SUPPLEMENT TO "RANK-BASED ESTIMATION UNDER ASYMPTOTIC DEPENDENCE AND INDEPENDENCE, WITH APPLICATIONS TO SPATIAL EXTREMES"

BY MICHAËL LALANCETTE [1,*], SEBASTIAN ENGELKE [2] AND STANISLAV VOLGUSHEV [1,†]

[1]*Department of Statistical Sciences, University of Toronto,* [*]*lalancette@utstat.toronto.edu;* [†]*stanislav.volgushev@utoronto.ca*

[2]*Research Center for Statistics, University of Geneva,* *sebastian.engelke@unige.ch*

This Supplementary Material to Lalancette, Engelke and Volgushev (2020+) is divided in six sections. Section S1 contains the proofs of all main results, with a number of necessary technical results deferred to Section S2. Sections S3 and S4 present proofs of several claims from different Examples. A brief discussion of computational complexity in spatial estimation is given in Section S5 and additional simulation results appear in Section S6. All references to sections, results and equations that do not start with the letter "S" are pointing to the aforementioned main paper.

S1. Proofs of main results. In this section are collected the proofs of Theorems 1 to 5. A number of more technical results, which are instrumental in the following, are collected in Section S2.

S1.1. *Bivariate estimation.* For the proofs concerning the bivariate estimators, we assume the framework of Sections 3.1 and 3.2, we define the transformed random variables $U = 1 - F_1(X)$, $V = 1 - F_2(Y)$ and note that Q is the distribution function of the random vector (U, V). Define the transformed observations $U_i = 1 - F_1(X_i)$, $V_i = 1 - F_2(Y_i)$ and denote by $U_{n,1}, \dots, U_{n,n}$ and $V_{n,1}, \dots, V_{n,n}$ the ordered versions thereof. Additionally define $U_{n,0} = V_{n,0} = 0$. For an intermediate sequence k, define the random functions u_n and v_n by

$$u_n(x) = \frac{n}{k} U_{n,\lfloor kx \rfloor} \quad \text{and} \quad v_n(y) = \frac{n}{k} V_{n,\lfloor ky \rfloor},$$

for $(x, y) \in [0, T]^2$. Recalling that $m = nq(k/n)$, it allows us to write

$$\widehat{c}_n(x, y) = \frac{n}{m} Q_n\left(\frac{k}{n} u_n(x), \frac{k}{n} v_n(y)\right)$$

where

$$Q_n(x, y) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{U_i \le x, V_i \le y\right\}$$

denotes the empirical distribution function of $(U_1, V_1), \dots, (U_n, V_n)$. We begin by discussing technical results that will be used in the proof of both Theorem 1 and Theorem 2. Consider the decomposition

$$\begin{aligned} W_n(x, y) =& \sqrt{m}\Big(\frac{n}{m} Q_n\Big(\frac{k}{n} u_n(x), \frac{k}{n} v_n(y)\Big) - \frac{n}{m} Q\Big(\frac{k}{n} u_n(x), \frac{k}{n} v_n(y)\Big)\Big) \\ &+ \sqrt{m}\Big(\frac{n}{m} Q\Big(\frac{k}{n} u_n(x), \frac{k}{n} v_n(y)\Big) - c(u_n(x), v_n(y))\Big) \\ &+ \sqrt{m}\big(c(u_n(x), v_n(y)) - c(x, y)\big). \end{aligned}$$

For the second term in the above decomposition, note that

$$\sqrt{m}\Big(\frac{n}{m}Q\Big(\frac{k}{n}x,\frac{k}{n}y\Big)-c(x,y)\Big)=O\left(\sqrt{m}q_1\left(\frac{k}{n}\right)\right)=o(1)$$

uniformly over all $(x,y)\in[0,2T]^2$; here the last equation follows from Condition 1(ii). By Corollary S1 we have $\mathbb{P}(u_n(T)\vee v_n(T)\le 2T)\to 1$, and thus

$$\sup_{x,y\in[0,T]}\sqrt{m}\Big|\frac{n}{m}Q\Big(\frac{k}{n}u_n(x),\frac{k}{n}v_n(y)\Big)-c(u_n(x),v_n(y))\Big|=o_P(1).$$

Next define for all $x,y\in[0,2T]$

$$H_n(x,y):=\sqrt{m}\Big(\frac{n}{m}Q_n\Big(\frac{k}{n}x,\frac{k}{n}y\Big)-\frac{n}{m}Q\Big(\frac{k}{n}x,\frac{k}{n}y\Big)\Big). \tag{S1.1}$$

By Lemma S4 this process converges, in $\ell^\infty([0,2T]^2)$, to the process W from Theorem 1 and by Corollary S1 u_n and v_n converge uniformly in probability to the identity function $I:[0,2T]\to[0,2T]$. Therefore, the triple (H_n,u_n,v_n) converges jointly in distribution to (W,I,I). This implies

$$\sup_{x,y\in[0,T]}\Big|H_n(u_n(x),v_n(y))-H_n(x,y)\Big|=o_P(1). \tag{S1.2}$$

Indeed, consider the map

$$f:\begin{cases}\ell^\infty([0,2T]^2)\times\mathcal{V}[0,T]\times\mathcal{V}[0,T]\to\mathbb{R}\\ (a,b_1,b_2)\mapsto\sup_{x,y\in[0,T]}|a(b_1(x),b_2(y))-a(x,y)|\end{cases}$$

where $\mathcal{V}[0,T]:=\{g\in\ell^\infty[0,T]:g([0,T])\subset[0,2T]\}$ and assume that the product space is equipped with the norm $\|a\|_\infty+\|b_1\|_\infty+\|b_2\|_\infty$. Observe that f is continuous at points (a,b_1,b_2) where a is a continuous function and that the sample paths of W are almost surely continuous. Thus, by the continuous mapping theorem, with probability converging to 1,

$$\sup_{x,y\in[0,T]}\Big|H_n(u_n(x),v_n(y))-H_n(x,y)\Big|=f(H_n,u_n,v_n)\rightsquigarrow f(W,I,I)=0.$$

Since the limit is constant a.s. Equation (S1.2) follows. Combining the equations above, we find

$$W_n(x,y)=H_n(x,y)+\sqrt{m}\big(c(u_n(x),v_n(y))-c(x,y)\big)+o_P(1), \tag{S1.3}$$

where the term $o_P(1)$ is uniform on $[0,T]^2$, and we recall that $H_n\rightsquigarrow W$ in $\ell^\infty([0,2T]^2)$.

S1.1.1. *Proof of Theorem 1.* Define

$$S_n(x,y):=\sqrt{m}\big(c(u_n(x),v_n(y))+c(x,y)\big).$$

In light of Equation (S1.3) it suffices to prove that $S_n\xrightarrow{P}0$ uniformly on $[0,T]^2$. From here on it is more convenient to study component-wise increments. That is, we write

$$\begin{aligned}S_n(x,y)&=\sqrt{m}(c(u_n(x),y)-c(x,y))+\sqrt{m}(c(u_n(x),v_n(y))-c(u_n(x),y))\\&=:S_n^{(a)}(x,y)+S_n^{(b)}(x,y)\end{aligned}$$

and we will show that both $S_n^{(a)}$ and $S_n^{(b)}$ converge to 0 in probability, starting with $S_n^{(a)}$.

By assumption, since with probability converging to 1 we have $u_n(x) \in [0, 2T]$ for every $x \le T$, we can write

$$\text{(S1.4)} \quad S_n^{(a)}(x,y) = \sqrt{m}\,(c(u_n(x),y) + c(x,y))$$
$$= \sqrt{m}\left\{\frac{n}{m}Q\left(\frac{k}{n}u_n(x), \frac{k}{n}y\right) - \frac{n}{m}Q\left(\frac{k}{n}x, \frac{k}{n}y\right) + O_P\left(q_1\left(\frac{k}{n}\right)\right)\right\}$$
$$\text{(S1.5)} \quad = \frac{n}{\sqrt{m}}\left(Q\left(\frac{k}{n}u_n(x), \frac{k}{n}y\right) - Q\left(\frac{k}{n}x, \frac{k}{n}y\right)\right) + o_P(1)$$

uniformly on $[0,T]^2$, since the sequence m was chosen so that $\sqrt{m}q_1(k/n) \to 0$. We will use both Equations (S1.4) and (S1.5) as representations of $S_n^{(a)}$ throughout the proof.

Let $\beta_n = (m/k)/(\log(k/m))$. From there, partition $[0,T]^2$ in $\Theta_n^{(1)} = [0, 1/k) \times [0,T]$, $\Theta_n^{(2)} = [1/k, \beta_n) \times [0,T]$ and $\Theta_n^{(3)} = [\beta_n, T] \times [0,T]$ (if $\beta_n < 1/k$, $\Theta_n^{(2)}$ is empty). These sets represent the "small", "intermediate" and "large" values of x, respectively. We will prove that the suprema of $S_n^{(a)}$ on $\Theta_n^{(1)}$, $\Theta_n^{(2)}$ and $\Theta_n^{(3)}$ all converge to 0 in probability. Equation (S1.5) yields

$$\begin{aligned}
\sup_{(x,y)\in\Theta_n^{(1)}} |S_n^{(a)}(x,y)| &= \frac{n}{\sqrt{m}} \sup_{0\le x<1/k} \left| Q\left(\frac{k}{n}u_n(x), \frac{k}{n}y\right) - Q\left(\frac{k}{n}x, \frac{k}{n}y\right)\right| + o_P(1) \\
&= \frac{n}{\sqrt{m}} \sup_{0\le x<1/k} Q\left(\frac{k}{n}x, \frac{k}{n}y\right) + o_P(1) \\
&\le \frac{n}{\sqrt{m}}\frac{1}{n} + o_P(1) \\
&= \frac{1}{\sqrt{m}} + o_P(1),
\end{aligned}$$

where we have once again used the facts that $u_n(x) = 0$ whenever $x < 1/k$ and that $Q(0,\cdot) = Q(\cdot,0) = 0$, in addition to the fact that $Q(u,v) \le u$. This proves that $\sup_{\Theta_n^{(1)}} |S_n^{(a)}| \to 0$ in probability.

Using Equation (S1.5) again, the supremum of $S_n^{(a)}$ on $\Theta_n^{(2)}$ can be expressed as

$$\begin{aligned}
\sup_{1/k\le x<\beta_n} \left|S_n^{(a)}(x,y)\right| &= \sup_{1/k\le x<\beta_n} \frac{n}{\sqrt{m}} \left| Q\left(\frac{k}{n}u_n(x), \frac{k}{n}y\right) - Q\left(\frac{k}{n}x, \frac{k}{n}y\right)\right| + o_P(1) \\
&\le \sup_{1/k\le x<\beta_n} \frac{n}{\sqrt{m}} \left|\frac{k}{n}u_n(x) - \frac{k}{n}x\right| + o_P(1) \\
&= \sup_{1/k\le x<\beta_n} \frac{k}{\sqrt{m}} |u_n(x) - x| + o_P(1) \\
&= O_P\left(\sup_{1/k\le x<\beta_n} \sqrt{\frac{k}{m}}\varphi(x)\right) + o_P(1),
\end{aligned}$$

where we have used Lipschitz continuity of Q and Lemma S3. The last bound holds for any function φ that satisfies the conditions in Lemma S3, but from now on we use $\varphi(x) := \sqrt{x \log\log(1/x)}$ on $(0, B]$ and $\varphi(x) := \sqrt{x}$ on $(B, T]$, where $B > 0$ is chosen small enough so that φ is well defined and non-decreasing. By monotonicity, the supremum is attained at

$x = \beta_n$. We then have

$$\sup_{1/k \le x < \beta_n} \left| S_n^{(a)}(x,y) \right| = O_P\left(\sqrt{\frac{k}{m}\beta_n \log\log(1/\beta_n)} \right) + o_P(1)$$

because since $\beta_n \to 0$, eventually $\beta_n \le B$, so eventually $\varphi(\beta_n) = \sqrt{\beta_n \log\log(1/\beta_n)}$. The last display converges in probability to 0 since

$$\frac{k}{m}\beta_n \log\log(1/\beta_n) = \frac{\log\log\left(\frac{k}{m}\log(k/m)\right)}{\log(k/m)} \longrightarrow 0$$

as $k/m \to \infty$, which proves that $\sup_{\Theta_n^{(2)}} |S_n^{(a)}| \to 0$ in probability.

Finally, when considering large values of x, Lemma S3 and a combination of Lemmas S7 and S8 imply that

$$\begin{aligned} \sup_{\beta_n \le x \le T} \left| S_n^{(a)}(x,y) \right| &= \sup_{\beta_n \le x \le T} \sqrt{m} |c(u_n(x), y) - c(x,y)| \\ &\lesssim \sqrt{m} \sup_{\beta_n \le x \le T} |u_n(x) - x| r(x \vee u_n(x)) \\ &= O_P\left(\sqrt{\frac{m}{k}} \sup_{\beta_n \le x \le T} \varphi(x) r(x \vee u_n(x)) \right), \end{aligned}$$

where $r(x) = (x\log(1/x))^{-1}$. By monotonicity of φ, the inside of the O_P can be upper bounded by

$$\sqrt{\frac{m}{k}} \sup_{\beta_n \le x \le T} \varphi(x \vee u_n(x)) r(x \vee u_n(x))$$

and since with probability converging to 1, for every $x \le T$, $u_n(x) \le 2T$, this can in turn be upper bounded (with probability converging to 1) by

$$\sqrt{\frac{m}{k}} \sup_{\beta_n \le x \le 2T} \varphi(x) r(x).$$

It can easily be checked (e.g. by differentiation) that the function $\varphi \times r$ is decreasing. Thus, the above supremum is attained at β_n. Finally, elementary computations yield

$$\sqrt{\frac{m}{k}} \varphi(\beta_n) r(\beta_n) \lesssim \sqrt{\frac{\log\log((k/m)^2)}{\log(k/m)}} \longrightarrow 0.$$

Overall, we have shown that $S_n^{(a)} \xrightarrow{P} 0$ uniformly over $[0,T]^2$. Note that all the bounds we derived are uniform over all values of $y \in [0,T]$, although it was removed from the notation for parsimony. In order to deal with $S_n^{(b)}$, we recall once again that with probability converging to 1, we have $u_n(x) \le 2T$ for every $x \le T$. Therefore, with probability converging to 1,

$$\begin{aligned} \sup_{(x,y)\in[0,T]^2} \left| S_n^{(b)}(x,y) \right| &= \sup_{(x,y)\in[0,T]^2} \sqrt{m} |c(u_n(x), v_n(y)) - c(u_n(x), y)| \\ &\le \sup_{x\in[0,2T], y\in[0,T]} \sqrt{m} |c(x, v_n(y)) - c(x,y)|. \end{aligned}$$

This can be shown to converge in probability to 0 using the exact same proof as for $S_n^{(a)}$. We finally conclude that $S_n \xrightarrow{P} 0$ in $\ell^\infty([0,T]^2)$, and the proof for deterministic $k = k_n$ is complete. It remains to show that the result continues to hold if we replace the deterministic sequence $k = k_n$ by data-dependent $\widehat{k}$ as outlined in Remark 1. This is established in Section S1.1.3. □

S1.1.2. *Proof of Theorem 2.* In view of Equation (S1.3), we require the joint asymptotic behavior of H_n, u_n and v_n. Define, for $(x, y) \in [0, \infty)^2$,

$$L_n^{(1)}(x) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{U_i \le \frac{k}{n} x\right\} \quad \text{and} \quad L_n^{(2)}(y) = \frac{1}{k} \sum_{i=1}^{n} \mathbb{1}\left\{V_i \le \frac{k}{n} y\right\},$$

a rescaled version of the marginal empirical distribution functions of U and V. We now show that the $\mathbb{D}$-valued process

$$(x, y) \mapsto \left(H_n(x, y), \sqrt{m}\left(L_n^{(1)}(x) - x\right), \sqrt{m}\left(L_n^{(2)}(y) - y\right)\right) \tag{S1.6}$$

converges in distribution to the Gaussian process $(W, W^{(1)}, W^{(2)})$ defined in Section 4.1.1 with covariance matrix Λ from Equation (4.2), where $\mathbb{D} := \left(\ell^\infty([0, 2T]^2)\right)^3$.

Again, let I denote the identity map on $\mathbb{R}$. The three processes H_n, $\sqrt{m}(L_n^{(1)} - I)$ and $\sqrt{m}(L_n^{(2)} - I)$ are individually tight (see Lemma S4) and hence it suffices to prove convergence of the marginal distributions. This in turn follows from convergence of the covariance function, by the multivariate Lindeberg-Feller theorem (see van der Vaart (2000), Theorem 2.27); verification of the Lindeberg condition is similar to condition (B) in the proof of Lemma S4. The convergence of $\mathbb{E}\left[H_n(x, y)H_n(x', y')\right]$ to $c(x \wedge x', y \wedge y')$ is already shown in Lemma S4. Using similar arguments and recalling that $m/k \to \chi > 0$, one easily deals with the other covariance terms and concludes that the processes in Equation (S1.6) weakly converge to $(W, W^{(1)}, W^{(2)})$ in $\mathbb{D}$.

Note that the random functions u_n and v_n are the generalized inverses of $L_n^{(1)} + 1/k$ and $L_n^{(2)} + 1/k$, respectively. Because $\sqrt{m}/k \to 0$, the term $1/k$ is negligible. Upon applying Vervaat's lemma (Vervaat (1972)), which states that the generalized inverse mapping is Hadamard differentiable around the identity function, we deduce that the processes G_n, defined by

$$G_n(x, y) = (H_n(x, y), \sqrt{m}(u_n(x) - x), \sqrt{m}(v_n(y) - y)),$$

weakly converge to $(W, -W^{(1)}, -W^{(2)})$ in $\mathbb{D}$. For $t > 0$, define the sets

$$\mathcal{V}(t) := \{b \in \ell^\infty([0, 2T]) : \forall x \in [0, T], x + tb(x) \in [0, 2T]\}. \tag{S1.7}$$

Let $\mathbb{D}_n \subset \mathbb{D}$ be the subset of functions $a = (a^{(0)}, a^{(1)}, a^{(2)})$ such that $a^{(1)}(x, y)$ is constant in y, $a^{(2)}(x, y)$ is constant in x and the functions $x \mapsto a^{(1)}(x, y)$ and $y \mapsto a^{(2)}(x, y)$ are elements of $\mathcal{V}(1/\sqrt{m})$. Let $\mathbb{E}$ be the space of equivalence classes $L^\infty([0, T]^2)$ equipped with the topology of hypi-convergence. Define the functionals $f_n : \mathbb{D}_n \to \mathbb{E}$ by

$$f_n(a)(x, y) := a^{(0)}(x, y) + \sqrt{m}\left(c\left(x + \frac{a^{(1)}(x, y)}{\sqrt{m}}, y + \frac{a^{(2)}(x, y)}{\sqrt{m}}\right) - c(x, y)\right).$$

Equation (S1.3) can be rephrased as $W_n = f_n(G_n) + o_P(1)$, assuming that $G_n \in \mathbb{D}_n$, which is true with probability

$$\mathbb{P}\left(u_n(T) \le 2T, v_n(T) \le 2T\right) \longrightarrow 1.$$

Let $\mathbb{D}_0 \subset \mathbb{D}$ be the subset of continuous functions a such that $a(0) = 0$. As soon as $a_n \in \mathbb{D}_n$ converges uniformly to $a \in \mathbb{D}_0$, by Lemma S9, $f_n(a_n)$ hypi-converges to $f(a)$, where $f : \mathbb{D}_0 \to \mathbb{E}$ satisfies

$$f(a) := a^{(0)} + \dot{c}_1 a^{(1)} + \dot{c}_2 a^{(2)}.$$

Note that $(W, -W^{(1)}, -W^{(2)})$ concentrates on $\mathbb{D}_0$. Therefore, by the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1),

$$W_n = f_n(G_n) + o_P(1) \rightsquigarrow f((W, -W^{(1)}, -W^{(2)})) = W - \dot{c}_1 W^{(1)} - \dot{c}_2 W^{(2)}$$

in $\mathbb{E}$. It remains to show that the result continues to hold if we replace the deterministic sequence $k = k_n$ by data-dependent $\widehat{k}$ as outlined in Remark 1. This is established in Section S1.1.3. □

S1.1.3. *Proof that Theorems 1 and 2 continue to hold with* $\widehat{k}$. Let $\widehat{c}_{n,\widehat{k}}$ be the estimator $\widehat{c}_n$ computed with the random quantity $\widehat{k}$ instead of k. We shall prove that $\sqrt{m}|\widehat{c}_{n,\widehat{k}} - \widehat{c}_n| \to 0$ in probability uniformly over $[0,T]^2$ (under asymptotic independence) or in the hypi semimetric (under asymptotic dependence).

