j = 1$ and write $W(s, t; 1, 1) \equiv W(s, t)$ and $\dot{R}_{s,t;j}(1, 1) \equiv \dot{R}_{s,t;j}$ and so on. It follows that

$$\begin{aligned} & \mathbb{E} \left[\hat{W}(s, t) \hat{W}(s', t') \right] \\ &= \mathbb{E}[W(s, t)W(s', t')] - \dot{R}_{s',t';1} \mathbb{E}[W(s, t)W(s')] - \dot{R}_{s',t';2} \mathbb{E}[W(s, t)W(t')] \\ & \quad - \dot{R}_{s,t;1} \mathbb{E}[W(s)W(s', t')] + \dot{R}_{s,t;1} \dot{R}_{s',t';1} \mathbb{E}[W(s)W(s')] + \dot{R}_{s,t;1} \dot{R}_{s',t';2} \mathbb{E}[W(s)W(t')] \\ & \quad - \dot{R}_{s,t;2} \mathbb{E}[W(t)W(s', t')] + \dot{R}_{s,t;2} \dot{R}_{s',t';1} \mathbb{E}[W(t)W(s')] + \dot{R}_{s,t;2} \dot{R}_{s',t';2} \mathbb{E}[W(t)W(t')] \\ &= R_{s,t,s',t'} - \dot{R}_{s',t';1} R_{s,t,s'} - \dot{R}_{s',t';2} R_{s,t,t'} \\ & \quad - \dot{R}_{s,t;1} R_{s,s',t'} + \dot{R}_{s,t;1} \dot{R}_{s',t';1} R_{s,s'} + \dot{R}_{s,t;1} \dot{R}_{s',t';2} R_{s,t'} \\ & \quad - \dot{R}_{s,t;2} R_{t,s',t'} + \dot{R}_{s,t;2} \dot{R}_{s',t';1} R_{t,s'} + \dot{R}_{s,t;2} \dot{R}_{s',t';2} R_{t,t'}. \end{aligned}$$

We calculate the quantities $\dot{R}_{s,t;j}(1, 1)$ and $R_t(1)$ for vectors $\mathbf{t} \in [0, 1]^d$ with d distinct elements for the Smith model in Example 1 with $r = 1$ and $\Sigma = 1$ and for the Pareto process in Example 2. In both cases, the bivariate tail copulas are of Hüsler–Reiss form, so that

$$\dot{R}_{s,t;j}(1, 1) = \begin{cases} \bar{\Phi}(|s - t|/2), & \text{Smith model,} \\ \bar{\Phi}(\sqrt{|s - t|}/2), & \text{Pareto process.} \end{cases}$$

Next we calculate the tail dependence coefficients $R_t(1)$. For the Smith model, we have, for a standard normal variable Z ,

$$R_t(1) = \mathbb{E} \left[\min_{j=1, \dots, d} \exp(t_j Z - t_j^2/2) \right] = 2\bar{\Phi}(\tfrac{1}{2}(t_{(d)} - t_{(1)}))$$

where $t_{(1)}$ and $t_{(d)}$ are the minimum and the maximum of $\mathbf{t}$, respectively. To show the second equality above, split the expectation according to the index j where the minimum is attained and use the identity $\mathbb{E}[\exp(uZ - u^2/2)f(Z)] = \mathbb{E}[f(Z + u)]$ for scalar u and measurable nonnegative function f .

For the Pareto process, we have

$$R_t(1) = \mathbb{E} \left[\min_{j=1, \dots, d} \exp\{W'(t_j) - t_j^2/2\} \right]$$

where W' is a standard Wiener process. The expectation can be calculated by means of the following lemma, using that $\text{cov}\{W'(s), W'(t)\} = \min(s, t)$ for $s, t \in [0, 1]$. If some t_j equals 0, then the covariance matrix of $(W'(t_j))_j$ is not positive definite, but the formula still holds, as can be seen for instance by replacing $\mathbf{t}$ by $\mathbf{t} + c$ for some $c > 0$ and exploiting the fact that W' has stationary and independent increments.

LEMMA 13. *For a d -dimensional ($d \geq 2$) centered Gaussian random vector X with positive definite covariance matrix $\Gamma = (\gamma_{jk})_{j,k=1}^d$, we have*

$$\mathbb{E} \left[\min_{j=1, \dots, d} \exp(X_j - \tfrac{1}{2}\gamma_{jj}) \right] = \sum_{j=1}^d \Phi_{d-1} \left((-\tfrac{1}{2} \text{var}(\Delta_k^{(j)}))_{k:k \neq j}; \text{cov}(\Delta^{(j)}) \right)$$

where $\Phi_r(\cdot; \Sigma)$ is the cdf of the r -variate centered normal distribution with covariance matrix Σ and where $\Delta^{(j)} = (\Delta_k^{(j)})_{k:k \neq j}$ with $\Delta_k^{(j)} = X_j - X_k$.

