

# Threshold selection for multivariate heavy-tailed data

Phyllis Wan<sup>1</sup>  · Richard A. Davis<sup>1</sup>

Received: 21 June 2017 / Revised: 23 March 2018 / Accepted: 26 March 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** Regular variation is often used as the starting point for modeling multivariate heavy-tailed data. A random vector is regularly varying if and only if its radial part  $R$  is regularly varying and is asymptotically independent of the angular part  $\Theta$  as  $R$  goes to infinity. The conditional limiting distribution of  $\Theta$  given  $R$  is large characterizes the tail dependence of the random vector and hence its estimation is the primary goal of applications. A typical strategy is to look at the angular components of the data for which the radial parts exceed some threshold. While a large class of methods has been proposed to model the angular distribution from these exceedances, the choice of threshold has been scarcely discussed in the literature. In this paper, we describe a procedure for choosing the threshold by formally testing the independence of  $R$  and  $\Theta$  using a measure of dependence called distance covariance. We generalize the limit theorem for distance covariance to our unique setting and propose an algorithm which selects the threshold for  $R$ . This algorithm incorporates a subsampling scheme that is also applicable to weakly dependent data. Moreover, it avoids the heavy computation in the calculation of the distance covariance, a typical limitation for this measure. The performance of our method is illustrated on both simulated and real data.

**Keywords** Heavy-tailed data · Multivariate regular variation · Threshold selection · Distance covariance

---

This research was supported in part by MURI ARO (grant no. W911NF-12-10385).

---

✉ Phyllis Wan  
phyllis@stat.columbia.edu  
Richard A. Davis  
rdavis@stat.columbia.edu

<sup>1</sup> Department of Statistics, Columbia University, 1255 Amsterdam Avenue, MC 4690, New York, NY 10027, USA

## 1 Introduction

For multivariate heavy-tailed data, the principal objective is often to study dependence in the ‘tail’ of the distribution. To achieve this goal, the assumption of multivariate regular variation is typically used as a starting point. A random vector  $\mathbf{X} \in \mathbb{R}^d$  is said to be *multivariate regularly varying* if the polar coordinates  $(R, \Theta) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|)$ , where  $\|\cdot\|$  is some norm, satisfy the conditions

- (a)  $R$  is univariate regularly varying, i.e.,  $\mathbb{P}(R > r) = L(r)r^{-\alpha}$ , where  $L(\cdot)$  is a slowly varying function at infinity;
- (b)  $\mathbb{P}(\Theta \in \cdot | R > r)$  converges weakly to a measure  $S(\cdot)$  as  $r \rightarrow \infty$ .

The  $\alpha$  is referred to as the index of the regular variation, while the  $S$  is called the angular distribution and characterizes the limiting tail dependence. There are other equivalent definitions of regular variation (Resnick 2002), but this one is the most convenient for our purposes.

Given observations  $\{\mathbf{X}_i\}_{i=1}^n$  and their corresponding polar coordinates  $\{(R_i, \Theta_i)\}_{i=1}^n$ , a straightforward procedure for estimating  $S$  is to look at angular components of the data for which the radii are greater than a large threshold  $r_0$ , that is,  $\Theta_i$  for which  $R_i > r_0$ . In most studies, one takes  $r_0$  to be a large empirical quantile of  $R$ . While there has been extensive research on choosing a threshold for which the distribution of  $R$  is regularly varying (i.e., limit condition (a)), little research has been devoted to ensuring the threshold is large enough for the independence of  $\Theta$  and  $R$  to be reasonable (i.e., limit condition (b)). To this end, de Haan and de Ronde (1998) fit a parametric extreme value distribution model to each marginal and examined the parameter stability plot of each coordinate. The Střaricá plot (Střaricá 1999) looked at the joint tail empirical measure, but was, in some way, equivalent to only examining the extremal behavior of  $R$ . Resnick (2007) suggested an automatic threshold selection from the Střaricá plot but observed that the thresholds were sometimes systematically underestimated. In their study of the threshold based inference for parametric max-stable processes, Jeon and Smith (2014) suggested choosing the threshold by minimizing the MSE of the estimated parameters.

In this paper, we propose an algorithm which selects the threshold for modeling  $S$ . Our motivation is the implied property that  $R$  and  $\Theta$  given  $R > r$  become independent as  $r \rightarrow \infty$ . Given a sequence of candidate threshold levels, we test the degree of dependence between  $R$  and  $\Theta$  for the truncated data above each level. The dependence measure we use is the distance covariance introduced by Székely et al. (2007). This measure has the ability to account for various types of dependence and to be applicable to data in higher dimensions. The resulting test statistics are given in the form of  $p$ -values and are compared across all levels through a subsampling scheme. This enables us to extract more accurate information from the test statistics while not overloading the computational burden.

The remainder of this paper is organized as follows. We first provide some theoretical backgrounds on multivariate regular variation in Section 2. The distance

covariance and its theoretical properties are introduced in Section 3. Applying this dependence measure in our conditioning setting, we propose a test statistic and prove relevant theoretical results in Section 4. Our proposed algorithm for threshold selection is presented in Section 5, and illustrated on simulated and real examples in Section 6. The paper concludes with a discussion.