Note that the definition of $\widehat{k}$ implies that $\widehat{c}_n(\widehat{k}/k, \widehat{k}/k) = 1$. By assumption, $\widehat{c}_n$ converges to c in probability uniformly in a neighborhood of $(1,1)$. Jointly with the fact that $c(x,x) = x^{1/\eta}$, this readily implies that $\widehat{k}/k \to 1$ in probability. Further note that

$$\widehat{c}_{n,\widehat{k}}(x,y) = \frac{q(k/n)}{q(\widehat{k}/n)} \widehat{c}_n(\widehat{k}x/k, \widehat{k}y/k).$$

We first discuss the case of asymptotic independence. By Theorem 1 and by Skorokhod's almost sure representation, we may assume that almost surely, $\widehat{c}_n = c + W/\sqrt{m} + o(1/\sqrt{m})$ and $\widehat{k}/k \to 1$. The object of interest is then equal, with probability one, to

$$\begin{aligned}
&\frac{q(k/n)}{q(\widehat{k}/n)}\sqrt{m}\left(\widehat{c}_n(\widehat{k}x/k, \widehat{k}y/k) - \frac{q(\widehat{k}/n)}{q(k/n)}\widehat{c}_n(x,y)\right) \\
&= \frac{q(k/n)}{q(\widehat{k}/n)}\Big\{\sqrt{m}\Big(c(\widehat{k}x/k, \widehat{k}y/k) - \frac{q(\widehat{k}/n)}{q(k/n)}c(x,y)\Big) + W(\widehat{k}x/k, \widehat{k}y/k) - W(x,y)\Big\} + o(1) \\
&= -\sqrt{m}c(x,y)\left(\frac{q(\widehat{k}/n)}{q(k/n)} - \left(\frac{\widehat{k}/n}{k/n}\right)^{1/\eta}\right)\frac{q(k/n)}{q(\widehat{k}/n)} + o(1),
\end{aligned} \tag{S1.8}$$

where we have used homogeneity of c, regular variation of q and the fact that almost surely, the sample paths of W are continuous, hence uniformly continuous on compact sets. The terms $o(1)$ are uniform over $[0,T]^2$. Finally, it is shown in Lemma S2 that uniformly over a in a neighborhood of 1, $q(at)/q(t) - a^{1/\eta} = O(q_1(t))$. Recalling that $\widehat{k}/k \to 1$ almost surely, the first term in Equation (S1.8) is then uniformly of the order of $\sqrt{m}q_1(k/n)$, which vanishes by Condition 1(ii).

In the case of asymptotic dependence, Theorem 2 ensures that $\widehat{c}_n = c + B/\sqrt{m} + o(1/\sqrt{m})$ in the hypi semimetric. We may apply the reasoning above except that, from the definition of the process B, we get the additional term

$$\begin{aligned}
&-\sum_{j=1}^{2}\Big(\dot{c}_j(\widehat{k}x/k, \widehat{k}y/k)W^{(j)}(\widehat{k}x/k, \widehat{k}y/k) - \dot{c}_j(x,y)W^{(j)}(x,y)\Big) \\
&\qquad = -\sum_{j=1}^{2}\dot{c}_j(x,y)\Big(W^{(j)}(\widehat{k}x/k, \widehat{k}y/k) - W^{(j)}(x,y)\Big);
\end{aligned} \tag{S1.9}$$

this follows from the fact that under asymptotic dependence, c is homogeneous of order 1 and the directional partial derivatives of such a function, when they exist, are constant along rays from the origin. The above term vanishes uniformly since $\dot{c}_j$ has to be locally bounded (only under asymptotic dependence) and since the sample paths of $W^{(j)}$ are almost surely continuous. We therefore obtain Equation (S1.8), except that this time the term $o(1)$ is understood in the hypi semimetric. From here on the proof is completed in the same way as under asymptotic independence. □

S1.1.4. *Proof of Theorem 3.* Recall the definition of Ψ_n from Section 3.2. Letting $\widehat{\sigma}_n = \frac{n}{m}\widehat{\zeta}_n$, the assumption that $(\widehat{\theta}_n, \widehat{\zeta}_n)$ minimizes the norm of Ψ_n^* becomes equivalent to $(\widehat{\theta}_n, \widehat{\sigma}_n)$ minimizing the norm of Ψ_n. The key is to note that for any θ, σ,

$$\Psi(\theta,\sigma) - \Psi_n(\theta,\sigma) = \int g(\widehat{c}_n - c)d\mu_{\mathrm{L}} = \frac{1}{\sqrt{m}}\int gW_n d\mu_{\mathrm{L}}, \tag{S1.10}$$

with W_n defined as in Theorems 1 and 2. By the dominated convergence theorem, and because g is integrable, one easily sees that the functional $f \mapsto \int gf d\mu_{\mathrm{L}}$ is continuous in $\ell^\infty([0,T]^2)$. By Lemma S10, this is also true in the topology of hypi-convergence on $\ell^\infty([0,T]^2)$ at points f that are continuous Lebesgue-almost everywhere on $[0,T]^2$. It is the case of both limiting Gaussian processes appearing in Theorems 1 and 2: W, $W^{(1)}$ and $W^{(2)}$ have almost surely continuous sample paths and under asymptotic dependence, the directional derivatives $\dot{c}_j$ are almost everywhere continuous. Those two results and the continuous mapping theorem then imply that

$$\int gW_n d\mu_{\mathrm{L}} \rightsquigarrow N(0,A).$$

We may therefore apply Lemma S11 with $\phi = \Psi$, $x_0 = (\theta_0, 1)$, $Y_n = \frac{1}{\sqrt{m}}\int gW_n d\mu_{\mathrm{L}}$ and $a_n = 1/\sqrt{m}$, and as required we obtain

$$\sqrt{m}((\widehat{\theta}_n, \widehat{\sigma}_n) - (\theta_0, 1)) = (J^\top J)^{-1}J^\top \int gW_n d\mu_{\mathrm{L}} + o_P(1) \rightsquigarrow N(0,\Sigma).$$

□

S1.2. *Spatial estimation.* For the proofs in the spatial setting, we assume the framework of Section 3.3, we define the transformed random variables $U^{(j)} = 1 - F^{(j)}(X^{(j)})$ and for a pair s, let $Q^{(s)}$ be the distribution function of the random vector $(U^{(s_1)}, U^{(s_2)})$. Define the transformed observations $U_i^{(j)} = 1 - F^{(j)}(X_i^{(j)})$ and denote by $U_{n,1}^{(j)}, \ldots, U_{n,n}^{(j)}$ the ordered versions thereof and define $U_{n,0}^{(j)} := 0$. For intermediate sequences $k^{(s)}$, we define the (weighted) empirical tail quantile functions $u_n^{(s,j)}$, $s \in \mathcal{P}, j \in \{1,2\}$, by

$$u_n^{(s,j)}(x) = \frac{n}{k^{(s)}} U_{n,\lfloor k^{(s)}x \rfloor}^{(s_j)}, \quad x \geq 0.$$

Recalling that $m^{(s)} = nq^{(s)}(k^{(s)}/n)$, it allows us to write

$$\widehat{c}_n^{(s)}(x,y) = \frac{n}{m^{(s)}} Q_n^{(s)}\left(\frac{k^{(s)}}{n}u_n^{(s,1)}(x), \frac{k^{(s)}}{n}u_n^{(s,2)}(y)\right).$$

where $Q_n^{(s)}$ denotes the empirical distribution function of $(U_1^{(s_1)}, U_1^{(s_2)}), \ldots, (U_n^{(s_1)}, U_n^{(s_2)})$. Following the discussion before the proof of Theorem 1, we may define

$$H_n^{(s)}(x,y) := \sqrt{m^{(s)}}\Big\{\frac{1}{m^{(s)}}\sum_{i=1}^{n}\mathbb{I}\Big\{U_i^{(s_1)} \le \frac{k^{(s)}}{n}x, U_i^{(s_2)} \le \frac{k^{(s)}}{n}y\Big\} - \frac{n}{m^{(s)}}\mathbb{P}\Big(U^{(s_1)} \le \frac{k^{(s)}}{n}x, U^{(s_2)} \le \frac{k^{(s)}}{n}y\Big)\Big\}.$$

and similarly obtain

(S1.11)

$$W_n^{(s)}(x,y) = H_n^{(s)}(x,y) + \sqrt{m^{(s)}}\left(c^{(s)}\left(u_n^{(s,1)}(x), u_n^{(s,2)}(y)\right) - c^{(s)}(x,y)\right) + o_P(1),$$

where $W_n^{(s)}$ is defined as in Theorem 4 and the term $o_P(1)$ is uniform over compact sets.

S1.2.1. *Proof of Theorem 4.* For asymptotically independent pairs, the second term of Equation (S1.11) vanishes uniformly, by the proof of Theorem 1. Define the $\mathbb{D}$-valued processes G_n by

$$G_n(x,y) := \left(\left(H_n^{(s)}(x,y)\right)_{s\in\mathcal{P}}, \left(\sqrt{m^{(s)}}\left(u_n^{(s,1)}(x) - x\right), \sqrt{m^{(s)}}\left(u_n^{(s,2)}(y) - y\right)\right)_{s\in\mathcal{P}_D}\right),$$

where $\mathbb{D} = \left(\ell^\infty([0,2T]^2)\right)^{|\mathcal{P}|+2|\mathcal{P}_D|}$. The proof now proceeds similarly to that of Theorem 2; we show that G_n converges in distribution, that the processes of interest $W_n^{(s)}$ can be approximately represented as a transformation of G_n, and we conclude by applying a continuous mapping theorem.

For $s \in \mathcal{P}$, $j \in \{1,2\}$, let

$$L_n^{(s,j)}(x) = \frac{1}{k^{(s)}}\sum_{i=1}^{n}\mathbb{1}\left\{U^{(s_j)} \le \frac{k^{(s)}}{n}x\right\}, \quad x \ge 0.$$

Recall that I denotes the identity mapping on $\mathbb{R}$. By standard arguments (see, e.g., the proofs of Theorems 1 and 2), we see that each of the processes $H_n^{(s)}$ and $\sqrt{m^{(s)}}\left(L_n^{(s,j)} - I\right)$ converge in distribution in $\ell^\infty([0,2T]^2)$, hence they are tight random elements in that space. It follows that the sequence of processes

(S1.12)

$$(x,y) \mapsto \left(\left(H_n^{(s)}(x,y)\right)_{s\in\mathcal{P}}, \left(\sqrt{m^{(s)}}\left(L_n^{(s,1)}(x) - x\right), \sqrt{m^{(s)}}\left(L_n^{(s,2)}(y) - y\right)\right)_{s\in\mathcal{P}_D}\right)$$

is tight in the product space $\mathbb{D}$. A Lindeberg-type condition (van der Vaart, 2000, Theorem 2.27) can easily be checked, so weak convergence of the process in Equation (S1.12) follows from convergence of $\mathbb{E}\left[G_n(x,y)G_n(x',y')^\top\right]$ to a suitable covariance matrix. This is simply a consequence of Condition 2; indeed, for suitable pairs $s, s' \in \mathcal{P}$, $j, j' \in \{1,2\}$ and $(x,y), (x',y') \in [0,\infty)^2$, this condition implies that

$$\lim_{n\to\infty}\mathbb{E}\left[H_n^{(s)}(x,y)H_n^{(s')}(x',y')\right] = \Gamma^{(s,s')}((x,y),(x',y')),$$

$$\lim_{n\to\infty}\mathbb{E}\left[H_n^{(s)}(x,y)\sqrt{m^{(s')}}\left(L_n^{(s',j)}(x') - x'\right)\right] = \Gamma^{(s,s',j)}((x,y),(x',y')),$$

$$\lim_{n\to\infty}\mathbb{E}\left[\sqrt{m^{(s)}}\left(L_n^{(s,j)}(x) - x\right)\sqrt{m^{(s')}}\left(L_n^{(s',j')}(x') - x'\right)\right] = \Gamma^{(s,j,s',j')}((x,y),(x',y')).$$

We deduce that in $\mathbb{D}$, the processes in Equation (S1.12) weakly converge to the Gaussian process

$$\left((W^{(s)})_{s\in\mathcal{P}}, (W^{(s,j)})_{s\in\mathcal{P}_D, j\in\{1,2\}}\right)$$

as defined in Section 4.2. Noting that $u_n^{(s,j)}$ is the generalized inverse function of $L_n^{(s,j)} + 1/k^{(s)}$ and that $\sqrt{m^{(s)}}/k^{(s)} \to 0$, we apply Vervaat's lemma (Vervaat, 1972) to obtain that

$$G_n \rightsquigarrow G := \left((W^{(s)})_{s\in\mathcal{P}}, (-W^{(s,j)})_{s\in\mathcal{P}_D, j\in\{1,2\}} \right) \tag{S1.13}$$

in $\mathbb{D}$.

Recall the definition of the sets $\mathcal{V}(t)$ in Equation (S1.7) and let $\mathbb{D}_n \subset \mathbb{D}$ be the subset of functions a of the form $\left((a^{(s)})_{s\in\mathcal{P}}, (a^{(s,j)})_{s\in\mathcal{P}_D, j\in\{1,2\}}\right)$ such that $a^{(s,1)}(x,y)$ is constant in y, $a^{(s,2)}(x,y)$ is constant in x and such that the functions $x \mapsto a^{(s,1)}(x,y)$ and $y \mapsto a^{(s,2)}(x,y)$ are elements of $\mathcal{V}\big(1/\sqrt{m^{(s)}}\big)$.

Defining $\mathbb{E}$ as the product space $\left(L^\infty([0,T]^2)\right)^{|\mathcal{P}|}$, with $L^\infty([0,T]^2)$ equipped with the topology of hypi-convergence, consider the following functionals $f_n : \mathbb{D}_n \to \mathbb{E}$. For an element $a = \left((a^{(s)})_{s\in\mathcal{P}}, (a^{(s,j)})_{s\in\mathcal{P}_D, j\in\{1,2\}}\right) \in \mathbb{D}_n$, $f_n(a) = (f_n(a)^{(s)})_{s\in\mathcal{P}}$ is a function such that $f_n(a)^{(s)} = a^{(s)}$ if $s \in \mathcal{P}_I$, and

$$f_n(a)^{(s)}(x,y) = a^{(s)}(x,y) + \sqrt{m^{(s)}} \left(c^{(s)}\left(x + \frac{a^{(s,1)}(x,y)}{\sqrt{m^{(s)}}}, y + \frac{a^{(s,2)}(x,y)}{\sqrt{m^{(s)}}} \right) - c^{(s)}(x,y) \right)$$

if $s \in \mathcal{P}_D$. Referring to Equation (S1.11) and recalling that the second term thereof vanishes if $s \in \mathcal{P}_I$, we notice that for every pair s, $W_n^{(s)} = f_n(G_n)^{(s)} + o_P(1)$. This representation, of course, holds only if $G_n \in \mathbb{D}_n$; this is satisfied with probability at least

$$\mathbb{P}\left(\forall s \in \mathcal{P}_D, j \in \{1,2\}, u_n^{(s,j)}(T) \le 2T \right) \longrightarrow 1$$

where the last convergence follows by Corollary S1 applied for each $s \in \mathcal{P}$. Define $f : \mathbb{D}_0 \to \mathbb{E}$, where $\mathbb{D}_0 \subset \mathbb{D}$ is the subset of continuous functions a such that $a(0) = 0$, as

$$f(a)^{(s)} = \begin{cases} a^{(s)}, & s \in \mathcal{P}_I \\ a^{(s)} + \dot{c}_1 a^{(s,1)} + \dot{c}_2 a^{(s,2)}, & s \in \mathcal{P}_D \end{cases}.$$

For a sequence $a_n \in \mathbb{D}_n$ that converges uniformly to a function $a \in \mathbb{D}_0$, $f_n(a_n) \to f(a)$ in $\mathbb{E}$. This can be seen by considering each pair separately; the result is obvious for asymptotically independent pairs, and for asymptotically dependent ones it follows from Lemma S9.

Finally, notice that the process G concentrates on $\mathbb{D}_0$. Therefore, by Equation (S1.13) and the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1),

$$\left(W_n^{(s)} \right)_{s\in\mathcal{P}} = f_n(G_n) + o_P(1) \rightsquigarrow f(G) = \left(B^{(s)} \right)_{s\in\mathcal{P}}$$

in $\mathbb{E}$. □

S1.2.2. *Proof of Theorem 5.* Similarly to the bivariate case, let

$$\Psi_n^{(s)}(\theta,\sigma) := (n/m)\Psi_n^{*(s)}(\theta, m\sigma/n).$$

As in the proof of Theorem 3, we may deduce that for every pair s, $\theta \in \widetilde{\Theta}$ and $\sigma > 0$,

$$\Psi^{(s)}(\theta,\sigma) - \Psi_n^{(s)}(\theta,\sigma) = \int g\left(\widehat{c}_n^{(s)} - c^{(s)}\right) d\mu_{\mathrm{L}} = \frac{1}{\sqrt{m}} \int g W_n^{(s)} d\mu_{\mathrm{L}},$$

with $W_n^{(s)}$ as defined in Theorem 4. By a similar argument to the bivariate case (involving the dominated convergence theorem and Lemma S10 to establish continuity of the mapping

$f \mapsto \int gf d\mu_{\mathrm{L}}$, see the proof of Theorem 3 for the applicability of Lemma S10), Theorem 4 and the continuous mapping theorem yield

$$\left(\int gW_n^{(s)} d\mu_{\mathrm{L}}\right)_{s\in\mathcal{P}} \rightsquigarrow \left(\int gB^{(s)} d\mu_{\mathrm{L}}\right)_{s\in\mathcal{P}}. \tag{S1.14}$$

The remaining proof consists of a number of successive applications of Lemma S11. We deal with each of the two estimators separately.

(i) For each pair s, applying Lemma S11 with $\phi = \Psi^{(s)}$, $x_0 = (h^{(s)}(\vartheta_0), 1)$, $a_n = 1/\sqrt{m}$ and $Y_n = \frac{1}{\sqrt{m}} \int gW_n^{(s)} d\mu_{\mathrm{L}}$ yields

$$\widehat{\theta}_n^{(s)} - h^{(s)}(\vartheta_0) = \frac{1}{\sqrt{m}} \mathcal{D}^{(s)} \int gW_n^{(s)} d\mu_{\mathrm{L}} + o_P\left(\frac{1}{\sqrt{m}}\right), \tag{S1.15}$$

where $\mathcal{D}^{(s)}$ is the block corresponding to the pair s in the matrix $\mathcal{D}$ defined in Equation (4.4); its existence, as well as the required smoothness of ϕ, are guaranteed by Condition 3. Now redefining ϕ as $\phi(\vartheta) = \big(h^{(s)}(\vartheta) - h^{(s)}(\vartheta_0)\big)_{s\in\mathcal{P}}$, we see that $\widehat{\vartheta}_n$ is in fact a minimizer of the norm of $\phi(\vartheta) - Y_n$, where Y_n is redefined as $\big(\widehat{\theta}_n^{(s)} - h^{(s)}(\vartheta_0)\big)_{s\in\mathcal{P}}$. Applying Lemma S11 again with ϕ and Y_n as above, $x_0 = \vartheta_0$ and $a_n = 1/\sqrt{m}$, we obtain

$$\begin{aligned} \widehat{\vartheta}_n - \vartheta_0 &= (J_1^\top J_1)^{-1} J_1^\top Y_n + o_P\left(\frac{1}{\sqrt{m}}\right) \\ &= \frac{1}{\sqrt{m}} (J_1^\top J_1)^{-1} J_1^\top \left(\mathcal{D}^{(s)} \int gW_n^{(s)} d\mu_{\mathrm{L}}\right)_{s\in\mathcal{P}} + o_P\left(\frac{1}{\sqrt{m}}\right), \end{aligned}$$

where the last equality follows from Equation (S1.15) and J_1 is defined as in Section 4.2 in the paragraph below Equation (4.4). The conclusion that $\sqrt{m}(\widehat{\vartheta}_n - \vartheta_0) \rightsquigarrow N(0, \Sigma_1)$ follows from this and Equation (S1.14).

(ii) Let $\widetilde{\sigma}_n = \frac{n}{m} \widetilde{\zeta}_n \in \mathbb{R}_+^{|\mathcal{P}|}$. Once more, we redefine

$$Y_n = \frac{1}{\sqrt{m}} \left(\int gW_n^{(s)} d\mu_{\mathrm{L}}\right)_{s\in\mathcal{P}} \quad \text{and} \quad \phi(\vartheta, \sigma) = \left(\Psi^{(s)}(h^{(s)}(\vartheta), \sigma^{(s)})\right)_{s\in\mathcal{P}}.$$

The estimator $(\widetilde{\vartheta}_n, \widetilde{\sigma}_n)$ can be seen to minimize the norm of $\phi - Y_n$. Therefore, applying Lemma S11 with $a_n = 1/\sqrt{m}$ and $x_0 = (\vartheta_0, 1, \ldots, 1)$, we obtain

$$(\widetilde{\vartheta}_n, \widetilde{\sigma}_n) - (\vartheta_0, 1, \cdots, 1) = \frac{1}{\sqrt{m}} (J_2^\top J_2)^{-1} J_2^\top \left(\int gW_n^{(s)} d\mu_{\mathrm{L}}\right)_{s\in\mathcal{P}} + o_P\left(\frac{1}{\sqrt{m}}\right),$$

which, combined with Equation (S1.14), implies $\sqrt{m}((\widetilde{\vartheta}_n, \widetilde{\sigma}_n) - (\vartheta_0, 1, \cdots, 1)) \rightsquigarrow N(0, \Sigma_2)$.