PROOF. For every vector $a \in \mathbb{R}^d$ and for every nonnegative measurable function f , we have $\mathbb{E}[\exp(a^\top X - \frac{1}{2}a^\top \Gamma a)f(X)] = \mathbb{E}[f(X + \Gamma a)]$. It follows that

$$\begin{aligned} & \mathbb{E} \left[\min_{j=1, \dots, d} \exp \left(X_j - \frac{1}{2} \gamma_{jj} \right) \right] \\ &= \sum_{j=1}^d \mathbb{E} \left[\exp \left(X_j - \frac{1}{2} \gamma_{jj} \right) \mathbb{1} \left\{ X_j - \frac{1}{2} \gamma_{jj} \leq \min_{k:k \neq j} (X_k - \frac{1}{2} \gamma_{kk}) \right\} \right] \\ &= \sum_{j=1}^d \mathbb{P} \left[X_j + \frac{1}{2} \gamma_{jj} \leq \min_{k:k \neq j} \{ X_k + \gamma_{kj} - \frac{1}{2} \gamma_{kk} \} \right] \end{aligned}$$

which can be rearranged into the stated expression. $\square$

Note that the expression in Lemma 13 is akin to the one of the stable tail dependence function or exponent function of the Hüsler–Reiss copula in Remark 2.5 in Nikoloulopoulos, Joe and Li (2009).

B.4. Nonparametric estimation the covariance matrix of $V^{(N)}$. To estimate the covariance matrix of $V^{(N)}$ in Eq. (19) we start as Section B.3 above and express $V^{(N)}$ as a linear transformation of the centered Gaussian process $\hat{W}$, which is in turn a linear transformation of the centered Gaussian process W . The covariance function of W in Eq. (5) is given entirely in terms of the tail copula, which we estimate by the empirical tail copula.

The coefficients appearing in the linear transformation are all known except for the partial derivatives $\dot{R}_{s,t;j}(1, 1)$. These can be estimated by a finite differencing via

$$\frac{1}{2\eta} \left\{ \hat{R}_{n;s,t}(1 + \eta, 1) - \hat{R}_{n;s,t}(1 - \eta, 1) \right\}$$

for $j = 1$ and similarly for $j = 2$; here $\eta = \eta_n > 0$ is such that $\eta > 0$ but $k^{1/2}\eta \rightarrow \infty$ as $n \rightarrow \infty$. Under the null hypothesis of stationarity, the partial derivative only depends on (s, t) via $s - t$. This information can be exploited by averaging the estimator over all pairs (s, t) with the same value of $s - t$. Also, since $0 \leq \dot{R} \leq 1$, we can enforce the same constraints on the estimate.

The estimate of the covariance matrix of a finite-dimensional distribution of W based on the empirical tail copula is automatically symmetric and positive semi-definite. This can be seen by writing the empirical tail copula as a matrix of second-order moments of indicator variables. The symmetric and positive semi-definite character is then inherited by the estimate of the covariance matrix of $V^{(N)}$ by using the formula that $\text{cov}(AX) = A \text{cov}(X) A'$ for a random vector X and a real matrix A with the appropriate number of columns.

Under the null hypothesis of stationarity, the covariance matrix of $V^{(N)}$ is Toeplitz, which is in general not the case for the estimate, although we found in the simulations that it was usually not far from being so, perhaps because of the regularisation of the estimate of $\dot{R}$ as explained above. To enforce the estimated covariance matrix to be Toeplitz, we averaged out over the subdiagonals. This operation never destroyed the positive definite property, although it is not guaranteed to do so. If needed, positive definiteness can be restored by adding a diagonal matrix with a small but sufficiently large diagonal depending on the smallest eigenvalue of the estimated matrix.

B.5. Calculating p -values. The p -value associated to the test statistic $D_n^{(N)}$ in Eq. (17) and computed from an estimate $\hat{\Sigma}_n^{(N)}$ of the covariance matrix of $V^{(N)}$ as described in Section B.4 is given by

$$1 - F(D_n^{(N)}; \hat{\Sigma}_n^{(N)}),$$

where for $q \in [0, \infty)$ and for a covariance matrix Γ of dimension $m \times m$ with $m \geq 2$, we put

$$F(q; \Gamma) = \mathbb{P}[\max(U) - \min(U) \leq q]$$

with U a centered multivariate normal random vector with covariance matrix Γ . The latter probability can be calculated as follows, provided Γ is positive definite: partitioning the event according to the index $i \in \{1, \dots, m\}$ at which the maximum is realized, we have

$$(45) \quad \begin{aligned} \mathbb{P}[\max(U) - \min(U) \leq q] &= \sum_{i=1}^m \mathbb{P}[\forall j \in \{1, \dots, m\} \setminus \{i\} : -q < U_j - U_i < 0] \\ &= \sum_{i=1}^m \Phi_{m-1}([-q, 0]^{m-1}; \text{cov}((U_j - U_i)_{j:j \neq i})), \end{aligned}$$

in terms of the r -variate centered Gaussian probability measure Φ_r with the given covariance matrix. Substituting $q = D_n^{(N)}$ and $\Gamma = \hat{\Sigma}_n^{(N)}$ then yields the p -value appearing in the middle row in Figure 1 in the paper. It is also on the basis of these p -values that the hypothesis test is implemented of which the power is shown on the bottom row of the same figure.

In the numerical experiments, it was the computation of the p -value that was the most time-consuming, especially if N is large: for $N = 20$ and $\Delta = 2$, the dimension of $V^{(N)}$ is $N - 2\Delta + 1 = 17 = m$, so that computing a single p -value requires $m = 17$ evaluations of the multivariate Gaussian probability measure of dimension $m - 1 = 16$. We used the function `sadmvn` of the R-package `mnormt` to perform this task, which is in turn based on an algorithm described in Genz (1992). The calculation of a single p -value then took about 1 second on a standard laptop, which is not that much, but of course quickly adds up if to be repeated over many samples.

B.6. Calculating the probability density function of the limit $D^{(N)}$. The probability density functions at the top row of Figure 1 are, upon rescaling by $2\Delta\sqrt{k}$, the ones of the limit variable $D^{(N)} = \max(V^{(N)}) - \min(V^{(N)})$ for the Smith model and the Pareto process. The cumulative distribution function of $D^{(N)}$ is given in Eq. (45), where for Γ we substitute the covariance matrix of $V^{(N)}$ for the given model, as calculated in Subsection B.3. From the cdf we then proceed to calculate the pdf by numeric differentiation (finite differencing). An alternative, which we did not try, would be to analytically differentiate Eq. (45) with respect to q and calculate the resulting expression numerically.

Acknowledgements. We are very grateful to an Associate Editor and three Referees for various insightful comments that led to this improved version of the manuscript.

John Einmahl holds the Arie Kapteyn Chair 2019–2022 and gratefully acknowledges the corresponding research support.

REFERENCES

- AZZALINI, A. and GENZ, A. (2020). The R package `mnormt`: The multivariate normal and t distributions (version 2.0.2).
 BÜCHER, A. and DETTE, H. (2013). Multiplier bootstrap of tail copulas with applications. *Bernoulli* **19** 1655–1687.

- CHIAPINO, M., SABOURIN, A. and SEGERS, J. (2019). Identifying groups of variables with the potential of being large simultaneously. *Extremes* **22** 193–222.
- COLES, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **55** 797–816.
- DE HAAN, L. and LIN, T. (2001). On convergence towards an extreme value distribution in $C[0, 1]$. *The Annals of Probability* **29** 467–483.
- DE HAAN, L. and LIN, T. (2003). Weak consistency of extreme value estimators in $C[0, 1]$. *The Annals of Statistics* **31** 1996–2012.
- DOMBRY, C. (2017). Personal communication.
- DOMBRY, C. and RIBATET, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface* **8** 9–17.
- DOMBRY, C., RIBATET, M. and STOEV, S. (2018). Probabilities of concurrent extremes. *Journal of the American Statistical Association* **113** 1565–1582.
- DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis* **64** 25–46.
- EINMAHL, J. H. J., GANTNER, M. and SAWITZKI, G. (2010). Asymptotics for the shorth plot. *Journal of Statistical Planning and Inference* **140** 3003–3012.
- EINMAHL, J. H. J., KIRILIOUK, A. and SEGERS, J. (2018). A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes* **21** 205–233.
- EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* **40** 1764–1793.
- EINMAHL, J. H. J. and LIN, T. (2006). Asymptotic normality of extreme value estimators on $C[0, 1]$. *The Annals of Statistics* **34** 469–492.
- FALK, M. (2019). *Multivariate Extreme Value Theory and D-Norms*. Springer, Cham.
- FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737.
- GENZ, A. (1992). Numerical Computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1** 141–149.
- GINÉ, E., HAHN, M. and VATAN, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probability Theory and Related Fields* **87** 139–165.
- GOMES, M. I., DE HAAN, L. and PESTANA, D. (2004). Joint exceedances of the ARCH process. *Journal of Applied Probability* **41** 919–926.
- GUDENDORF, G. and SEGERS, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference* **142** 3073–3085.
- HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters* **7** 283–286.
- KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability* **37** 2042–2065.
- KOCH, E. (2017). Spatial risk measures and applications to max-stable processes. *Extremes* **20** 635–670.
- MERCADIER, C. and ROUSTANT, O. (2019). The tail dependograph. *Extremes* **22** 343–372.
- NIKOULOPOULOS, A. K., JOE, H. and LI, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12** 129–148.
- PENG, L. and QI, Y. (2008). Bootstrap approximation of tail dependence function. *Journal of Multivariate Analysis* **99** 1807–1824.
- RESSEL, P. (2013). Homogeneous distributions—And a spectral representation of classical mean values and stable tail dependence functions. *Journal of Multivariate Analysis* **117** 246–256.
- RIBATET, M. (2020). *SpatialExtremes: Modelling Spatial Extremes* R package version 2.0-9.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.
- RUYMGAART, F. H. (1973). *Asymptotic Theory of Rank Tests for Independence*. Mathematisch Centrum, Amsterdam Mathematical Centre Tracts, 43.
- SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44.
- SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics* **33** 307–355.
- SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782.
- SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.
- R CORE TEAM (2018). *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer Sciences+Business Media, New York.