□

S2. Technical results used in Section S1. Throughout the paper, particularly the proof of Lemma S2 below, we use (without reference when obvious) the following results on regularly varying functions at 0.

LEMMA S1. *Suppose the functions f_1 and f_2 are regularly varying at 0 with indices ρ_1 and ρ_2, respectively.*

(i) *If $\rho_1 > 0$ (respectively $\rho_1 < 0$), $\lim_{t\to 0} f_1(t) = 0$ (respectively ∞).*
(ii) *For any $\alpha \in \mathbb{R}$, f_1^α is $(\alpha\rho_1)$–RV at 0.*

(iii) *The product* $f_1 f_2$ *is* $(\rho_1+\rho_2)$*–RV at 0.*
(iv) *If* $\lim_{t\to 0} f_2(t)=0$*, then* $f_1\circ f_2$ *is* $(\rho_1\rho_2)$*–RV at 0.*
(v) *If* $\rho_1>0$*, then* f_1^{-1} *is* $(1/\rho_1)$*–RV at 0, where we define the generalized inverse of* f_1 *as*

$$f_1^{-1}(t)=\inf\{u>0 : f_1(u)\geq t\}.$$

PROOF. The assertions (ii) and (iii) are trivial consequences of the definition of regular variation. As for (i), (iv) and (v), analogue versions for regularly varying functions at ∞ are proved in Proposition 0.8 of Resnick (1987). The proof can readily be adapted, using the fact that f is ρ–RV at 0 if and only if $u\mapsto 1/f(1/u)$ is ρ–RV at ∞.

□

LEMMA S2. (i) *Assume Equation* (3.1)*. Then there exists* $\eta\in(0,1]$ *such that* q *is a regularly varying (RV) function at 0 with index* $1/\eta$ *and* c *is* $1/\eta$*-homogeneous.*
(ii) *Assume Condition 1(i) and suppose that* q_1 *is non-decreasing and that there exists* $b>1$ *such that* $q_1(bt)=O(q_1(t))$ *as* $t\to 0$*. Then Equation* (3.1) *holds locally uniformly on* $[0,\infty)^2$*.*

REMARK S1. In part (ii) of the previous result, the monotonicity condition on q_1 is artificial; it can be removed at the cost of replacing $q_1(t)$ by the non-decreasing function $\bar{q_1}(t):=\sup_{0<s\leq t} q_1(s)$. Indeed, if Condition 1 is satisfied with q_1, it is trivially satisfied with $\bar{q_1}$. Moreover, if $q_1(bt)=O(q_1(t))$, $\bar{q_1}$ also satisfies the same property.

Because q_1 is positive non-decreasing, that required property implies that $q_1(bt)=O(q_1(t))$ holds for every $b\geq 1$ (Bingham, Goldie and Teugels, 1987, Corollary 2.0.6). The function q_1 is then said to be O-regularly varying at 0.

PROOF. (i) Recall that we assume $c(1,1)=1$. For any $x>0$, Equation (3.1) implies that $Q(tx,tx)=q(t)(c(x,x)+o(1))$ and $Q(tx,tx)=q(tx)(1+o(1))$. This can be manipulated into

$$\frac{q(tx)}{q(t)}=\frac{c(x,x)+o(1)}{1+o(1)}\longrightarrow c(x,x).$$

By Karamata's characterization theorem (Bingham, Goldie and Teugels, 1987, Theorem 1.4.1), q has to be ρ–RV and $c(x,x)=x^\rho$, for some $\rho\in\mathbb{R}$. However, since $q(t)\leq t$, we must have $\rho\geq 1$. Moreover, for any $a,x,y>0$,

$$c(ax,ay)=\lim_{t\to 0}\frac{Q(atx,aty)}{q(t)}=\lim_{t\to 0}\frac{Q(tx,ty)}{q(t/a)}=\lim_{t\to 0}\frac{Q(tx,ty)}{q(t)}\frac{q(t)}{q(t/a)}=a^\rho c(x,y).$$

Defining $\eta=1/\rho$, this proves (i).
(ii) For arbitrary $(x,y)\in[0,\infty)^2$, we write $(x,y)=a(u,v)$. We will prove that Equation (3.1) holds uniformly over all $(u,v)\in\mathcal{S}^+$ and over $a\in(0,b]$, for an arbitrary $b\in[1,\infty)$.

We have

$$\frac{Q(tx,ty)}{q(t)}=\frac{Q(atu,atv)}{q(t)}=\frac{q(at)}{q(t)}\frac{Q(atu,atv)}{q(at)}. \tag{S2.1}$$

First, the term $Q(atu,atv)/q(at)$ is equal to $c(u,v)+O(q_1(at))$ uniformly in $(u,v)\in\mathcal{S}^+$. In order to control the term $q(at)/q(t)$, we note that since q is $1/\eta$-RV, there exists a slowly varying function L such that for any $a > 0$,

$$\begin{aligned}\frac{L(at)}{L(t)} - 1 &= a^{-1/\eta}\left(\frac{q(at)}{q(t)} - c(a,a)\right)\\ &= a^{-1/\eta}\left(\frac{Q(at,at)(1+O(q_1(at)))}{q(t)} - c(a,a)\right)\\ &= a^{-1/\eta}\left(\frac{Q(at,at)}{q(t)} - c(a,a) + O(q_1(at))\right)\\ &= O(q_1(t)+q_1(at)) = O(q_1(bt)) = O(q_1(t)),\end{aligned}$$

where we have used the fact that $Q(at,at) = q(at)(1+O(q_1(at)))$, which can be reversed into $q(at) = Q(at,at)(1+O(q_1(at)))$. The function L is thus *slowly varying with remainder* (Bingham, Goldie and Teugels, 1987, Section 3.12). By theorem 3.12.1 of that book, the previous relation holds uniformly over all $a \in (1/2, b]$, so we henceforth focus on values $a \in (0, 1/2]$. Using Theorem 3.12.2 of the same book (which we adapt for slow variation at 0), we obtain that for some constants $C \in \mathbb{R}, T_0 \in (0,\infty)$ and for t small enough,

$$L(t) = \exp\left\{C + \delta_1(t) + \int_t^{T_0} \frac{\delta_2(s)}{s} ds\right\},$$

where the functions δ_j are real-valued, measurable and satisfy $|\delta_j(t)| \le K q_1(t)$ for some constant $K \in (0,\infty)$. The ratio $L(at)/L(t)$ becomes

$$\frac{L(at)}{L(t)} = \exp\left\{\delta_1(at) - \delta_1(t) + \int_{at}^{t} \frac{\delta_2(s)}{s} ds\right\}.$$

As $t \to 0$, we can use the monotonicity of q_1 to control the integral in the previous display:

$$\left|\int_{at}^{t} \frac{\delta_2(s)}{s} ds\right| \le K \int_{at}^{t} \frac{q_1(s)}{s} ds \le K q_1(t) \int_{at}^{t} \frac{ds}{s} = K q_1(t) \log\left(\frac{1}{a}\right).$$

Because $a \le 1/2$, $\log(1/a)$ is lower bounded, so K can be chosen large enough so that $K q_1(t)\log(1/a)$ also upper bounds the absolute value of $\delta_1(at) - \delta_1(t) + \int_{at}^{t} \frac{\delta_2(s)}{s} ds$. Therefore, using the fact that for every $h \in \mathbb{R}$, $|e^h - 1| \le e^{|h|} - 1$, we obtain

$$\left|\frac{L(at)}{L(t)} - 1\right| \le \exp\left\{K q_1(t) \log\left(\frac{1}{a}\right)\right\} - 1 = a^{-K q_1(t)} - 1.$$

What we are interested in is bounding $q(at)/q(t) - a^{1/\eta}$. This can be done by recalling that

$$\left|\frac{q(at)}{q(t)} - a^{1/\eta}\right| = a^{1/\eta}\left|\frac{L(at)}{L(t)} - 1\right| \le a^{1/\eta}\left(a^{-K q_1(t)} - 1\right) =: \tau(a,t). \tag{S2.2}$$

By simple differentiation, it is straightforward to see that for a fixed value of t small enough so that $K q_1(t) < 1/\eta$, the function τ is differentiable in its first argument and that

$$\frac{\partial}{\partial a}\tau(a,t) = a^{1/\eta - 1}\left((1/\eta - K q_1(t)) a^{-K q_1(t)} - 1/\eta\right).$$

This suggests that the function attains its unique maximum at the point $a_{\max}(t) := (1-\eta K q_1(t))^{1/(Kq_1(t))}$. Considering Equation (S2.2), we obtain that for all $a \in (0, 1/2]$,

$$\begin{aligned}\left|\frac{q(at)}{q(t)} - a^{1/\eta}\right| &\le \tau(a_{\max}(t), t) \\ &= (1-\eta K q_1(t))^{1/(\eta K q_1(t))}\left(\frac{1}{1-\eta K q_1(t)} - 1\right) \\ &= O(q_1(t))\end{aligned}$$

as $t \to 0$, since $(1-\eta K q_1(t))^{1/(\eta K q_1(t))} \to e^{-1}$ and since the function $x \mapsto 1/(1-x)$ is continuously differentiable at 0. Finally, this allows us to rewrite Equation (S2.1) as

$$\frac{Q(tx, ty)}{q(t)} = \left(a^{1/\eta} + O(q_1(t))\right)(c(u,v) + O(q_1(at))) = a^{1/\eta}c(u,v) + O(q_1(t)),$$

and the last equation holds uniformly over $a \in (0, b]$ and $(u,v) \in \mathcal{S}^+$. The proof is over since $a^{1/\eta}c(u,v) = c(x,y)$.

□

LEMMA S3. *Let $\varphi : (0, T] \to (0, \infty)$ be a non-decreasing function such that $\varphi(t)/\sqrt{t} \to \infty$ as $t \to 0$ and assume there exists $c > 0$ such that*

$$\int_0^T \frac{1}{x} \exp\left\{-c\frac{\varphi^2(x)}{x}\right\} dx < \infty.$$

Then under the assumptions of Theorem 1, for every $\lambda \in (0,1)$ we have

$$\sup_{\lambda/k \le x \le T} \frac{\sqrt{k}}{\varphi(x)}|u_n(x) - x| = O_P(1),$$

where u_n is defined as in Section S1.1. In particular, note that $\varphi(x) := 1$, as well as any function that satisfies $\varphi(x) := \sqrt{x \log\log(1/x)}$ in a neighborhood of 0, are valid choices.

PROOF. This is essentially proved in Csörgő and Horváth (1987), up to a slight difference between their definition of the quantiles and ours. We prove here that this difference does not change the result. More precisely, their Theorem 2.6 (ii) states that

$$\sup_{\lambda/k \le x \le T} \frac{|w_n(x)|}{\varphi(x)} = O_P(1), \tag{S2.3}$$

where we denote w_n what they call v_n (to avoid confusion with our definitions). From their definitions, one easily sees that

$$w_n(x) = \frac{n}{\sqrt{k}}\left(\frac{k}{n}x - U_{n,\lceil kx\rceil}\right) = \sqrt{k}\left(x - \frac{n}{k}U_{n,\lceil kx\rceil}\right).$$

Then, by the reverse triangle inequality,

$$\begin{aligned}|\sqrt{k}|u_n(x) - x| - |w_n(x)|| &\le |\sqrt{k}(u_n(x) - x) + w_n(x)| \\ &= \sqrt{k}\left|u_n(x) - \frac{n}{k}U_{n,\lceil kx\rceil}\right| = \frac{n}{\sqrt{k}}\left(U_{n,\lceil kx\rceil} - U_{n,\lfloor kx\rfloor}\right).\end{aligned}$$

Using this and the inequality $\lfloor x \rfloor \geq \lceil x \rceil - 1$, we have

$$\begin{aligned}
&\left| \sup_{\lambda/k \leq x \leq T} \frac{\sqrt{k}}{\varphi(x)} |u_n(x) - x| - \sup_{\lambda/k \leq x \leq T} \frac{|w_n(x)|}{\varphi(x)} \right| \\
&\leq \frac{n}{\sqrt{k}} \sup_{\lambda/k \leq x \leq T} \frac{1}{\varphi(x)} \left(U_{n,\lceil kx \rceil} - U_{n,\lfloor kx \rfloor}\right) \\
&\leq \frac{n}{\sqrt{k}} \sup_{\lambda/k \leq x \leq T} \frac{1}{\varphi(x)} \left(U_{n,\lceil kx \rceil} - U_{n,\lceil kx \rceil - 1}\right) \\
&\leq \frac{n}{\sqrt{k}} \sup_{\lambda/k \leq x \leq (1+\lambda)/k} \frac{1}{\varphi(x)} \left(U_{n,\lceil kx \rceil} - U_{n,\lceil kx \rceil - 1}\right) \\
&\quad + \frac{n}{\sqrt{k}} \sup_{(1+\lambda)/k \leq x \leq T} \frac{1}{\varphi(x)} \left(U_{n,\lceil kx \rceil} - U_{n,\lceil kx \rceil - 1}\right). \qquad \text{(S2.4)}
\end{aligned}$$

In the first term, since $\lambda/k \leq x \leq (1+\lambda)/k$ and $\lambda \in (0,1)$, we must have $\lceil kx \rceil \in \{1, 2\}$. Therefore, we end up studying $U_{n,i} - U_{n,i-1}$, for some $i \in \{1, 2\}$. It is a well known fact that those differences, regardless of the value of i, have a Beta distribution with parameters 1 and n. In particular, they are both $O_P(1/n)$. It follows that the first supremum on the right hand side of Equation (S2.4) is asymptotically bounded in probability by

$$\frac{1}{\sqrt{k}} \sup_{\lambda/k \leq x \leq (1+\lambda)/k} \frac{1}{\varphi(x)} = \frac{1}{\sqrt{k}\varphi(\lambda/k)} \longrightarrow 0$$

by assumption on φ. As for the second term in Equation (S2.4), it is equal to

$$\begin{aligned}
&\frac{n}{\sqrt{k}} \sup_{(1+\lambda)/k \leq x \leq T} \frac{1}{\varphi(x)} \left(U_{n,\lceil kx \rceil} - U_{n,\lceil k(x-1/k) \rceil}\right) \\
&= \frac{n}{\sqrt{k}} \sup_{\lambda/k \leq x \leq T - 1/k} \frac{1}{\varphi(x+1/k)} \left(U_{n,\lceil k(x+1/k) \rceil} - U_{n,\lceil kx \rceil}\right)
\end{aligned}$$

after shifting x to the right by $1/k$. Using Equation (S2.3), this is in turn equal to

$$\begin{aligned}
&\frac{n}{\sqrt{k}} \sup_{\lambda/k \leq x \leq T - 1/k} \frac{1}{\varphi(x+1/k)} \left(\frac{k}{n}\left(x + \frac{1}{k}\right) - \frac{k}{n}x\right) + \frac{n}{\sqrt{k}} O_P\left(\frac{\sqrt{k}}{n}\right) \\
&= \frac{1}{\sqrt{k}} \sup_{\lambda/k \leq x \leq T - 1/k} \frac{1}{\varphi(x+1/k)} + O_P(1) \\
&= \frac{1}{\sqrt{k}\varphi((1+\lambda)/k)} + O_P(1) \\
&= O_P(1)
\end{aligned}$$

once again by the properties of φ. We have shown that the difference between the quantity we are interested in and the term appearing in Equation (S2.3) is $O_P(1)$. We may thus conclude, by Equation (S2.3), that

$$\sup_{\lambda/k \leq x \leq T} \frac{\sqrt{k}}{\varphi(x)} |u_n(x) - x| = \sup_{\lambda/k \leq x \leq T} \frac{|w_n(x)|}{\varphi(x)} + O_P(1) = O_P(1).$$

□

COROLLARY S1. *Define the random functions u_n and v_n as in Section S1.1. Then, as $n \to \infty$,*

$$\sup_{0 \le x \le 2T} |u_n(x) - x| \quad \textit{and} \quad \sup_{0 \le y \le 2T} |v_n(y) - y|$$

are both $O_P\left(1/\sqrt{k}\right)$.

PROOF. Note that by definition, $u_n(z) = v_n(z) = 0$ whenever $z < 1/k$. It follows that

$$\begin{aligned}
\sup_{0 \le x \le 2T} |u_n(x) - x| &\le \sup_{0 \le x < 1/k} |u_n(x) - x| + \sup_{1/k \le x \le 2T} |u_n(x) - x| \\
&= \sup_{0 \le x < 1/k} x + \sup_{1/k \le x \le 2T} |u_n(x) - x| \\
&= \frac{1}{k} + \sup_{1/k \le x \le 2T} |u_n(x) - x|.
\end{aligned}$$

This is $O_P\left(1/\sqrt{k}\right)$ by the preceding Lemma S3 with the function $\varphi(x) = 1$. The same proof holds with u_n replaced by v_n.

□

LEMMA S4. *Under Condition 1 the process H_n as defined in Equation (S1.1) converges to the process W from Theorem 1 in $\ell^\infty([0, 2T]^2)$.*

PROOF. Denoting $f_{n,(x,y)}(u, v) := \sqrt{\frac{n}{m}} \mathbb{1}\left\{u \le \frac{k}{n}x, v \le \frac{k}{n}y\right\}$, we see that H_n can be written as

$$H_n(x, y) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f_{n,(x,y)}(U_i, V_i) - \mathbb{E}\left[f_{n,(x,y)}(U, V)\right]\right).$$

Therefore, convergence of the process H_n to a Gaussian process in $\ell^\infty([0, 2T]^2)$ is equivalent to checking that the sequence of function classes

$$\mathcal{F}_n = \{f_{n,(x,y)} : (x, y) \in [0, 2T]^2\}$$

are Donsker classes for the distribution of (U, V). This is guaranteed by Theorem 11.20 of Kosorok (2008), provided that we can check the six conditions. Note that $\mathcal{F}_n$ admits the envelope function $F_n = f_{n,(2T,2T)}$.

(0) First, the AMS condition is trivially satisfied; by right continuity of indicator functions, for any $n \in \mathbb{N}$, $(x, y) \in [0, 2T]^2$ and $(u, v) \in [0, 1]^2$,

$$\inf_{(x',y') \in \mathbb{Q}^2} |f_{n,(x',y')}(u, v) - f_{n,(x,y)}(u, v)| = 0.$$

It follows that Equation (11.7) of Kosorok (2008) is satisfied with $T_n = \mathbb{Q}^2$, which is countable. Hence the classes $\mathcal{F}_n$ are AMS.

(A) For every n, it is easily checked that $\mathcal{F}_n$ is a VC class with VC-index 2. Therefore, condition (A) is a direct consequence of Lemma 11.21 of Kosorok (2008).

(B) For $(x, y), (x', y') \in [0, 2T]^2$ arbitrary, it follows from the definition of H_n that

$$\begin{aligned}
\mathbb{E}\left[H_n(x, y)H_n(x', y')\right] &= \mathbb{E}\left[f_{n,(x,y)}(U, V)f_{n,(x',y')}(U, V)\right] \\
&\quad - \mathbb{E}\left[f_{n,(x,y)}(U, V)\right]\mathbb{E}\left[f_{n,(x',y')}(U, V)\right]
\end{aligned}$$

$$= \frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}(x \wedge x'), V \le \frac{k}{n}(y \wedge y')\right)$$
$$- \frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}x, V \le \frac{k}{n}y\right)\mathbb{P}\left(U \le \frac{k}{n}x', V \le \frac{k}{n}y'\right).$$

Recall that $n/m = 1/q(k/n)$. Therefore, the first term of the last display converges to $c(x \wedge x', y \wedge y')$. The second term vanishes since both probabilities are of the order of m/n. The convariance functions of H_n thus converge pointwise to the covariance function of W.

(C) By definition of the envelope functions and by assumption, we have

$$\limsup_{n\to\infty} \mathbb{E}\left[F_n^2(U,V)\right] = \limsup_{n\to\infty} \frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}2T, V \le \frac{k}{n}2T\right) = c(2T,2T) < \infty.$$

(D) For every $\varepsilon > 0$,

$$\mathbb{E}\left[F_n^2(U,V)\mathbb{1}\left\{F_n(U,V) > \varepsilon\sqrt{n}\right\}\right] \le \frac{n}{m}\mathbb{1}\left\{\sqrt{\frac{n}{m}} > \varepsilon\sqrt{n}\right\},$$

which is equal to 0 as soon as $m \ge \varepsilon^{-2}$.

(E) We first recall that for arbitrary events A, B,

$$\mathbb{P}\left(\mathbb{1}_A \ne \mathbb{1}_B\right) = \mathbb{P}\left(A \backslash B\right) + \mathbb{P}\left(B \backslash A\right) = \mathbb{P}\left(A\right) + \mathbb{P}\left(B\right) - 2\mathbb{P}\left(A \cap B\right).$$

A direct application of this fact yields

$$\begin{aligned}
\rho_n^2((x,y),(x',y')) &:= \mathbb{E}\left[(f_{n,(x,y)}(U,V) - f_{n,(x',y')}(U,V))^2\right] \\
&= \frac{n}{m}\mathbb{P}\left(\mathbb{1}\left\{U \le \frac{k}{n}x, V \le \frac{k}{n}y\right\} \ne \mathbb{1}\left\{U \le \frac{k}{n}x', V \le \frac{k}{n}y'\right\}\right) \\
&= \frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}x, V \le \frac{k}{n}y\right) + \frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}x', V \le \frac{k}{n}y'\right) \\
&\quad - 2\frac{n}{m}\mathbb{P}\left(U \le \frac{k}{n}(x \wedge x'), V \le \frac{k}{n}(y \wedge y')\right) \\
&\longrightarrow c(x,y) + c(x',y') - 2c(x \wedge x', y \wedge y') \\
&=: \rho^2((x,y),(x',y')).
\end{aligned}$$

Moreover, by Lemma S2(ii), this convergence is uniform over $[0,2T]^4$. This means that for any sequences x_n, y_n, x_n', y_n' in $[0,2T]$ such that $\rho((x_n,y_n),(x_n',y_n')) \to 0$, $\rho_n((x_n,y_n),(x_n',y_n'))$ is equal to

$$\begin{aligned}
&\{\rho_n((x_n,y_n),(x_n',y_n')) - \rho((x_n,y_n),(x_n',y_n'))\} + \rho((x_n,y_n),(x_n',y_n')) \\
&\le \sup_{(x,y,x',y') \in [0,2T]^4} |\rho_n((x,y),(x',y')) - \rho((x,y),(x',y'))| \\
&\quad + \rho((x_n,y_n),(x_n',y_n')) \\
&\longrightarrow 0.
\end{aligned}$$

Finally, the theorem implies that $H_n \rightsquigarrow W$ in $\ell^\infty([0,2T]^2)$.

$\square$

LEMMA S5. *Let Q be a bivariate copula. If there exists a positive function q and a finite function c that is not everywhere 0 such that for every $(x,y) \in [0,\infty)^2$, as $n \to \infty$,*

$$\frac{Q(x/n, y/n)}{q(1/n)} \longrightarrow c(x,y),$$

then there exists a measure ν such that for every $(x,y) \in [0,\infty)^2$, $c(x,y) = \nu((0,x] \times (0,y])$. Note that Equation (3.1) satisfies this setting.

PROOF. Define the measures ν_n by

$$\nu_n((0,x] \times (0,y]) = \frac{Q(x/n, y/n)}{q(1/n)}$$

and fix $a \in (0,\infty)$. Note that since c is not everywhere 0, $c(a,a)$ is eventually positive, so for n and a large enough, $\nu_n((0,a]^2) > 0$. Then clearly

$$P_{n,a} := \left(\nu_n((0,a]^2)\right)^{-1} \nu_n$$

is a probability measure on $[0,a]^2$. Since it is supported on the same compact set for every n, the sequence $\{P_{n,a} : n \in \mathbb{N}\}$ is tight. Thus, by Helly's selection theorem there exists a probability measure P_a also supported on $[0,a]^2$ and a subsequence $\{n_j : j \in \mathbb{N}\}$ such that $P_{n_j,a} \rightsquigarrow P_a$. However, by definition of ν_n, we have for every $(x,y) \in [0,a]^2$

$$P_{n_j,a}((0,x] \times (0,y]) \longrightarrow \frac{c(x,y)}{c(a,a)}.$$

Therefore, we must have $P_a((0,x] \times (0,y]) = c(x,y)/c(a,a)$, so choosing $\nu_a = c(a,a)P_a$, the result holds for every $(x,y) \in [0,a]^2$. However, the value of $\nu_a((0,x] \times (0,y])$ is independent of a (as long as $x \vee y \leq a$), so ν_a can be uniquely extended to a measure ν on the bounded Borel sets of $[0,\infty)^2$.

□

LEMMA S6 (similar to Theorem 1 in Ramos and Ledford (2009)). *Define the function c as in Equation (3.1). Then there exists a finite measure H on $[0,1]$ such that, for every $(x,y) \in [0,\infty)^2$,*

$$c(x,y) = \int_{[0,1]} \left(\frac{x}{1-w} \wedge \frac{y}{w}\right)^{1/\eta} H(dw).$$

It is also useful to note that this integral is equal to

$$\int_{\left[0,\frac{y}{x+y}\right]} \left(\frac{x}{1-w}\right)^{1/\eta} H(dw) + \int_{\left(\frac{y}{x+y},1\right]} \left(\frac{y}{w}\right)^{1/\eta} H(dw).$$

PROOF. By Lemma S5, we can write

$$c(x,y) = \nu((0,x] \times (0,y]) = \int_{[0,\infty)^2} \mathbb{1}_{(0,x]\times(0,y]} d\nu = \int_{[0,\infty)^2\setminus\{0\}} \mathbb{1}_{[0,x]\times[0,y]} d\nu. \tag{S2.5}$$

In the last equality, nothing changed since $\nu((0,x] \times \{0\} \cup \{0\} \times (0,y]) \leq c(x,0) + c(0,y) = 0$. Then, through the mapping $f : [0,\infty)^2\setminus\{0\} \to (0,\infty) \times [0,1]$ defined by $f(x,y) = (x+y, \frac{y}{x+y})$, define the push-forward measure $\mu = \nu \circ f^{-1}$. By homogeneity of ν, we see that μ is a product measure:

$$\mu((0,r] \times (0,w]) = r^{1/\eta}\mu((0,1] \times (0,w]) =: G((0,r])H((0,w]),$$

where G is a measure on $(0,\infty)$ and H is a measure on $[0,1]$. Finally, for any (x,y), define the function $g:(0,\infty)\times[0,1]\to\mathbb{R}$ as

$$g(r,w)=\mathbb{1}\left\{r\le\frac{x}{1-w}\wedge\frac{y}{w}\right\},$$

so that $g\circ f=\mathbb{1}_{[0,x]\times[0,y]}$. Using Equation (S2.5) and Theorem 9.15 from Teschl (1998), we have

$$\begin{aligned}c(x,y)&=\int_{[0,\infty)^2\setminus\{0\}}g\circ f d\nu\\&=\int_{(0,\infty)\times[0,1]}g d\mu\\&=\int_{[0,1]}\int_{(0,\infty)}\mathbb{1}_{(0,\frac{x}{1-w}\wedge\frac{y}{w}]}(r)G(dr)H(dw)\\&=\int_{[0,1]}\left(\frac{x}{1-w}\wedge\frac{y}{w}\right)^{1/\eta}H(dw),\end{aligned}$$

where we used Fubini's theorem to write the integral with respect to the product measure μ as a double integral. Moreover, note that H is finite since

$$H([0,1])=\mu((0,1]\times[0,1])=\nu\left(\left\{(x,y)\in[0,\infty)^2:x+y\le1\right\}\right)\le c(1,1)=1.$$

□

LEMMA S7. *Define the function c as in Equation (3.1). Then for every $(x,y)\in[0,T]^2$ and $h>0$,*

$$c(x+h,y)-c(x,y)\le\frac{1}{\eta}h\frac{c(x+h,y)}{x+h}.$$

PROOF. By Lemma S6, write

$$c(x,y)=\int_{[0,1]}\left(\frac{x}{1-w}\wedge\frac{y}{w}\right)^{1/\eta}H(dw)=:\int_{[0,1]}f(x,y,w)H(dw).$$

Clearly, it is sufficient to prove that for every x,y,h,w,

$$f(x+h,y,w)-f(x,y,w)\le\frac{1}{\eta}h\frac{f(x+h,y,w)}{x+h},\tag{S2.6}$$

because then the result follows by integrating both sides. To prove Equation (S2.6), first note that for any y,w,

$$f(x,y,w)=\begin{cases}\left(\frac{x}{1-w}\right)^{1/\eta}&,\quad x\le\frac{1-w}{w}y\\\left(\frac{y}{w}\right)^{1/\eta}&,\quad x\ge\frac{1-w}{w}y\end{cases}.$$

As a function of x, this is continuously differentiable everywhere on $(0,T]$ except at the change point $x=\frac{1-w}{w}y$ and its derivative with respect to x, f', is equal to $f(x,y,h)/(\eta x)$ on the first part and 0 on the second. From here we consider three different cases, depending on the position of the change point with respect to x and $x+h$.

First, if $x+h\le\frac{1-w}{w}y$,

$$f(x+h,y,w)-f(x,y,w)=hf'(\xi,y,w)=h\frac{f(\xi,y,w)}{\eta\xi},$$

for some $\xi \in [x, x+h]$, by Taylor's theorem. By monotonicity, this is upper bounded by

$$\frac{1}{\eta} h \frac{f(x+h,y,w)}{x+h}.$$

Next, if $\frac{1-w}{w} y \leq x$, $f(x+h,y,w) - f(x,y,w) = 0$ so the result is trivial.
Finally, if $x < \frac{1-w}{w} y < x+h$,

$$f(x+h,y,w) - f(x,y,w) = f\left(\frac{1-w}{w} y, y, w\right) - f(x,y,w) = \left(\frac{1-w}{w} y - x\right) \frac{f(\xi,y,w)}{\eta \xi},$$

for ξ between x and $\frac{1-w}{w} y$, once again by Taylor's theorem. By monotonicity, we have

$$\frac{f(\xi,y,w)}{\eta\xi} \leq \frac{1}{\eta \frac{1-w}{w} y} \left(\frac{y}{w}\right)^{1/\eta} = \frac{1}{\eta \frac{1-w}{w} y} f(x+h,y,w).$$

Moreover,

$$\frac{\frac{1-w}{w} y - x}{\frac{1-w}{w} y} \leq \frac{(x+h) - x}{(x+h)} = \frac{h}{x+h},$$

because the function $t \mapsto (t-x)/t$ is non-decreasing. Piecing everything together, we have

$$f(x+h,y,w) - f(x,y,w) \leq \frac{1}{\eta} h \frac{f(x+h,y,w)}{x+h}.$$

We have proved that Equation (S2.6) holds for every $(x,y) \in [0,T]^2$, $h > 0$ and $w \in [0,1]$. □

LEMMA S8. *Define the function c as in Equation (3.1) and assume Condition 1(i). Then there exists a constant $K := K_T < \infty$ such that for every $(x,y) \in [0,T]^2$,*

$$c(x,y) \leq \frac{K}{\log(1/x)}.$$

PROOF. We will prove that as $x \to 0$,

$$c(x,y) \lesssim \frac{1}{\log(1/x)}$$

uniformly for all $y \in [0,T]$. Since c is locally bounded, the result will follow.

Since Condition 1(i) is satisfied, we may assume it is satisfied with the function $q_1(t) = 1/\log(1/t)$. Recall that as $t \downarrow 0$, by Lemma S2,

$$Q(tx,ty) = q(t)c(x,y) + O(q(t)q_1(t))$$

uniformly over all $(x,y) \in [0,T]^2$. That is,

$$c(x,y) = \frac{Q(tx,ty)}{q(t)} + O(q_1(t)) \leq \frac{tx}{q(t)} + O(q_1(t)) \tag{S2.7}$$

uniformly, by Lipschitz continuity of the copula Q. The previous relation holds whenever $t \to 0$, and in particular it holds when t and x are related and both tend to 0.

Define $g(t) = q(t)q_1(t)/t \to 0$ as $t \to 0$. We argue, in the following, that for any x small enough, there exists $t(x) > 0$ such that $x \leq g(t(x)) \leq 2^{1/\eta} x$. Plugging $t(x)$ into Equation (S2.7), we find that as $x \to 0$,

$$c(x,y) \leq \frac{t(x)x}{q(t(x))} + O(q_1(t(x))) = O(q_1(t(x))), \tag{S2.8}$$

because, since we assume $x \le g(t(x))$,

$$\frac{t(x)x}{q(t(x))} \le \frac{t(x)}{q(t(x))} g(t(x)) = q_1(t(x)).$$

Moreover, since the function g is ρ-RV at 0, $\rho := 1/\eta - 1$, for small enough t we have $g(t) \ge t^\alpha$, as long as $\alpha > \rho$. This means that

$$q_1(t(x)) = \frac{1}{\log(1/t(x))} = \frac{\alpha}{\log(1/t(x)^\alpha)} \lesssim \frac{1}{\log(1/g(t(x)))}.$$

Finally, by the assumption that $g(t(x)) \le 2^{1/\eta} x$, we get

$$q_1(t(x)) \lesssim \frac{1}{\log(1/g(t(x)))} \lesssim \frac{1}{\log(1/x)}$$

which, in conjunction with Equation (S2.8), yields the desired bound for $c(x, y)$ as $x \to 0$, uniformly over bounded y.

The only thing left is to prove the existence of a point $t(x)$ such that $g(t(x)) \in [x, 2^{1/\eta} x]$ for every small enough x. This can be done by using the fact that the function g is ρ-RV at 0. Applying Theorem 1.5.6(iii) in Bingham, Goldie and Teugels (1987) (adapted to functions of regular variation at 0) with any $\delta \in (0, 1)$ and $A = 2^{1-\delta}$, we find that there exists $T_0 \in (0, \infty)$ such that for every $t \le T_0$,

$$\frac{g(t)}{g(t/2)} \le 2^{1-\delta} 2^{\rho+\delta} = 2^{1/\eta}.$$

We now construct a non-increasing sequence the follwing way: take $t_0 = T_0$ and for $n \in \mathbb{N}$, define $t_n = t_{n-1}/2$ if $g(t_{n-1}/2) \le g(t_{n-1})$. Otherwise, $t_n = t_{n-1}/4$ if $g(t_{n-1}/4) \le g(t_{n-1})$. Otherwise, we try $t_{n-1}/8$, etc. In general

$$t_n = \max \left\{ \frac{t_{n-1}}{2^k} : k \in \mathbb{N}, g\left(\frac{t_{n-1}}{2^k}\right) \le g(t_{n-1}) \right\}.$$

Therefore, the sequence satisfies, for every natural n,

$$1 \le \frac{g(t_n)}{g(t_{n+1})} \le 2^{1/\eta}. \tag{S2.9}$$

Now choose any $x \in (0, T_0/2]$ and let $t = \min_{n \in \mathbb{N}} \{t_n : g(t_n) \ge x\}$. Clearly, $g(t) \ge x$, and $g(t)$ has to be $\le 2^{1/\eta} x$. Indeed, suppose the opposite. Then by Equation (S2.9), $g(t_{n+1}) \ge g(t)/2^{1/\eta} > x$, which contradicts the definition of t. We conclude that for every $x \in (0, T_0/2]$, the desired $t(x)$ exists.

□

LEMMA S9. *Assume the setting of Theorem 2. For arbitrary positive t and T, let*

$$\mathcal{V}(t) := \{b \in \ell^\infty([0, 2T]) : \forall x \in [0, T], x + tb(x) \in [0, 2T]\}.$$

Let $t_n \downarrow 0$ and assume that $b_n := (b_n^{(1)}, b_n^{(2)}) \in \mathcal{V}(t_n)^2$ converges uniformly to a continuous function $b = (b^{(1)}, b^{(2)})$ such that $b^{(1)}(0) = b^{(2)}(0) = 0$. Then, the functions $g_n : [0, T] \to \mathbb{R}$ defined by

$$g_n(x, y) := \frac{c\left(x + t_n b_n^{(1)}(x), y + t_n b_n^{(2)}(y)\right) - c(x, y)}{t_n}$$

hypi-converge to $\dot{c}_1(x, y) b^{(1)}(x) + \dot{c}_2(x, y) b^{(2)}(y)$, where $\dot{c}_1$ and $\dot{c}_2$ are defined as in Section 4.1.1.

PROOF. Let ℓ be the stable tail dependence function associated to the random vector (X,Y). Because we assume asymptotic dependence, we know that $\chi := \lim_{t\downarrow 0} q(t)/t > 0$ and that $c(x,y) = (x+y-\ell(x,y))/\chi$. Then,

$$g_n(x,y) = \chi^{-1}\left(b_n^{(1)}(x) + b_n^{(2)}(y) - \frac{\ell\left(x+t_n b_n^{(1)}(x), y+t_n b_n^{(2)}(y)\right) - \ell(x,y)}{t_n}\right).$$

By assumption, the sum of the first two terms converges uniformly to $b^{(1)}(x)+b^{(2)}(y)$. Let $\mathcal{S}\subset[0,\infty)^2$ be the set of points where ℓ is differentiable. Since ℓ is convex, the complement of $\mathcal{S}$ is Lebesgue-null and the gradient of ℓ is continuous on $\mathcal{S}$ (Rockafellar, 1970, Theorem 25.5). By Lemma F.3 of Bücher, Segers and Volgushev (2014), the last term hypi-converges to

$$\mathcal{L}_1(x,y) := \sup_{\varepsilon>0}\inf\left\{\dot{\ell}_1(x',y')b^{(1)}(x') + \dot{\ell}_2(x',y')b^{(2)}(y') : (x',y')\in\mathcal{S}, \|(x,y)-(x',y')\|<\varepsilon\right\},$$

where $\dot{\ell}_j$ are defined like $\dot{c}_j$: $\dot{\ell}_1(x,y)$ is the first partial derivative at (x,y) from the left, except if $x=0$ in which case it is from the right, and $\dot{\ell}_2$ is always the second partial derivative from the right. We argue below that the hypi-distance between the functions $\mathcal{L}_1$ and $\mathcal{L}_2$, defined by $\mathcal{L}_2(x,y) = \dot{\ell}_1(x,y)b^{(1)}(x) + \dot{\ell}_2(x,y)b^{(2)}(y)$, is 0. That is, $\mathcal{L}_1$ and $\mathcal{L}_2$ belong to the same equivalence class in the space $L^\infty([0,2T]^2)$ and hypi-convergence to $\mathcal{L}_1$ is equivalent to hypi-convergence to $\mathcal{L}_2$. It follows that $g_n(x,y)$ hypi-converges to

$$\frac{b^{(1)}(x)+b^{(2)}(y)-\mathcal{L}_2(x,y)}{\chi} = \dot{c}_1(x,y)b^{(1)}(x) + \dot{c}_2(x,y)b^{(2)}(y), \tag{S2.10}$$

where the last equality is a consequence of the relation $\dot{\ell}_j(x,y) = 1-\chi\dot{c}_j(x,y)$, $j\in\{1,2\}$.

To prove the equivalence between $\mathcal{L}_1$ and $\mathcal{L}_2$, first note that by continuity of $b^{(1)}$ and $b^{(2)}$,

$$\mathcal{L}_1(x,y) := \sup_{\varepsilon>0}\inf\left\{\dot{\ell}_1(x',y')b^{(1)}(x) + \dot{\ell}_2(x',y')b^{(2)}(y) : (x',y')\in\mathcal{S}, \|(x,y)-(x',y')\|<\varepsilon\right\}.$$

Let $\dot{\ell}_j^-$ and $\dot{\ell}_j^+$ denote the directional partial derivatives of ℓ from the left and right, respectively. The function $\mathcal{L}_2$ can then be expressed the following way, and we analogously define $\mathcal{L}_3$:

$$\mathcal{L}_2(x,y) = \dot{\ell}_1^-(x,y)b^{(1)}(x) + \dot{\ell}_2^+(x,y)b^{(2)}(y), \quad \mathcal{L}_3(x,y) := \dot{\ell}_1^+(x,y)b^{(1)}(x) + \dot{\ell}_2^-(x,y)b^{(2)}(y).$$

The main tool is the homogeneity property of ℓ ($\ell(ax,ay) = a\ell(x,y)$, $a\geq 0$). It implies that the directional derivatives $\dot{\ell}_j^\pm$ are constant along rays of the form $\{az : a>0\}$, $z\in(0,\infty)^2$ and therefore that $\mathcal{S}$ consists exactly of a dense union of such rays.

Fix a point $(x,y)\in(0,\infty)^2$. For any sufficiently small $\varepsilon>0$, the open ε-ball $B(\varepsilon)$ around (x,y) can be partitioned into the two open "half-balls"

$$B_1(\varepsilon) := \{(x',y')\in B(\varepsilon) : y'/x' > y/x\}, \quad B_2(\varepsilon) := \{(x',y')\in B(\varepsilon) : y'/x' < y/x\}$$

and the line $B_3(\varepsilon) := \{(x',y')\in B(\varepsilon) : y'/x' = y/x\}$. Provided that ε is sufficiently small, there exists a positive $\delta=\delta(\varepsilon)$ such that $\delta(\varepsilon)\to 0$ as $\varepsilon\to 0$, such that each point in $B_1(\varepsilon)$ is on the same ray as some $u\in(x-\delta,x]\times\{y\}$ and some $v\in\{x\}\times[y,y+\delta)$ and such that each point in $B_2(\varepsilon)$ is on the same ray as some $u\in(x,x+\delta)\times\{y\}$ and some $v\in\{x\}\times(y-\delta,y)$. By Rockafellar (1970), Theorem 24.1, we have

$$\lim_{\delta\downarrow 0}\dot{\ell}_1^\pm(x-\delta,y) = \dot{\ell}_1^-(x,y), \quad \lim_{\delta\downarrow 0}\dot{\ell}_1^\pm(x+\delta,y) = \dot{\ell}_1^+(x,y),$$

$$\lim_{\delta\downarrow 0}\dot{\ell}_2^\pm(x,y-\delta) = \dot{\ell}_2^-(x,y), \quad \lim_{\delta\downarrow 0}\dot{\ell}_2^\pm(x,y+\delta) = \dot{\ell}_2^+(x,y).$$

Then, as $\varepsilon \to 0$, the vectors $(\dot\ell_1^\pm(x',y'), \dot\ell_2^\pm(x',y'))$ converge to $(\dot\ell_1^-(x,y), \dot\ell_2^+(x,y))$ for $(x',y') \in B_1(\varepsilon)$ and to $(\dot\ell_1^+(x,y), \dot\ell_2^-(x,y))$ for $(x',y') \in B_2(\varepsilon)$. It follows by continuity of b that for any sufficiently small $\varepsilon > 0$,

$$\lim_{(x',y')\to(x,y),(x',y')\in B_1(\varepsilon)} \mathcal{L}_2(x',y') = \lim_{(x',y')\to(x,y),(x',y')\in B_1(\varepsilon)} \mathcal{L}_3(x',y') = \mathcal{L}_2(x,y) \tag{S2.11}$$

$$\lim_{(x',y')\to(x,y),(x',y')\in B_2(\varepsilon)} \mathcal{L}_2(x',y') = \lim_{(x',y')\to(x,y),(x',y')\in B_2(\varepsilon)} \mathcal{L}_3(x',y') = \mathcal{L}_3(x,y) \tag{S2.12}$$

In particular, since $\dot\ell_j^\pm$ are constant on $B_3(\varepsilon)$, the semicontinuous hulls of $\mathcal{L}_2$ are

$$\mathcal{L}_{2,\wedge}(x,y) := \sup_{\varepsilon>0} \inf \left\{ \mathcal{L}_2(x',y') : (x',y') \in B(\varepsilon) \right\} = \mathcal{L}_2(x,y) \wedge \mathcal{L}_3(x,y),$$

$$\mathcal{L}_{2,\vee}(x,y) := \inf_{\varepsilon>0} \sup \left\{ \mathcal{L}_2(x',y') : (x',y') \in B(\varepsilon) \right\} = \mathcal{L}_2(x,y) \vee \mathcal{L}_3(x,y),$$

and since $B_1(\varepsilon) \cap \mathcal{S}$ and $B_2(\varepsilon) \cap \mathcal{S}$ are always nonempty, the preceding relations also hold if $B(\varepsilon)$ is intersected with $\mathcal{S}$, whence

$$\mathcal{L}_1(x,y) = \sup_{\varepsilon>0} \inf \left\{ \mathcal{L}_2(x',y') : (x',y') \in B(\varepsilon) \cap \mathcal{S} \right\} = \mathcal{L}_{2,\wedge}(x,y).$$

One easily argues that $\mathcal{L}_1$ is lower semicontinuous, i.e. its lower semicontinuous hull is equal to $\mathcal{L}_1$ itself, which is also equal to the lower semicontinuous hull of $\mathcal{L}_2$.

Next observe that

$$\begin{aligned} \mathcal{L}_{1,\vee}(x,y) &= \inf_{\varepsilon>0} \sup \left\{ \mathcal{L}_2(x',y') \wedge \mathcal{L}_3(x',y') : (x',y') \in B(\varepsilon) \right\} \\ &= \mathcal{L}_2(x,y) \vee \mathcal{L}_3(x,y) = \mathcal{L}_{2,\vee}(x,y). \end{aligned}$$

where the first equality follows from the definition of $\mathcal{L}_{1,\vee}$, the fact that $\mathcal{L}_1 = \mathcal{L}_{2,\wedge}$ as shown earlier and the representation for $\mathcal{L}_{2,\wedge}$ derived above while the second equality follows from Equations (S2.11) and (S2.12).

The previous argument assumes $(x,y) \in (0,\infty)^2$. It remains to show that the semicontinuous hulls of $\mathcal{L}_1$ also correspond to those of $\mathcal{L}_2$ on the axes. For this, assume now that $x > 0, y = 0$. The ball $B(\varepsilon)$ around $(x,0)$ now becomes a "half-ball" (we intersect if with $[0,\infty)^2$). Let (x',y') be a point in that ball. Then (x',y') is on the same ray as (x,δ), for some $\delta \geq 0$ that can be made to converge to 0 as $\varepsilon \to 0$. We have $\dot\ell_2^\pm(x',y') = \dot\ell_2^\pm(x,\delta) \to \dot\ell_2^+(x,0)$ as $\varepsilon \to 0$. For the first derivative, the known bounds $x \vee y \leq \ell(x,y) \leq x + y$ imply that $x \leq \ell(x,\delta) \leq x + \delta$. The convexity and homogeneity properties then imply that

$$\dot\ell_1(x,0) = 1 \geq \dot\ell_1^\pm(x,\delta) \geq \frac{\ell(x,\delta) - \ell(0,\delta)}{x} \geq \frac{x-\delta}{x} \longrightarrow 1$$

as $\varepsilon \to 0$. By uniform boundedness of $\dot\ell_1^\pm, \dot\ell_2^\pm$ it follows easily that $\mathcal{L}_1$ and $\mathcal{L}_2$ are continuous at $(x,0)$ and that $\mathcal{L}_1(x,0) = \mathcal{L}_2(x,0) = b^{(1)}(x)$, whence those two functions have the same semicontinuous hulls at that point.

Because $\dot\ell_1(0,y)$ was defined as the partial derivative from the right, one deals with a point $(0,y)$ in the same way.

Finally, note that since $b^{(1)}(0) = b^{(2)}(0) = 0$, and by uniform boundedness of $\dot\ell_1^\pm, \dot\ell_2^\pm$ the functions $\mathcal{L}_1$ and $\mathcal{L}_2$ are both continuous and take the value 0 at $(0,0)$. Their semicontinuous hulls are therefore also equal at that point.

We have shown that everywhere on $[0,\infty)^2$, $\mathcal{L}_{1,\wedge} = \mathcal{L}_{2,\wedge}$ and $\mathcal{L}_{1,\vee} = \mathcal{L}_{2,\vee}$. By definition (see Bücher, Segers and Volgushev, 2014, Proposition 2.1), this means that $d_{\text{hypi}}(\mathcal{L}_1, \mathcal{L}_2) = 0$. □

LEMMA S10. *Let $f : [0,T]^2 \to \mathbb{R}$ be continuous Lebesgue-almost everywhere, $g := (g_1, \ldots, g_q)^\top : [0,T]^2 \to \mathbb{R}^q$ be a vector of integrable functions and assume that f_n are measurable and hypi-converge to f on $[0,T]^2$. Then $\int g f_n d\mu_L \to \int g f d\mu_L$, where μ_L denotes the Lebesgue measure on $[0,T]^2$.*

PROOF. For every $j \in \{1, \ldots, q\}$ and $M < \infty$, we have

$$\begin{aligned}\int |g_j f_n - g_j f| d\mu_{\mathrm{L}} &= \int |g_j||f_n - f| \mathbb{1}\{|g_j| \leq M\} \, d\mu_{\mathrm{L}} + \int |g_j||f_n - f| \mathbb{1}\{|g_j| > M\} \, d\mu_{\mathrm{L}} \\ &\leq M \int |f_n - f| d\mu_{\mathrm{L}} + \sup_{(x,y)\in[0,T]^2} |f_n(x,y) - f(x,y)| \int |g_j| \mathbb{1}\{|g_j| > M\} \, d\mu_{\mathrm{L}} \\ &\leq M \int |f_n - f| d\mu_{\mathrm{L}} \\ &\quad + \left(\sup_{(x,y)\in[0,T]^2} |f_n(x,y)| + \sup_{(x,y)\in[0,T]^2} |f(x,y)| \right) \int |g_j| \mathbb{1}\{|g_j| > M\} \, d\mu_{\mathrm{L}}.\end{aligned}$$

The first term on the right hand side converges to 0 by Proposition 2.4 of Bücher, Segers and Volgushev (2014) and since f is assumed continuous almost everywhere. By Proposition 2.3 of that paper, $\sup_{(x,y)\in[0,T]^2} |f_n(x,y)| \to \sup_{(x,y)\in[0,T]^2} |f(x,y)|$. Therefore, we have

$$\lim_{n\to\infty} \int |g_j f_n - g_j f| d\mu_{\mathrm{L}} \leq 2 \sup_{(x,y)\in[0,T]^2} |f(x,y)| \int |g_j| \mathbb{1}\{|g_j| > M\} \, d\mu_{\mathrm{L}},$$

which can be made arbitrarily small by choosing M large enough, since g_j is integrable. The claim follows.

□

LEMMA S11. *Let $\phi : \mathbb{R}^p \to \mathbb{R}^q$, $p \leq q$, have a unique, well separated zero at a point $x_0 \in \mathbb{R}^p$ and be continuously differentiable at x_0 with Jacobian matrix $J := J_\phi(x_0)$ of full rank p. Let Y_n be a random vector in $\mathbb{R}^q$ such that $a_n^{-1} Y_n$ weakly converges to a random vector Y, for some sequence $a_n \to 0$. Then if $X_n = \arg\min_x \|\phi(x) - Y_n\|$, we have*

$$X_n - x_0 = (J^\top J)^{-1} J^\top Y_n + o_P(a_n).$$

PROOF. Let $h_n := a_n^{-1}(X_n - x_0 - (J^\top J)^{-1} J^\top Y_n)$. By definition of X_n, h_n is a minimizer of the random function $M_n : \mathbb{R}^p \to \mathbb{R}_+$ defined as

$$M_n(h) := a_n^{-1} \Big\| \phi\big(x_0 + (J^\top J)^{-1} J^\top Y_n + a_n h\big) - Y_n \Big\|.$$

By differentiability of ϕ, $M_n(h)$ is the norm of

$$\big(J(J^\top J)^{-1} J^\top - I\big) a_n^{-1} Y_n + Jh + o(1)$$

uniformly over bounded h, where I is the $q \times q$ identity matrix. The above display, seen as a function of h, weakly converges to

$$\big(J(J^\top J)^{-1} J^\top - I\big) Y + Jh$$

in $(\ell^\infty(\mathcal{K}))^q$, for any compact set $\mathcal{K}$. The mapping $f \mapsto \{h \mapsto \|f(h)\|\}$ being continuous from $(\ell^\infty(\mathcal{K}))^q$ onto $\ell^\infty(\mathcal{K})$, it follows that $M_n \rightsquigarrow M$ in $\ell^\infty(\mathcal{K})$, for

$$M(h) := \Big\| \big(J(J^\top J)^{-1} J^\top - I\big) Y + Jh \Big\|.$$

The function M^2 is strictly convex and has derivative $\partial(M^2(h))/\partial h = 2J^\top Jh$ which, since J has full rank, has a unique zero at $h=0$. It follows that M^2, and thus M, has a unique minimizer at the point 0. Therefore, if we can show that the sequence $\{h_n\}$ is uniformly tight, Corollary 5.58 of van der Vaart (2000) will ensure that h_n converges in distribution (and hence in probability) to 0, which in turn implies the result.

It is known by Prohorov's theorem that $\{a_n^{-1}Y_n\}$ is uniformly tight. Therefore, it is sufficient to establish tightness of $\{a_n^{-1}(X_n - x_0)\}$. First, define for $\delta > 0$

$$\varepsilon(\delta) = \inf_{x \notin B(x_0,\delta)} \|\phi(x)\|,$$

where $B(x_0,\delta)$ denotes an open δ-ball around x_0. By assumption, $\varepsilon(\delta) > 0$ for every positive δ. Choose $\delta_0 > 0$ small enough so that for every $x \in B(x_0,\delta_0)$,

$$\|\phi(x) - J(x - x_0)\| < \frac{1}{2}\|J(x - x_0)\|,$$

which is possible by differentiability of ϕ (recall that J is the Jacobian at x_0). By the reverse triangle inequality, this implies that $|\|\phi(x)\| - \|J(x-x_0)\||$ has the same upper bound. Then, for $\delta \le \delta_0$,

$$\varepsilon(\delta) > \frac{1}{2} \inf_{x \in B(x_0,\delta)} \|J(x - x_0)\| = \frac{\sigma_1(J)}{2}\delta,$$

where $\sigma_1(J)$, the smallest singular value of J, is positive since J has full rank.

Now, fix an arbitrary $\eta > 0$. Because the sequence $\{a_n^{-1}Y_n\}$ is uniformly tight, there exists a finite $K = K(\eta)$ such that for $\delta_n := Ka_n$ and for n large enough so that $\delta_n \le \delta_0$,

$$\mathbb{P}\left(\|Y_n\| \ge \frac{\varepsilon(\delta_n)}{2}\right) \le \mathbb{P}\left(\|Y_n\| \ge \frac{K\sigma_1(J)}{4}a_n\right) \le \eta$$

Hence with probability at least $1-\eta$, $\|Y_n\| < \varepsilon(\delta_n)/2$. The last inequality implies two things. First, letting $\phi_n = \phi - Y_n$ and recalling that $\phi(x_0) = 0$, we have $\|\phi_n(x_0)\| = \|Y_n\| < \varepsilon(\delta_n)/2$. Second, for any $x \notin B(x_0,\delta_n)$, we have $\|\phi(x)\| \ge \varepsilon(\delta_n)$ so

$$\|\phi_n(x)\| = \|\phi(x) - Y_n\| \ge |\|\phi(x)\| - \|Y_n\|| > \frac{\varepsilon(\delta_n)}{2}.$$

That is, with probability at least $1-\eta$, $X_n = \arg\min_x \|\phi_n(x)\| \in B(x_0,\delta_n)$. Since $\delta_n = O(a_n)$ and η was arbitrary, we conclude that $\{a_n^{-1}(X_n - x_0)\}$ is uniformly tight, and so is $\{h_n\}$.

□

S3. Proof of the claims in Examples 8, 11 and 12.

S3.1. *Example 8*. Recall that the random vector $Z := (1-X, 1-Y)$ is assumed max-stable with uniform margin and stable tail dependence function ℓ, hence its distribution function is given by Equation (2.1). Let $(x,y) \in (0,1]^2$ (the result is trivial if x or y is zero). Note that we can without loss of generality focus on $(x,y) \in (0,1]^2$ instead of general bounded sets since any bounded set can be rescaled to be contained in $[0,1]^2$ at the cost of absorbing the scaling into t. The survival copula Q of (X,Y) satisfies

$$\begin{aligned}
Q(tx,ty) &:= \mathbb{P}\left(X \ge 1 - tx, Y \ge 1 - ty\right) \\
&= \mathbb{P}\left(1 - X \le tx, 1 - Y \le ty\right) \\
&= \exp\{-\ell(-\log(tx), -\log(ty))\} \\
&= \exp\left\{\log(t)\ell\left(1 + \frac{\log(x)}{\log(t)}, 1 + \frac{\log(y)}{\log(t)}\right)\right\},
\end{aligned}$$

where we have used the homogeneity property of ℓ in the last line. By the assumed expansion of the function ℓ,

$$\ell\left(1+\frac{\log(x)}{\log(t)}, 1+\frac{\log(y)}{\log(t)}\right) = \ell(1,1) + \dot{\ell}_1(1,1)\frac{\log(x)}{\log(t)} + \dot{\ell}_2(1,1)\frac{\log(y)}{\log(t)} + \delta(t,x,y),$$

where $\dot{\ell}_1$ and $\dot{\ell}_2$ are the right partial derivatives of ℓ with respect to its first and second argument, respectively, and

$$\delta(t,x,y) \lesssim \left(\frac{\log(x)}{\log(t)}\right)^2 + \left(\frac{\log(y)}{\log(t)}\right)^2.$$

This is a linear approximation of the function ℓ; since that function is convex, it lies above its sub gradient, so the error term $\delta(t,x,y)$ is non-negative. Plugging this in our expression for $Q(tx,ty)$ yields

$$Q(tx,ty) = t^{\ell(1,1)} x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)} e^{\delta'(t,x,y)},$$

where $\delta'(t,x,y) = \log(t)\delta(t,x,y)$ satisfies

$$\frac{\log(x)^2 + \log(y)^2}{\log(t)} \lesssim \delta'(t,x,y) \leq 0.$$

Letting $q(t) = t^{\ell(1,1)}$ and $c(x,y) = x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)}$, we obtain

$$\begin{aligned}\left|\frac{Q(tx,ty)}{q(t)} - c(x,y)\right| &= x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)}\left(1 - e^{\delta'(t,x,y)}\right)\\ &\leq x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)} |\delta'(t,x,y)|\\ &\lesssim \frac{x^{\dot{\ell}_1(1,1)} y^{\dot{\ell}_2(1,1)}(\log(x)^2 + \log(y)^2)}{\log(1/t)},\end{aligned}$$

where we used the fact that $0 \leq 1 - e^x \leq |x|$ for all $x \leq 0$. Since $\dot{\ell}_1(1,1)$ and $\dot{\ell}_2(1,1)$ are positive it follows that this upper bound is of order $1/\log(1/t)$ uniformly over x, y in bounded sets. The claim in Example 8 is proved. □

S3.2. *Example 11*. Now, recall the setting of Example 11. The expression for $\Gamma^{(s,s)}$ is trivial. We shall treat the case where s and s' are two pairs that share an element, i.e. $s = (s_1, s_2)$ and $s' = (s_1, s_3)$. One similarly deals with different combinations of s, s', including the case where they are disjoint.

Let ℓ be the stable tail dependence function of the max-stable, trivariate random vector $(1 - X^{(s_1)}, 1 - X^{(s_2)}, 1 - X^{(s_3)})$. By assumption and by the calculations above for the bivariate case, the pairs $(X^{(s_1)}, X^{(s_2)})$ and $(X^{(s_1)}, X^{(s_3)})$ satisfy Condition 1(i) with scaling functions $q^{(s)}(t) = t^{\ell(1,1,0)}$ and $q^{(s')}(t) = t^{\ell(1,0,1)}$, respectively. Since those functions are invertible, we may choose any diverging sequence $m = o(\log(n)^2)$ and invert them, setting $k^{(s)}/n = (m/n)^{1/\ell(1,1,0)}$ and $k^{(s')}/n = (m/n)^{1/\ell(1,0,1)}$. In fact, we may do so with every pair and obtain, as claimed, a universal sequence m.

Without loss of generality, let us assume that $\ell(1,1,0) \leq \ell(1,0,1)$ so that $k^{(s)} \leq k^{(s')}$. Let $t_n = k^{(s)}/n$ and $\alpha = \ell(1,1,0)/\ell(1,0,1) \in (0,1]$; observe that $k^{(s')}/n = t_n^{\alpha}$. By definition, for fixed $x^1, x^2 \in (0,1]^2$ (we can restrict our attention to this setting by similar arguments as in the bivariate case), we have

$$\Gamma^{(s,s')}(x^1,x^2) = \lim_{n\to\infty} \frac{n}{m}\mathbb{P}\left(1 - X^{(s_1)} \leq t_n x, 1 - X^{(s_2)} \leq t_n y, 1 - X^{(s_3)} \leq t_n^{\alpha} z\right), \tag{S3.1}$$

where x is equal to $x_1^1 \wedge x_2^1$ if $\alpha = 1$ and to x_1^1 otherwise, $y = x_2^1$ and $z = x_2^2$. Using the same reasoning as in the bivariate case above (including the homogeneity property of ℓ), the probability in Equation (S3.1) can be written as

$$\begin{aligned}
&\exp\{-\ell(-\log(t_n x), -\log(t_n y), -\log(t_n^\alpha z))\} \\
&\quad = \exp\left\{\log(t_n)\ell\left(1 + \frac{\log(x)}{\log(t_n)}, 1 + \frac{\log(y)}{\log(t_n)}, \alpha + \frac{\log(z)}{\log(t_n)}\right)\right\} \\
&\quad = t_n^{\ell(1,1,\alpha)} \exp\left\{\log(t_n)\left[\ell\left(1 + \frac{\log(x)}{\log(t_n)}, 1 + \frac{\log(y)}{\log(t_n)}, \alpha + \frac{\log(z)}{\log(t_n)}\right) - \ell(1,1,\alpha)\right]\right\}.
\end{aligned}$$

Eventually, $\log(t_n)$ is negative, which makes the difference in the square brackets nonnegative by monotonicity of ℓ. This eventually upper bounds the exponential by 1 and the entire expression by $t_n^{\ell(1,1,\alpha)}$, for any $x, y, z \in (0, 1]$. Considering Equation (S3.1), it follows that for every fixed $x^1, x^2 \in (0, 1]^2$,

$$\Gamma^{(s,s')}(x^1, x^2) \le \lim_{n\to\infty} \frac{n}{m} t_n^{\ell(1,1,\alpha)} = \lim_{n\to\infty} \left(\frac{m}{n}\right)^{\frac{\ell(1,1,\alpha)}{\ell(1,1,0)} - 1} = 0,$$

since the assumption that ℓ is component-wise strictly increasing means that $\ell(1,1,\alpha) > \ell(1,1,0)$. □

S3.3. *Example 12.* We present here the main ideas, as most of the precise calculations are similar to the preceding section. As before, let $X^{(j)} = Y(u_j)$, and write $Z^{(j)}$ and $Z'^{(j)}$ for $Z(u_j)$ and $Z'(u_j)$. Consider a pair $s := (s_1, s_2)$ and let F be the distribution function of the unit Fréchet distribution. Recall that $X^{(j)} = \max\{aZ^{(j)}, (1-a)Z'^{(j)}\}$. We have for $t \downarrow 0$

$$\begin{aligned}
&\mathbb{P}\left(F(X^{(s_1)}) \ge 1 - tx, F(X^{(s_2)}) \ge 1 - ty\right) \\
&= \mathbb{P}\left(F(Z^{(s_1)})^{1/a} \vee F(Z'^{(s_1)})^{1/(1-a)} \ge 1 - tx, F(Z^{(s_2)})^{1/a} \vee F(Z'^{(s_2)})^{1/(1-a)} \ge 1 - ty\right) \\
&= \mathbb{P}\left(F(Z^{(s_1)}) \ge (1-tx)^a, F(Z^{(s_2)}) \ge (1-ty)^a\right) \\
&\qquad + \mathbb{P}\left(F(Z'^{(s_1)}) \ge (1-tx)^{1-a}, F(Z'^{(s_2)}) \ge (1-ty)^{1-a}\right) + O(t^2),
\end{aligned} \tag{S3.2}$$

where the term $O(t^2)$ is uniform over bounded x, y. Note that $(1-tx)^a = 1 - t(ax + O(tx^2))$. The first term of Equation (S3.2) is equal to

$$a\chi^{Z,(s)} t(x + y - \ell^{Z,(s)}(x, y)) + O(t^2)$$

uniformly over bounded x, y, where $\chi^{Z,(s)}$ and $\ell^{Z,(s)}$ are the extremal dependence coefficient and stable tail dependence function, respectively, corresponding to the random vector $(Z^{(s_1)}, Z^{(s_2)})$. From previous calculations, the second term of Equation (S3.2) is equal to

$$((1-a)t)^{\ell^{Z',(s)}(1,1)} x^{\dot{\ell}_1^{Z',(s)}(1,1)} y^{\dot{\ell}_2^{Z',(s)}(1,1)} + O\left(t^{\ell^{Z',(s)}(1,1)}/\log(1/t)\right),$$

uniformly over bounded x, y, where $\ell^{Z',(s)}$ is the stable tail dependence function corresponding to the max-stable random vector $(1/Z'^{(s_1)}, 1/Z'^{(s_2)})$. It follows that Condition 1(i) is satisfied for every pair of locations; depending on whether $(Z^{(s_1)}, Z^{(s_2)})$ is dependent or independent, either the first of the second of the last two expressions dominates. This determines that $q^{(s)}(t)$ is proportional to t for asymptotically dependent pairs and to $t^{1/\eta'^{(s)}}$ for asymptotically independent ones, where $\eta'^{(s)}$ is the coefficient of tail dependence of

$(1/Z'^{(s_1)}, 1/Z'^{(s_2)})$, satisfying $1 < 1/\eta'^{(s)} < 2$ by assumption — for any inverted max-stable distribution, its coefficient of tail dependence η is always in $[1/2, 1)$, and can only be equal to $1/2$ under perfect independence. The coefficient of tail dependence $\eta^{(s)}$ of $(X^{(s_1)}, X^{(s_2)})$ is equal to 1 if $\chi^{Z,(s)} > 0$ and to $\eta'^{(s)}$ otherwise.

We now show how to obtain an expression for the functions $\Gamma^{(s,s')}$. First, since the functions $q^{(s)}$ are proportional to simple powers, for a sufficiently slow intermediate sequence m, we let $k^{(s)}/n$ be proportional to m/n if s is an asymptotically dependent pair and to $(m/n)^{\eta^{(s)}}$ otherwise, so that all $m^{(s)}$ are equal to m.

The case $s = s'$ follows trivially from the previous developments; $\Gamma^{(s,s)}$ can be derived from $c^{(s)}$. Next consider the case where s, s' share one element, i.e. $s = (s_1, s_2)$ and $s' = (s_1, s_3)$. Letting $t_n = k^{(s)}/n$ and $t'_n = k^{(s')}/n$, assume without loss of generality that $t'_n \lesssim t_n$. The probability of interest is of the form

$$
\begin{aligned}
&\mathbb{P}\left(F(X^{(s_1)}) \geq 1 - (t_n x \wedge t'_n x'), F(X^{(s_2)}) \geq 1 - t_n y, F(X^{(s_3)}) \geq 1 - t'_n z\right) \\
&= \mathbb{P}\left(F(Z^{(s_1)}) \geq (1 - (t_n x \wedge t'_n x'))^a, F(Z^{(s_2)}) \geq (1 - t_n y)^a, F(Z^{(s_3)}) \geq (1 - t'_n z)^a\right) \\
&\quad + \mathbb{P}\Big(F(Z'^{(s_1)}) \geq (1 - (t_n x \wedge t'_n x'))^{1-a}, F(Z'^{(s_2)}) \geq (1 - t_n y)^{1-a}, \\
&\qquad F(Z'^{(s_3)}) \geq (1 - t'_n z)^{1-a}\Big) + O(t_n^2).
\end{aligned}
$$

Indeed, the third term above is the probability of a certain event that requires at least one of the Z and one of the Z' to be large, which has probability at most $O(t_n^2)$ since Z and Z' are assumed independent (recall that we assumed $t'_n = O(t_n)$). We note that the term in front of this probability in the definition of $\Gamma^{(s,s')}$ is equal to $q^{(s)}(t_n)^{-1} = t_n^{-1/\eta^{(s)}}$. However $t_n^2 = o(t_n^{1/\eta^{(s)}})$ since $\eta^{(s)} > 1/2$, and the second probability above is also $o(t_n^{1/\eta^{(s)}})$, following the calculations for Example 11. Therefore, in this case, $\Gamma^{(s,s')}((x,y),(x',z))$ is equal to the limit

$$
\begin{aligned}
&\lim_{n\to\infty} t_n^{-1/\eta^{(s)}} \mathbb{P}\Big(F(Z^{(s_1)}) \geq (1 - (t_n x \wedge t'_n x'))^a, \\
&\qquad F(Z^{(s_2)}) \geq (1 - t_n y)^a, F(Z^{(s_3)}) \geq (1 - t'_n z)^a\Big) \\
&= \lim_{n\to\infty} t_n^{-1/\eta^{(s)}} \mathbb{P}\Big(F(Z^{(s_1)}) \geq 1 - a(t_n x \wedge t'_n x'), \\
&\qquad F(Z^{(s_2)}) \geq 1 - a t_n y, F(Z^{(s_3)}) \geq 1 - a t'_n z\Big)
\end{aligned}
$$

which is non-zero if and only if $(Z^{(s_1)}, Z^{(s_2)}, Z^{(s_3)})$ is fully dependent (i.e., it contains no pairwise independence).

For the case where the pairs $s = (s_1, s_2)$ and $s' = (s_3, s_4)$ are disjoint, let $t_n = k^{(s)}/n$ and $t'_n = k^{(s')}/n$ and assume as before that $t'_n \lesssim t_n$. By similar arguments as above, one obtains that $\Gamma^{(s,s')}((x,y),(x',y'))$ is equal to the limit

$$
\begin{aligned}
&\lim_{n\to\infty} t_n^{-1/\eta^{(s)}} \mathbb{P}\Big(F(Z^{(s_1)}) \geq 1 - a t_n x, F(Z^{(s_2)}) \geq 1 - a t_n y, \\
&\qquad F(Z^{(s_3)}) \geq 1 - a t'_n x', F(Z^{(s_4)}) \geq 1 - a t'_n y'\Big),
\end{aligned}
$$

which is non-zero if and only if $(Z^{(s_1)}, Z^{(s_2)}, Z^{(s_3)}, Z^{(s_4)})$ has no independent pairs.

Using the same ideas and after straightforward computations, one may calculate the limits $\Gamma^{(s,s',j)}$, for $s' \in \mathcal{P}_D$. First, consider the case where $s = (s_1, s_2)$ and $s'_j = s_1$, that is the element s'_j is in the pair s. Defining t_n and t'_n as above, we still have $t'_n \lesssim t_n$ since s' is an asymptotically dependent pair. Then $\Gamma^{(s,s',j)}((x,y),(x',y'))$ is equal to

$$\chi^{Z,(s')} \lim_{n\to\infty} t_n^{-1/\eta^{(s)}} \mathbb{P}\left(F(Z^{(s_1)}) \geq 1 - a(t_n x \wedge t'_n x'), F(Z^{(s_2)}) \geq 1 - a t_n y\right),$$

which is non-zero if and only if $(Z^{(s_1)}, Z^{(s_2)})$ is dependent. Now if $s_3 := s'_j$ is not an element of s, $\Gamma^{(s,s',j)}((x,y),(x',y'))$ becomes

$$\chi^{Z,(s')} \lim_{n\to\infty} t_n^{-1/\eta^{(s)}} \mathbb{P}\left(F(Z^{(s_1)}) \geq 1 - a t_n x, F(Z^{(s_2)}) \geq 1 - a t_n y, F(Z^{(s_3)}) \geq 1 - a t'_n x'\right),$$

which is non-zero if and only if $(Z^{(s_1)}, Z^{(s_2)}, Z^{(s_3)})$ is fully dependent.

Finally, for $s, s' \in \mathcal{P}_D$, again letting $t_n = k^{(s)}/n$ and $t'_n = k^{(s')}/n$, note that this time t'_n/t_n is constant. Without loss of generality, let $j = j' = 1$. Then $\Gamma^{(s,j,s',j')}((x,y),(x',y'))$ is equal to

$$\chi^{Z,(s)} \chi^{Z,(s')} \lim_{n\to\infty} t_n^{-1} \mathbb{P}\left(F(Z^{(s_1)}) \geq 1 - t_n x, F(Z^{(s'_1)}) \geq 1 - t'_n y'\right),$$

which is non-zero if and only if $(Z^{(s_1)}, Z^{(s'_1)})$ is dependent. □

S4. Proof of the claims in Example 9. The multiplicative constant appearing in the scaling function q, as a function of λ, is given by

$$K_\lambda = \begin{cases} 2\frac{1-\lambda}{2-\lambda}, & \lambda \in (0,1) \\ 2, & \lambda = 1 \\ \left(1 - \frac{1}{\lambda}\right)^{\lambda - 1} \frac{2(\lambda-1)}{\lambda(2-\lambda)}, & \lambda \in (1,2) \\ \frac{1}{2}, & \lambda = 2 \\ \frac{\left(1-\frac{1}{\lambda}\right)^2}{1-\frac{2}{\lambda}}, & \lambda \in (2,\infty) \end{cases}; \tag{S4.1}$$

it can be deduced from the proof.

The argument must be separated in two cases depending on whether $\lambda = 1$.

S4.1. *The case* $\lambda \neq 1$. For now, assume that $\alpha_R \neq \alpha_W$. Let $\bar{F}_R$ denote the survival function of R. Then $\bar{F}_R(x) = x^{-\alpha_R}$ for $x > 1$, and $\bar{F}_R(x) = 1$ for $x \leq 1$. The first step in calculationg Q is to find an expression for the survival function $\bar{F}$ of X (and equivalently of Y) and its inverse. We have, for $x \geq 1$,

$$\begin{aligned} \bar{F}(x) &= \mathbb{P}(RW_1 > x) \\ &= \mathbb{P}\left(R > \frac{x}{W_1}\right) \\ &= \mathbb{E}\left[\bar{F}_R\left(\frac{x}{W_1}\right)\right] \\ &= \mathbb{P}(W_1 > x) + \int_1^x \left(\frac{w}{x}\right)^{\alpha_R} \frac{\alpha_W}{w^{\alpha_W+1}} dw \\ &= x^{-\alpha_W} + \alpha_W x^{-\alpha_R} \frac{x^{\alpha_R - \alpha_W} - 1}{\alpha_R - \alpha_W} \\ &= \frac{\alpha_R}{\alpha_R - \alpha_W} x^{-\alpha_W} - \frac{\alpha_W}{\alpha_R - \alpha_W} x^{-\alpha_R} \\ &= \frac{\alpha_\vee}{\alpha_\vee - \alpha_\wedge} x^{-\alpha_\wedge}\left(1 - \frac{\alpha_\wedge}{\alpha_\vee - \alpha_\wedge} x^{\alpha_\wedge - \alpha_\vee}\right), \end{aligned}$$

where $\alpha_\wedge$ and $\alpha_\vee$ denote the smallest and the largest of the two α's, respectively. Although not easily invertible, this function is close to $\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}x^{-\alpha_\wedge}$, which has an analytical inverse. We now argue that this inverse is close to that of $\bar{F}$. First, for any $X \in (1,\infty)$, we have for $x \in [X,\infty)$

$$\underbrace{\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}x^{-\alpha_\wedge}\left(1-\frac{\alpha_\wedge}{\alpha_\vee-\alpha_\wedge}X^{\alpha_\wedge-\alpha_\vee}\right)}_{f_1(x)} \leq \bar{F}(x) \leq \underbrace{\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}x^{-\alpha_\wedge}}_{f_2(x)}.$$

Now note that for two decreasing, invertible functions g_1 and g_2, $g_1 \leq g_2$ is equivalent to $g_1^{-1} \leq g_2^{-1}$. This means that as soon as $y \leq f_1(X)$, $f_1^{-1}(y) \leq \bar{F}^{-1}(y) \leq f_2^{-1}(y)$. In other words, for such y,

$$\left(1-\frac{\alpha_\wedge}{\alpha_\vee-\alpha_\wedge}X^{\alpha_\wedge-\alpha_\vee}\right)^{1/\alpha_\wedge}\left(\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}\right)^{1/\alpha_\wedge} y^{-1/\alpha_\wedge} \leq \bar{F}^{-1}(y) \leq \left(\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}\right)^{1/\alpha_\wedge} y^{-1/\alpha_\wedge}.$$

Because these inequalities are true as soon as $y \leq f_1(X)$, they are true if $y = f_1(X)$. If y is small enough, choosing $X = \left(\frac{1}{2}\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}\right)^{1/\alpha_\wedge} y^{-1/\alpha_\wedge}$ is sufficient to have $y \leq f_1(X)$. Therefore, if y is small enough, the first inequality in the last display becomes

$$\bar{F}^{-1}(y) \geq \left(1-O\left(y^{\frac{\alpha_\vee}{\alpha_\wedge}-1}\right)\right)\left(\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}\right)^{1/\alpha_\wedge} y^{-1/\alpha_\wedge}.$$

Combining this with the upper bound (the second inequality) yields

$$\bar{F}^{-1}(y) = (1+O(y^\tau))\left(\frac{\alpha_\vee}{\alpha_\vee-\alpha_\wedge}\right)^{1/\alpha_\wedge} y^{-1/\alpha_\wedge}, \tag{S4.2}$$

where $\tau = \frac{\alpha_\vee}{\alpha_\wedge} - 1$.

The copula Q can now be expressed as

$$\begin{aligned} Q(tx,ty) &= \mathbb{P}\left(X \geq \bar{F}^{-1}(tx), Y \geq \bar{F}^{-1}(ty)\right) \\ &= \mathbb{P}\left(RW_1 \geq \bar{F}^{-1}(tx), RW_2 \geq \bar{F}^{-1}(ty)\right) = \mathbb{P}(R \geq Z) = \mathbb{E}\left[\bar{F}_R(Z)\right], \end{aligned}$$

where

$$Z := Z(tx,ty) = \frac{\bar{F}^{-1}(tx)}{W_1} \vee \frac{\bar{F}^{-1}(ty)}{W_2}.$$

Recalling the definition of $\bar{F}_R$, we have

$$\begin{aligned} Q(tx,ty) &= \mathbb{P}(Z \leq 1) + \mathbb{E}\left[Z^{-\alpha_R}; Z > 1\right] \\ &= \mathbb{P}(Z \leq 1) + \int_0^\infty \mathbb{P}\left(Z^{-\alpha_R} > a, Z > 1\right) da \\ &= \mathbb{P}(Z \leq 1) + \int_0^\infty \mathbb{P}\left(1 < Z \leq a^{-1/\alpha_R}\right) da \\ &= \mathbb{P}(Z \leq 1) + \int_0^1 \mathbb{P}\left(1 < Z \leq a^{-1/\alpha_R}\right) da \\ &= \mathbb{P}(Z \leq 1) + \int_0^1 \left(\mathbb{P}\left(Z \leq a^{-1/\alpha_R}\right) - \mathbb{P}(Z \leq 1)\right) da \\ &= \int_0^1 \mathbb{P}\left(Z \leq a^{-1/\alpha_R}\right) da. \end{aligned}$$

In order to compute the previous integral, we need to derive the CDF of Z. From the definition of Z and by independence of W_1 and W_2, it is clear that, for any $z > 0$,

$$\mathbb{P}(Z \le z) = \mathbb{P}\left(W_1 \ge \frac{\bar{F}^{-1}(tx)}{z}\right)\mathbb{P}\left(W_2 \ge \frac{\bar{F}^{-1}(ty)}{z}\right).$$

From now on, assume without loss of generality that $x \ge y$ since $c(x,y) = c(y,x)$ (because the random variables X and Y are exchangeable). Then $\bar{F}^{-1}(tx) \le \bar{F}^{-1}(ty)$. The previous probability can take 3 different forms:

$$\mathbb{P}(Z \le z) = \begin{cases} \left(\bar{F}^{-1}(tx)\bar{F}^{-1}(ty)\right)^{-\alpha_W} z^{2\alpha_W}, & \text{if } z \le \bar{F}^{-1}(tx) \\ \left(\bar{F}^{-1}(ty)\right)^{-\alpha_W} z^{\alpha_W}, & \text{if } \bar{F}^{-1}(tx) < z \le \bar{F}^{-1}(ty) \\ 1, & \text{if } z > \bar{F}^{-1}(ty) \end{cases}.$$

When substituting $z = a^{-1/\alpha_R}$, for $a \in (0,1)$, notice that we are in the three preceding cases, respectively, when

$$\begin{cases} a \ge \left(\bar{F}^{-1}(tx)\right)^{-\alpha_R} \\ \left(\bar{F}^{-1}(ty)\right)^{-\alpha_R} \le a < \left(\bar{F}^{-1}(tx)\right)^{-\alpha_R} \\ a < \left(\bar{F}^{-1}(ty)\right)^{-\alpha_R} \end{cases}.$$

This allows us to write

$$Q(tx,ty) = \int_0^{\left(\bar{F}^{-1}(ty)\right)^{-\alpha_R}} da + \left(\bar{F}^{-1}(ty)\right)^{-\alpha_W}\int_{\left(\bar{F}^{-1}(ty)\right)^{-\alpha_R}}^{\left(\bar{F}^{-1}(tx)\right)^{-\alpha_R}} a^{-\frac{\alpha_W}{\alpha_R}}\,da$$

$$+ \left(\bar{F}^{-1}(tx)\bar{F}^{-1}(ty)\right)^{-\alpha_W}\int_{\left(\bar{F}^{-1}(tx)\right)^{-\alpha_R}}^{1} a^{-2\frac{\alpha_W}{\alpha_R}}\,da. \tag{S4.3}$$

Since we only need Equation (3.1) to hold uniformly over a sphere, we may assume that $y \le x \le 1$. Then, Equation (S4.2) yields

$$\bar{F}^{-1}(tx) = (1 + O(t^\tau))\left(\frac{\alpha_\vee}{\alpha_\vee - \alpha_\wedge}\right)^{1/\alpha_\wedge}(tx)^{-1/\alpha_\wedge}$$

and the same for $\bar{F}^{-1}(ty)$. Moreover, the term $O(t^\tau)$ is uniform over all $(x,y) \in [0,1]^2$. The first term in Equation (S4.3) is then equal to

$$\left(\bar{F}^{-1}(ty)\right)^{-\alpha_R} = (1 + O(t^\tau))\left(1 - \frac{\alpha_\wedge}{\alpha_\vee}\right)^{\frac{\alpha_R}{\alpha_\wedge}} t^{\frac{\alpha_R}{\alpha_\wedge}} y^{\frac{\alpha_R}{\alpha_\wedge}} =: Q^{(1)}(tx,ty),$$

the second one is equal to

$$\begin{aligned}
&\left(\bar{F}^{-1}(ty)\right)^{-\alpha_W} \frac{a^{1-\frac{\alpha_W}{\alpha_R}}}{1 - \frac{\alpha_W}{\alpha_R}}\Bigg|_{a=\left(\bar{F}^{-1}(ty)\right)^{-\alpha_R}}^{\left(\bar{F}^{-1}(tx)\right)^{-\alpha_R}} \\
&= \frac{1}{1 - \frac{\alpha_W}{\alpha_R}}\left(\bar{F}^{-1}(ty)\right)^{-\alpha_W}\left(\bar{F}^{-1}(tx)^{-(\alpha_R-\alpha_W)} - \bar{F}^{-1}(ty)^{-(\alpha_R-\alpha_W)}\right) \\
&= (1 + O(t^\tau))\frac{\left(1 - \frac{\alpha_\wedge}{\alpha_\vee}\right)^{\frac{\alpha_R}{\alpha_\wedge}}}{1 - \frac{\alpha_W}{\alpha_R}} t^{\frac{\alpha_R}{\alpha_\wedge}} y^{\frac{\alpha_W}{\alpha_\wedge}}\left(x^{\frac{\alpha_R-\alpha_W}{\alpha_\wedge}} - y^{\frac{\alpha_R-\alpha_W}{\alpha_\wedge}}\right) \\
&=: Q^{(2)}(tx,ty)
\end{aligned}$$

and finally the third one is equal to

$$\begin{aligned}
&\left(\bar F^{-1}(tx)\bar F^{-1}(ty)\right)^{-\alpha_W} \left.\frac{a^{1-2\frac{\alpha_W}{\alpha_R}}}{1-2\frac{\alpha_W}{\alpha_R}}\right|_{a=\left(\bar F^{-1}(tx)\right)^{-\alpha_R}}^{1} \\
&= \frac{1}{1-2\frac{\alpha_W}{\alpha_R}} \left(\bar F^{-1}(tx)\bar F^{-1}(ty)\right)^{-\alpha_W} \left(1-\left(\bar F^{-1}(tx)\right)^{2\alpha_W-\alpha_R}\right) \\
&= (1+O(t^\tau)) \frac{\left(1-\frac{\alpha_\wedge}{\alpha_\vee}\right)^{2\frac{\alpha_W}{\alpha_\wedge}}}{1-2\frac{\alpha_W}{\alpha_R}} t^{2\frac{\alpha_W}{\alpha_\wedge}} (xy)^{\frac{\alpha_W}{\alpha_\wedge}} \left(1-\left(1-\frac{\alpha_\wedge}{\alpha_\vee}\right)^{\frac{\alpha_R-2\alpha_W}{\alpha_\wedge}} (tx)^{\frac{\alpha_R-2\alpha_W}{\alpha_\wedge}}\right) \\
&=: Q^{(3a)}(tx,ty)
\end{aligned}$$

in the case where $\alpha_R \neq 2\alpha_W$, and if $\alpha_R = 2\alpha_W$, it is equal to

$$\begin{aligned}
&-\left(\bar F^{-1}(tx)\bar F^{-1}(ty)\right)^{-\alpha_W} \log\left(\left(\bar F^{-1}(tx)\right)^{-\alpha_R}\right) \\
&= (1+O(t^\tau)) \left(1-\frac{\alpha_\wedge}{\alpha_\vee}\right)^{2\frac{\alpha_W}{\alpha_\wedge}} t^{2\frac{\alpha_W}{\alpha_\wedge}} (xy)^{\frac{\alpha_W}{\alpha_\wedge}} \\
&\quad \times \left(-\log\left((1+O(t^\tau))\left(1-\frac{\alpha_\wedge}{\alpha_\vee}\right)^{\frac{\alpha_R}{\alpha_\wedge}}\right) + \frac{\alpha_R}{\alpha_\wedge}\left(\log(1/x)+\log(1/t)\right)\right) \\
&= \frac{1}{2} t^2 \log(1/t) xy + O(t^2) \\
&=: Q^{(3b)}(tx,ty),
\end{aligned}$$

where the term $O(t^2)$ is uniform over $(x,y) \in [0,1]^2$. We now divide the possible values of $\lambda = \alpha_R/\alpha_W$ in four ranges and determine which of the three terms $Q^{(1)}$, $Q^{(2)}$ or $Q^{(3)}$ dominates.

S4.1.1. $\lambda \in (0,1)$. This is the case where we obtain asymptotic dependence. All three terms are of the order of t, so they all matter. In this case, $\alpha_\wedge = \alpha_R$, $\alpha_\vee = \alpha_W$ and $\tau = 1/\lambda - 1$. Therefore,

$$\begin{aligned}
Q^{(1)}(tx,ty) &= (1+O(t^\tau))\left(1-\frac{\alpha_R}{\alpha_W}\right) ty = (1-\lambda)ty + O\left(t^{1+\tau}\right), \\
Q^{(2)}(tx,ty) &= (1+O(t^\tau)) \frac{1-\frac{\alpha_R}{\alpha_W}}{1-\frac{\alpha_W}{\alpha_R}} ty^{\frac{\alpha_W}{\alpha_R}} \left(x^{1-\frac{\alpha_W}{\alpha_R}} - y^{1-\frac{\alpha_W}{\alpha_R}}\right) \\
&= (1+O(t^\tau)) \frac{\alpha_R}{\alpha_W} t \left(y - x^{1-\frac{\alpha_W}{\alpha_R}} y^{\frac{\alpha_W}{\alpha_R}}\right) \\
&= \lambda t \left(y - x^{1-1/\lambda} y^{1/\lambda}\right) + O\left(t^{1+\tau}\right), \\
Q^{(3a)}(tx,ty) &= (1+O(t^\tau)) \frac{\left(1-\frac{\alpha_R}{\alpha_W}\right)^{2\frac{\alpha_W}{\alpha_R}}}{2\frac{\alpha_W}{\alpha_R}-1} t^{2\frac{\alpha_W}{\alpha_R}} (xy)^{\frac{\alpha_W}{\alpha_R}} \left(\left(1-\frac{\alpha_R}{\alpha_W}\right)^{1-2\frac{\alpha_W}{\alpha_R}} (tx)^{1-2\frac{\alpha_W}{\alpha_R}} - 1\right) \\
&= (1+O(t^\tau)) \frac{1-\frac{\alpha_R}{\alpha_W}}{2\frac{\alpha_W}{\alpha_R}-1} tx^{1-\frac{\alpha_W}{\alpha_R}} y^{\frac{\alpha_W}{\alpha_R}} + O\left(t^{2\frac{\alpha_W}{\alpha_R}}\right)
\end{aligned}$$

$$= \lambda \frac{1-\lambda}{2-\lambda} t x^{1-1/\lambda} y^{1/\lambda} + O\left(t^{1+\tau} + t^{2/\lambda}\right) = \lambda \frac{1-\lambda}{2-\lambda} t x^{1-1/\lambda} y^{1/\lambda} + O\left(t^{1+\tau}\right),$$

where in the last line we have used $1+\tau = \alpha_\vee/\alpha_\wedge = 1/\lambda < 2/\lambda$. Therefore in this case we get

$$\begin{aligned} Q(tx,ty) &= Q^{(1)}(tx,ty) + Q^{(2)}(tx,ty) + Q^{(3a)}(tx,ty) \\ &= (1-\lambda)ty + \lambda t\left(y - x^{1-1/\lambda} y^{1/\lambda}\right) + \lambda \frac{1-\lambda}{2-\lambda} t x^{1-1/\lambda} y^{1/\lambda} + O\left(t^{1+\tau}\right) \\ &= t\left(y + \left(-\lambda + \lambda \frac{1-\lambda}{2-\lambda}\right) x^{1-1/\lambda} y^{1/\lambda}\right) + O\left(t^{1+\tau}\right) \\ &= t\left(y - \frac{\lambda}{2-\lambda} x^{1-1/\lambda} y^{1/\lambda}\right) + O\left(t^{1+\tau}\right). \end{aligned}$$

S4.1.2. $\lambda \in (1,2)$. Here again, all three terms are of the order of t^λ so they all matter. Note that here and in the next two cases, $\alpha_\wedge = \alpha_W$, $\alpha_\vee = \alpha_R$ and $\tau = \lambda - 1$. Through similar calculations as before, we obtain this time

$$Q^{(1)}(tx,ty) = (1+O(t^\tau))\left(1 - \frac{\alpha_W}{\alpha_R}\right)^{\frac{\alpha_R}{\alpha_W}} t^{\frac{\alpha_R}{\alpha_W}} y^{\frac{\alpha_R}{\alpha_W}} = \left(1 - \frac{1}{\lambda}\right)^{\lambda} t^\lambda y^\lambda + O\left(t^{\lambda+\tau}\right),$$

$$\begin{aligned} Q^{(2)}(tx,ty) &= (1+O(t^\tau)) \frac{\left(1 - \frac{\alpha_W}{\alpha_R}\right)^{\frac{\alpha_R}{\alpha_W}}}{1 - \frac{\alpha_W}{\alpha_R}} t^{\frac{\alpha_R}{\alpha_W}} y \left(x^{\frac{\alpha_R}{\alpha_W}-1} - y^{\frac{\alpha_R}{\alpha_W}-1}\right) \\ &= \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} t^\lambda \left(x^{\lambda-1} y - y^\lambda\right) + O\left(t^{\lambda+\tau}\right), \end{aligned}$$

$$\begin{aligned} Q^{(3a)}(tx,ty) &= (1+O(t^\tau)) \frac{\left(1 - \frac{\alpha_W}{\alpha_R}\right)^2}{2\frac{\alpha_W}{\alpha_R} - 1} t^2 xy \left(\left(1 - \frac{\alpha_W}{\alpha_R}\right)^{\frac{\alpha_R}{\alpha_W}-2} (tx)^{\frac{\alpha_R}{\alpha_W}-2} - 1\right) \\ &= (1+O(t^\tau)) \frac{\left(1 - \frac{1}{\lambda}\right)^2}{\frac{2}{\lambda} - 1} t^2 xy \left(\left(1 - \frac{1}{\lambda}\right)^{\lambda-2} (tx)^{\lambda-2} - 1\right) \\ &= (1+O(t^\tau)) \frac{\left(1 - \frac{1}{\lambda}\right)^\lambda}{\frac{2}{\lambda} - 1} t^\lambda x^{\lambda-1} y + O\left(t^2\right) \\ &= \lambda \frac{\left(1 - \frac{1}{\lambda}\right)^\lambda}{2-\lambda} t^\lambda x^{\lambda-1} y + O\left(t^{\lambda+\tau} + t^2\right) \\ &= \lambda \frac{\left(1 - \frac{1}{\lambda}\right)^\lambda}{2-\lambda} t^\lambda x^{\lambda-1} y + O\left(t^{(2\lambda-1)\wedge 2}\right). \end{aligned}$$

Therefore, Q can be calculated as

$$\begin{aligned} Q(tx,ty) &= Q^{(1)}(tx,ty) + Q^{(2)}(tx,ty) + Q^{(3a)}(tx,ty) \\ &= \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} t^\lambda \left(\left(1 - \frac{1}{\lambda}\right) y^\lambda + x^{\lambda-1} y - y^\lambda + \lambda \frac{1 - \frac{1}{\lambda}}{2-\lambda} x^{\lambda-1} y\right) + O\left(t^{(2\lambda-1)\wedge 2}\right) \\ &= \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} t^\lambda \left(-\frac{1}{\lambda} y^\lambda + \left(1 + \lambda \frac{1 - \frac{1}{\lambda}}{2-\lambda}\right) x^{\lambda-1} y\right) + O\left(t^{(2\lambda-1)\wedge 2}\right) \end{aligned}$$

$$= \left(1 - \frac{1}{\lambda}\right)^{\lambda-1} t^{\lambda} \left(\frac{1}{2-\lambda} x^{\lambda-1} y - \frac{1}{\lambda} y^{\lambda}\right) + O\left(t^{(2\lambda-1)\wedge 2}\right).$$

S4.1.3. $\lambda = 2$. In this case, $\alpha_R/\alpha_\wedge = 2$, so we easily see that both $Q^{(1)}(tx, ty)$ and $Q^{(2)}(tx, ty)$ are $O(t^2)$. Because the term $Q^{(3b)}$ is of the order of $t^2 \log(1/t)$, it dominates the preceding two by a factor of $\log(1/t)$. Therefore,

$$Q(tx, ty) = Q^{(3b)}(tx, ty) + O\left(t^2\right) = \frac{1}{2} t^2 \log(1/t) xy + O\left(t^2\right).$$

S4.1.4. $\lambda \in (2, \infty)$. Once again, the terms $Q^{(1)}$ and $Q^{(2)}$ are dominated by the third term; they are both of the order of t^{λ}, whereas the third term is of the order of t^2. Therefore,

$$\begin{aligned}
Q(tx, ty) &= Q^{(3a)}(tx, ty) + O\left(t^{\frac{\alpha_R}{\alpha_W}}\right) \\
&= (1 + O(t^{\tau})) \frac{\left(1 - \frac{\alpha_W}{\alpha_R}\right)^2}{1 - 2\frac{\alpha_W}{\alpha_R}} t^2 xy \left(1 - \left(1 - \frac{\alpha_W}{\alpha_R}\right)^{\frac{\alpha_R}{\alpha_W} - 2} (tx)^{\frac{\alpha_R}{\alpha_W} - 2}\right) + O\left(t^{\frac{\alpha_R}{\alpha_W}}\right) \\
&= (1 + O(t^{\tau})) \frac{\left(1 - \frac{1}{\lambda}\right)^2}{1 - \frac{2}{\lambda}} t^2 xy + O\left(t^{\lambda}\right) \\
&= \frac{\left(1 - \frac{1}{\lambda}\right)^2}{1 - \frac{2}{\lambda}} t^2 xy + O\left(t^{(2+\tau)\wedge\lambda}\right) \\
&= \frac{\left(1 - \frac{1}{\lambda}\right)^2}{1 - \frac{2}{\lambda}} t^2 xy + O\left(t^{\lambda}\right),
\end{aligned}$$

because, in the last line, $2 + \tau = \lambda + 1 > \lambda$.

S4.2. *The case* $\lambda = 1$. From now on, we assume that $\alpha_R = \alpha_W = \alpha$. That is, R, W_1, W_2 are iid with a Pareto (α) distribution. Like before, we denote by $\bar{F}_R$ and $\bar{F}$ the survival functions of R and of X (and equivalently Y), respectively. As before, we first find an expression for $\bar{F}$. For any $x \geq 1$,

$$\begin{aligned}
\bar{F}(x) &= \mathbb{P}\left(RW_1 > x\right) \\
&= \mathbb{P}\left(R > \frac{x}{W_1}\right) \\
&= \mathbb{E}\left[\bar{F}_R\left(\frac{x}{W_1}\right)\right] \\
&= \mathbb{P}\left(W_1 > x\right) + \int_1^x \left(\frac{w}{x}\right)^{\alpha} \frac{\alpha}{w^{\alpha+1}} dw \\
&= x^{-\alpha} + \alpha x^{-\alpha} \int_1^x \frac{dw}{w} \\
&= x^{-\alpha}\left(1 + \alpha \log(x)\right).
\end{aligned}$$

The inverse of this function is given by

$$\bar{F}^{-1}(y) = \left(\frac{-W_{-1}(-y/e)}{y}\right)^{1/\alpha},$$

where W_{-1} denotes the lower branch of the Lambert W function; for $y \in [-e^{-1}, 0)$, $W_{-1}(y)$ denotes the only solution in $x \in (-\infty, -1]$ of the equation $y = xe^x$. Indeed, it can be seen by a simple plug-in argument that for any $y \in (0, 1]$,

$$\bar{F}\left(\left(\frac{-W_{-1}(-y/e)}{y}\right)^{1/\alpha}\right) = y.$$

Repeating the steps leading to Equation (S4.3), we obtain the following similar integral representation for Q:

$$\begin{aligned}
Q(tx, ty) &= \int_0^{\left(\bar{F}^{-1}(ty)\right)^{-\alpha}} da + \left(\bar{F}^{-1}(ty)\right)^{-\alpha} \int_{\left(\bar{F}^{-1}(ty)\right)^{-\alpha}}^{\left(\bar{F}^{-1}(tx)\right)^{-\alpha}} a^{-1} da \\
&\quad + \left(\bar{F}^{-1}(tx)\bar{F}^{-1}(ty)\right)^{-\alpha} \int_{\left(\bar{F}^{-1}(tx)\right)^{-\alpha}}^{1} a^{-2} da \\
&= \left(\bar{F}^{-1}(ty)\right)^{-\alpha} + \left(\bar{F}^{-1}(ty)\right)^{-\alpha} \log\left(\frac{\left(\bar{F}^{-1}(tx)\right)^{-\alpha}}{\left(\bar{F}^{-1}(ty)\right)^{-\alpha}}\right) \\
&\quad + \left(\bar{F}^{-1}(tx)\bar{F}^{-1}(ty)\right)^{-\alpha}\left(\left(\bar{F}^{-1}(tx)\right)^{\alpha} - 1\right) \\
&= \left(\bar{F}^{-1}(ty)\right)^{-\alpha}\left(2 + \log\left(\frac{\left(\bar{F}^{-1}(tx)\right)^{-\alpha}}{\left(\bar{F}^{-1}(ty)\right)^{-\alpha}}\right)\right) - \left(\bar{F}^{-1}(tx)\bar{F}^{-1}(ty)\right)^{-\alpha}.
\end{aligned}$$

The last term in this expression is negligible, compared to the first one, by a factor of at least $\left(\bar{F}^{-1}(ty)\right)^{-\alpha}$, which (we shall see) is small enough to be absorbed by the term $O(q_1(t))$.

Now, by Corless et al. (1996), Section 4, we may obtain the following expansion of $\left(\bar{F}^{-1}(t)\right)^{-\alpha}$ as $t \to 0$:

$$\begin{aligned}
\left(\bar{F}^{-1}(t)\right)^{-\alpha} &= \frac{t}{-W_{-1}(-t/e)} \\
&= \frac{t}{\log(e/t) + \log\log(e/t) + o(1)} \\
&= \frac{t}{\log(1/t) + \log\log(1/t) + O(1)} \\
&= \left(1 + O\left(\frac{1}{\log(1/t)}\right)\right) \frac{t}{\log(1/t) + \log\log(1/t)}.
\end{aligned}$$

Note that, since we are only interested in $(x, y) \in (0, 1]^2$ and since we assume $y \leq x$, $1/\log(1/ty) \leq 1/\log(1/tx) \leq 1/\log(1/t)$. Plugging the expansion in our expression for Q yields

$$\begin{aligned}
Q(tx, ty) &= \left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \frac{ty}{\log(1/ty) + \log\log(1/ty)} \\
&\quad \times \left(2 + \log\left(\frac{\left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \frac{tx}{\log(1/tx) + \log\log(1/tx)}}{\left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \frac{ty}{\log(1/ty) + \log\log(1/ty)}}\right)\right) \\
&\quad + O\left(\left(\frac{t}{\log(1/t) + \log\log(1/t)}\right)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \frac{ty}{\log(1/t) + \log\log(1/t) + O(\log(1/y))} \\
&\quad \times \left(2 + \log\left(\left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \frac{x}{y} \frac{\log(1/t) + \log\log(1/t) + O(\log(1/y))}{\log(1/t) + \log\log(1/t) + O(\log(1/x))}\right)\right) \\
&\quad + O\left(\left(\frac{t}{\log(1/t) + \log\log(1/t)}\right)^2\right)
\end{aligned} \tag{S4.4}
$$

Note that the first term thereof can be written as

$$
\begin{aligned}
\frac{ty}{\log(1/t) + \log\log(1/t) + O(\log(1/y))} &= \frac{ty}{\log(1/t) + \log\log(1/t)} \left\{1 + O\left(\frac{\log(1/y)}{\log(1/t)}\right)\right\} \\
&= \frac{ty}{\log(1/t) + \log\log(1/t)} \left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\}
\end{aligned}
$$

because as y approaches 0, the term $\log(1/y)$ gets absorbed by the term y on the numerator. Furthermore,

$$
\begin{aligned}
\frac{x}{y} \frac{\log(1/t) + \log\log(1/t) + O(\log(1/y))}{\log(1/t) + \log\log(1/t) + O(\log(1/x))} &= \frac{x}{y} \left\{1 + O\left(\frac{\log(1/x) + \log(1/y)}{\log(1/t)}\right)\right\} \\
&= \frac{x}{y} \left\{1 + O\left(\frac{\log(1/y)}{\log(1/t)}\right)\right\}.
\end{aligned}
$$

Thus the $\log$ term in Equation (S4.4) equals

$$
\log\left(\frac{x}{y} + O\left(\frac{x}{y}\frac{\log(1/y)}{\log(1/t)}\right)\right) = \log\left(\frac{x}{y}\right) + O\left(\frac{\frac{x}{y}\frac{\log(1/y)}{\log(1/t)}}{x/y}\right) = \log\left(\frac{x}{y}\right) + O\left(\frac{\log(1/y)}{\log(1/t)}\right),
$$

where we have used the fact that, for any $a \geq 1$ and $b \geq 0$, $\log(a+b) \leq \log(a) + b/a$ (recall that $x/y \geq 1$). Piecing everything together, Equation (S4.4) may be rewritten as

$$
\begin{aligned}
Q(tx, ty) &= \frac{ty}{\log(1/t) + \log\log(1/t)} \left(2 + \log\left(\frac{x}{y}\right) + O\left(\frac{\log(1/y)}{\log(1/t)}\right)\right) \left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\} \\
&= \frac{ty}{\log(1/t) + \log\log(1/t)} \left(2 + \log\left(\frac{x}{y}\right)\right) \left\{1 + O\left(\frac{1}{\log(1/t)}\right)\right\},
\end{aligned}
$$

once again because the term $\log(1/y)$ is absorbed by y as y approaches 0. Recalling that we assumed $y \leq x$, the claim follows. $\square$

S5. A few words on the computational complexity of the method in spatial problems. Both estimators we propose in the spatial setting (defined in Equations (3.8) and (3.9)) essentially rely on the evaluation of bivariate functions and as such are much faster than methods based on full likelihood (especially if the number of locations is large). A comparison with pairwise likelihood depends on the cost of likelihood evaluations in the particular model under consideration and the type of weight functions that we choose. For the sake of brevity we will focus on the estimator $\widehat{\vartheta}$ from Equation (3.8); similar arguments apply to $\widetilde{\vartheta}$ from Equation (3.9) with obvious modifications.

Typically, we expect that $\widehat{\vartheta}$ can be computed faster than a pairwise likelihood-based estimator. The main computational burden arises when computing the pairwise empirical integrals $\int g(x,y)\widehat{Q}^{(s)}(kx/n, ky/n)dxdy$ and the corresponding estimators $\widehat{\theta}_n^{(s)}$. In computing those estimators, when finding the minimizer of

$$\left\| \int g(x,y)\widehat{Q}^{(s)}(kx/n, ky/n)dxdy - \zeta \int g(x,y)c_\theta(x,y)dxdy \right\|$$

through numerical optimization, only population level integrals $\int g(x,y)c_\theta(x,y)dxdy$ need to be re-computed for each optimization step. For specific models (such as the inverted Brown–Resnick process considered in our application) those integrals have simple analytic expressions, which additionally speeds up the computation. In comparison, the likelihood of a bivariate extreme value model may be substantially more costly to compute, and it needs to be evaluated at every optimization step.

The above procedure only needs to be completed once and can easily be parallelized by considering pairs independently. Once the estimators $\widehat{\theta}_n^{(s)}$ are available, the objective function in Equation (3.8) only depends on evaluating the low-dimensional functions $h^{(s)}$. Again, in our example those are very simple analytic functions.

To give a rough idea of the computation times for the proposed methods in a specific example, we report below average computation times for the spatial simulation study in Section 5.2, with $d = 40$ locations (corresponding to 780 pairs), $n = 5000$ and a few different values of m. All computation times are for computing both spatial estimators simultaneously (but the time to compute only one is not so different since most of the "pairwise" steps leading to each estimator are the same). The values given are averaged based on 100 repetitions and the values in parenthesis are standard deviations. All computations were executed on a personal laptop with a 2.5GHz Intel Core i5-7200U processor without utilizing parallel computation.

m	25	100	250	500	1000
time (seconds)	9.6 (0.6)	9.5 (0.3)	9.6 (0.4)	9.8 (0.3)	9.8 (0.3)

S6. Additional simulation results. This section contains additional simulation results not included in Section 5.

S6.1. *Bivariate distributions.* The following scatter plots represent data from each of the three bivariate models M1–M3 found in Section 5.1. For illustration purposes, there is no additive noise and the marginals are transformed to unit exponential.

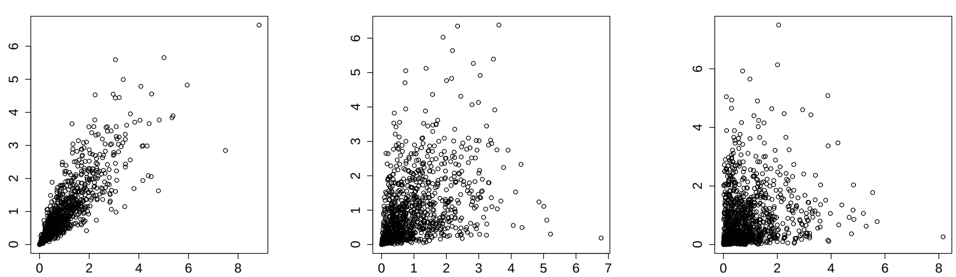

FIG S1. *Samples of 1 000 data points from the inverted Hüsler–Reiss distribution with parameter θ equal to 0.6, 0.75 and 0.9, from left to right. The marginal distributions are scaled to unit exponential.*

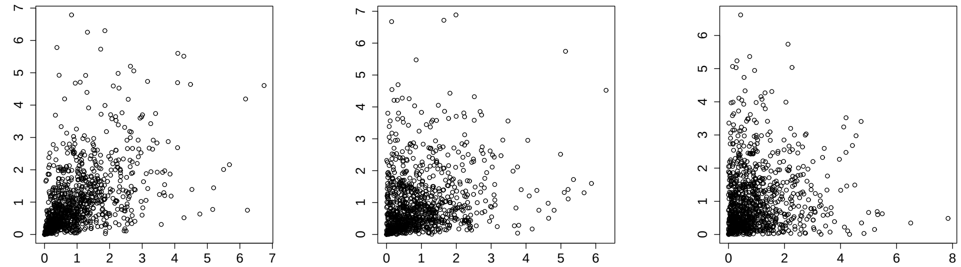

FIG S2. *Samples of 1 000 data points from the inverted asymmetric logistic distribution with parameter θ equal to* $(0.72, 0.72)$, $(0.75, 0.91)$ *and* $(0.91, 0.91)$, *from left to right. The marginal distributions are unit exponential.*

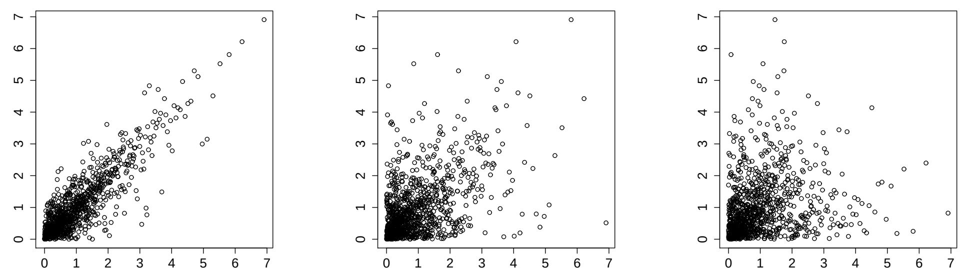

FIG S3. *Samples of 1 000 data points from the Pareto random scale model with parameter λ equal to 0.4, 1 and 1.6, from left to right. The marginal distributions are approximately unit exponential.*

S6.1.1. *Sensitivity with respect to the weight function.* Recall the weight function in Equation (5.1) that is used throughout the paper. It is composed of the weighted indicator functions of the five rectangles $I_1 := [0,1]^2$, $I_2 := [0,2]^2$, $I_3 := [1/2, 3/2]^2$, $I_4 := [0,1] \times [0,3]$ and $I_5 := [0,3] \times [0,1]$. As explained in Section 5.1, those rectangles are chosen specifically to ensure identifiability in every model, so that a unique weight function may be used for all simulations.

We now consider different subsets of the five rectangles above and repeat the simulation study with each of the associated lower dimensional weight functions. Precisely, we define $g^{(1)}$ as the function g in Equation (5.1) and by the same principle we construct $g^{(2)}, \dots, g^{(7)}$, using the rectangles in Table S1.

Weight fct.	$g^{(1)}$	$g^{(2)}$	$g^{(3)}$	$g^{(4)}$	$g^{(5)}$	$g^{(6)}$	$g^{(7)}$
Rectangles	I_1, I_2, I_3, I_4, I_5	I_1, I_2	I_1, I_3	I_1, I_4, I_5	I_1, I_2, I_3	I_1, I_2, I_4, I_5	I_1, I_3, I_4, I_5

TABLE S1
Rectangles used to construct each weight function.

We repeat the simulation study from Section 5.1; 1 000 data sets of size $n = 5\,000$ are drawn from each of the three models, with the same noise mechanism as before, and from each data set seven estimators are computed based on the seven weight functions. We use the values k that were deemed good previously, that is 800 for the two inverted max-stable models (M1 and M2) and 400 for the Pareto random scale model (M3). For each model and each parameter value, we compare the weight functions based on the estimated RMSE of the M-estimator in Figure S4.

In the inverted Hüsler–Reiss model, the parameter has a one-to-one relation with the coefficient of homogeneity $1/\eta$ of c. In order to identify that coefficient, it is sufficient to compare the integral of c over the rectangles I_1 and I_2. It can moreover be deduced from the developments in Section S3 that in this model, the bias arising from the pre-asymptotic approximation of c is largest around the axes. Thus, as can be observed below, adding the non required rectangles I_4 and I_5, which contain a large portion of the axes, adds bias to the estimator. The best strategy for this model seems to be using I_1, I_2 and possibly I_3.

In contrast, the parameter in the inverted asymmetric logistic model is not identifiable if the rectangles used are all symmetric, since then (θ_1, θ_2) cannot be distinguished from (θ_2, θ_1). Therefore the estimator is not uniquely defined when neither I_4 nor I_5 is used, so the functions $g^{(2)}$, $g^{(3)}$ and $g^{(5)}$ were not included. It is to be noted that $g^{(4)}$ does not include either of I_2 and I_3, and as such is not able to estimate the homogeneity coefficient $\theta_1 + \theta_2$ well, even if it is able to recover the asymmetry. This explains the monotonic behavior of the error with respect to $\theta_1 + \theta_2$. The other three weight functions perform similarly to each other.

Finally, in the Pareto random scale model, the weight function $g^{(2)}$ only estimates the homogeneity and as such, it is unable to distinguish the parameters in the range $(0, 1)$, corresponding to asymptotic dependence. It was thus ignored. Among the other functions, the ones that use I_4 and I_5 ($g^{(1)}$, $g^{(4)}$, $g^{(6)}$, $g^{(7)}$) all have a similar performance whereas the other two ($g^{(3)}$ and $g^{(5)}$) incur a noticeably larger error. It seems that those rectangles help estimating characteristics that are strongly different from the coefficient of homogeneity, which explains why they significantly reduce the RMSE under asymptotic dependence ($\lambda < 1$).

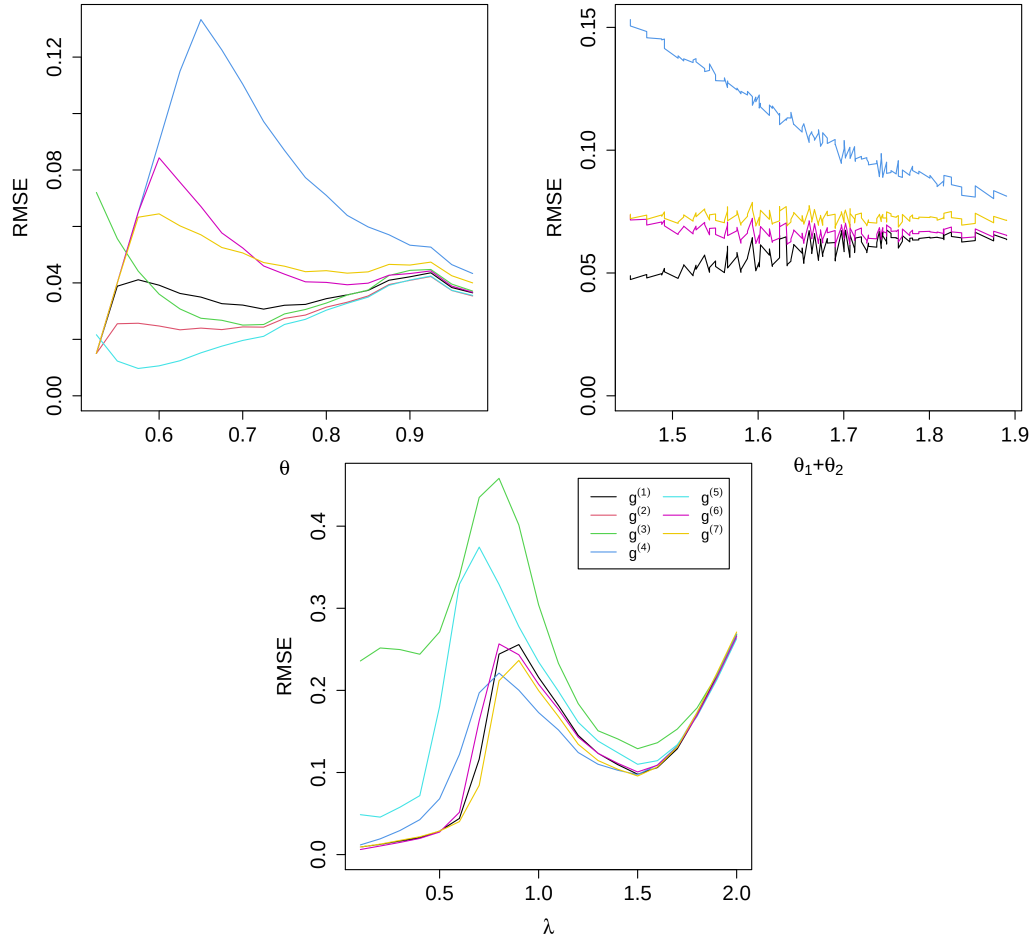

FIG S4. *RMSE of the M-estimator in the models M1–M3 as a function of the parameter, based on 1 000 data sets of size* $n = 5\,000$, $k = 800$ *(for M1 and M2) and* $k = 400$ *(for M3). Colors represent the seven weight functions from Table S1.*

S6.2. *Spatial models.* Figure S5 shows the distribution of the distances of all the pairs that are used in the analysis in Section 5.2. Figures S6 and S7 present the same results as in Section 5.2 when the estimator (3.9) is used instead of (3.8).

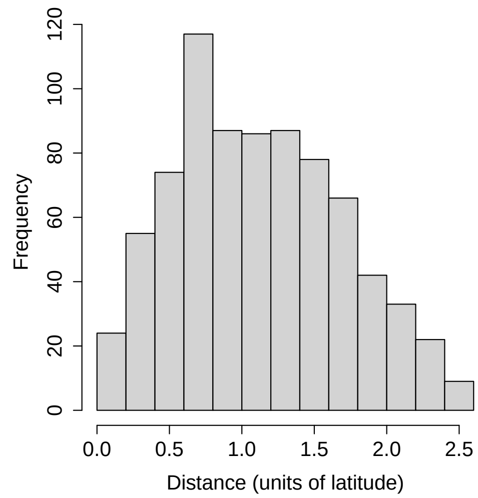

FIG S5. *Distribution of the distances* $\Delta^{(s)}$ *for the 780 pairs used.*

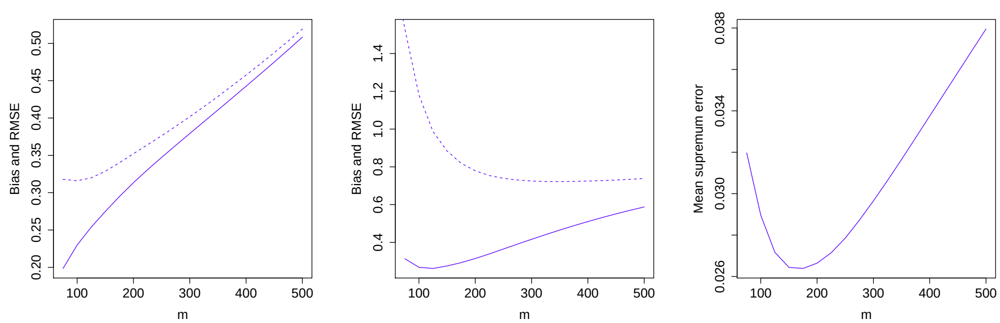

FIG S6. *Left and middle columns: Bias (solid line) and RMSE (dotted line) of the estimators of the two spatial parameters* α *(left) and* β *(middle) as a function of* m*. Right: Mean of the supremum error* $\sup_{0\leq\Delta\leq 3}|\theta(\Delta;\widehat{\alpha},\widehat{\beta}) - \theta(\Delta;\alpha,\beta)|$ *as a function of* m*.*

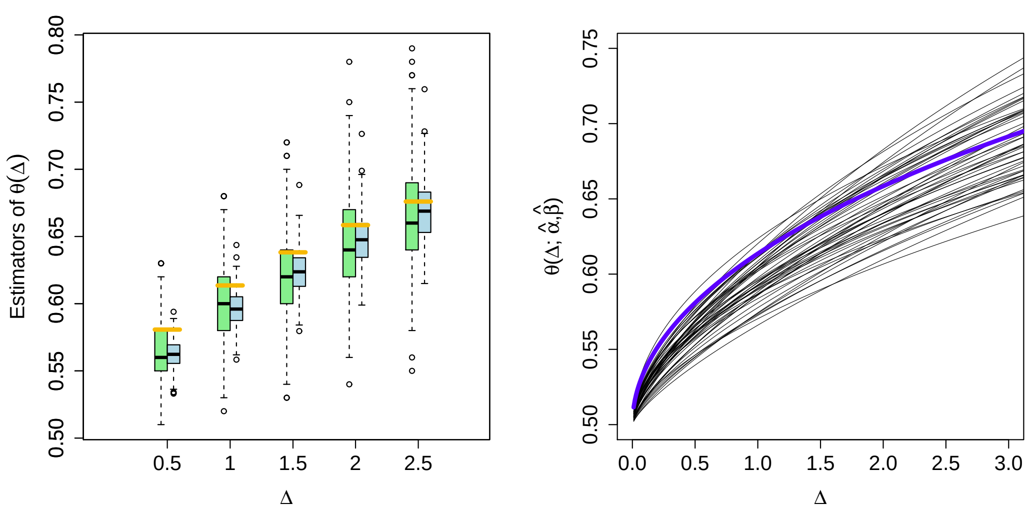

FIG S7. *Left panel: Estimators of* $\theta(\Delta)$ *for 5 different distances. For each distance, bivariate M-estimator* $\widehat{\theta}_n^{(s)}$ *(green) and spatial estimator* $\theta(\Delta^{(s)};\widehat{\alpha},\widehat{\beta})$ *(blue) based on the* $d = 40$ *locations. Right panel: 50 sampled curves* $\theta(\cdot;\widehat{\alpha},\widehat{\beta})$*. Blue represents the true curve* $\theta(\cdot;\alpha,\beta)$*.*

REFERENCES

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge University Press.

BÜCHER, A., SEGERS, J. and VOLGUSHEV, S. (2014). When Uniform Weak Convergence Fails: Empirical Processes for Dependence Functions and Residuals via Epi- and Hypographs. *Ann. Stat.* **42** 1598-1634.

CORLESS, R. M., GONNET, G. H., HARE, D. E., JEFFREY, D. J. and KNUTH, D. E. (1996). On the LambertW function. *Advances in Computational mathematics* **5** 329–359.

CSÖRGŐ, M. and HORVÁTH, L. (1987). Approximation of Intermediate Quantile Processes. *Journal of multivariate analysis* **21** 250–262.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer.

LALANCETTE, M., ENGELKE, S. and VOLGUSHEV, S. (2020+). Rank-based Estimation under Asymptotic Dependence and Independence, with Applications to Spatial Extremes. *Ann. Stat.*

RAMOS, A. and LEDFORD, A. (2009). A new class of models for bivariate joint tails. *J. R. Stat. Soc. B* **71** 219-241.

RESNICK, S. I. (1987). *Extreme values, regular variation and point processes*. Springer.

ROCKAFELLAR, R. T. (1970). *Convex Analysis* **28**. Princeton university press.

TESCHL, G. (1998). *Topics in Real and Functional Analysis*. Unpublished, available online at https://www.mat.univie.ac.at/gerald/ftp/book-fa.

VAN DER VAART, A. W. (2000). *Asymptotic statistics*. Cambridge university press.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer.

VERVAAT, W. (1972). Functional Central Limit Theorems for Processes with Positive Drift and Their Inverses. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **23** 245-253.