## 2 Multivariate regular variation and problem set-up

One way to approach multivariate heavy-tailed data is through the notion of multivariate regular variation. For a detailed review, see, for example, Chapter 6 of Resnick (2007). Let  $\mathbf{X} = (X_1, \dots, X_d)$  be a  $d$ -dimensional random variable defined on the cone  $\mathbb{R}_+^d = [\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$ . Define the polar coordinate transformation

$$T(\mathbf{X}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|) =: (R, \Theta), \quad (1)$$

where  $\|\cdot\|$  denotes some norm. Then  $\mathbf{X}$  is regularly varying if and only if there exists a probability measure  $S(\cdot)$  on  $\mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$ , and a function  $b(t) \rightarrow \infty$ , such that

$$t\mathbb{P}[(R/b(t), \Theta) \in \cdot] \xrightarrow{v} \nu_\alpha \times S, \quad t \rightarrow \infty, \quad (2)$$

where  $\xrightarrow{v}$  denotes vague convergence, and  $\nu_\alpha$  is a measure defined on  $(0, \infty]$  such that

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \quad x > 0.$$

Here  $b(t)$  can be chosen as the  $1 - t^{-1}$ -quantile, i.e.,

$$b(t) = \inf\{s | \mathbb{P}(R \leq s) \geq 1 - t^{-1}\}.$$

The convergence (2) implies that

$$\mathbb{P}\left[\left(\frac{R}{r}, \Theta\right) \in \cdot \mid R > r\right] \xrightarrow{w} \nu_\alpha \times S, \quad r \rightarrow \infty, \quad \text{on } [1, \infty) \times \mathbb{S}^{d-1}, \quad (3)$$

where  $\xrightarrow{w}$  denotes weak convergence. In other words, given that  $R > r$  for  $r$  large, the conditional distribution of  $R/r$  and  $\Theta$  are independent in the limit. In view of Eq. 3, we restrict the measure  $\nu_\alpha$  to  $[1, \infty)$  throughout the remainder of the paper. The angular measure  $S$  characterizes the tail dependence structure of  $\mathbf{X}$ . If  $S$  is concentrated on  $\{e_i, i = 1, \dots, d\}$ , where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ , then the components of  $\mathbf{X}$  are asymptotically independent in the tail, a case known as asymptotic independence. If  $S$  has mass lying in the subspace  $\{(t_1, \dots, t_d) \in \mathbb{S}^{d-1} | t_i > 0, t_j > 0, i \neq j\}$ , then an extreme observation in the  $X_i$  direction implicates a positive probability of an extreme observation in the  $X_j$  direction, a case known as asymptotic dependence. Hence the estimation of  $S$  from observations is an important problem, and often the primary goal, in multivariate heavy-tailed modeling.

The following convergence is implied from Eq. 3:

$$\mathbb{P}(\Theta \in \cdot | R > r) \xrightarrow{w} S(\cdot), \quad r \rightarrow \infty. \quad (4)$$

This suggests estimating  $S$  using the angular data  $(\Theta_i)$  whose radial parts satisfy  $R_i > r_0$  for  $r_0$  large. The motivation behind our method is to seek  $r_0$  such that when

$R > r_0$ ,  $R$  and  $\Theta$  are virtually independent. Given a candidate threshold sequence  $\{r_k\}$ , we formally test the independence between  $(R_i, \Theta_i)$  on the index set  $\{i | R_i > r_k\}$ . The use of Pearson's correlation as the dependence measure is unsuitable in this case, for two reasons. First, correlation is only applicable to univariate random variables, whereas  $\Theta$  lies on the sphere of dimension  $d - 1$ . Second, correlation only describes the linear relationship between two random variables, thus having zero correlation is not a sufficient condition for independence. Instead, we use a more powerful dependence measure, the *distance covariance*, which is introduced in the next section.

### 3 Distance covariance

In this section, we briefly review the definition and some properties of the distance covariance. More detailed descriptions and proofs can be found in Székely et al. (2007) and Davis et al. (2017).

Let  $X \in \mathbb{R}^p$  and  $Y \in \mathbb{R}^q$  be two random vectors, then the distance covariance between  $X$  and  $Y$  is defined as

$$T(X, Y; \mu) = \iint_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(s, t) - \varphi_X(s) \varphi_Y(t)|^2 \mu(ds, dt), \quad (s, t) \in \mathbb{R}^{p+q} \quad (5)$$

where  $\varphi_{X,Y}(s, t)$ ,  $\varphi_X(s)$ ,  $\varphi_Y(t)$  denote the joint and marginal characteristic functions of  $(X, Y)$  and  $\mu$  is a suitable measure on  $\mathbb{R}^{p+q}$ . In order to ensure that  $T(X, Y; \mu)$  is well-defined, one of the following conditions is assumed to be satisfied throughout the paper (Davis et al. 2017):

1.  $\mu$  is a finite measure on  $\mathbb{R}^{p+q}$ ;
2.  $\mu$  is an infinite measure on  $\mathbb{R}^{p+q}$  such that

$$\iint_{\mathbb{R}^{p+q}} (1 \wedge |s|^\alpha)(1 \wedge |t|^\alpha) \mu(ds, dt) < \infty$$

and

$$\mathbb{E}[|XY|^\alpha + |X|^\alpha + |Y|^\alpha] < \infty$$

for some  $\alpha \in (0, 2]$ .

One advantage of distance covariance over, say, Pearson's covariance, is that, if  $\mu$  has a positive Lebesgue density on  $\mathbb{R}^{p+q}$ , then  $X$  and  $Y$  are independent if and only if  $T(X, Y; \mu) = 0$ . Another attractive property of this dependence measure is that it readily applies to random vectors of different dimensions.

To estimate  $T(X, Y; \mu)$  from observations  $(X_1, Y_1), \dots, (X_n, Y_n)$ , define the empirical distance covariance

$$T_n(X, Y; \mu) = \iint_{\mathbb{R}^{p+q}} |\hat{\varphi}_{X,Y}(s, t) - \hat{\varphi}_X(s) \hat{\varphi}_Y(t)|^2 \mu(ds, dt),$$

where  $\hat{\varphi}_{X,Y}(s, t) = \frac{1}{n} \sum_{j=1}^n e^{i s \cdot X_j + i t \cdot Y_j}$  and  $\hat{\varphi}_X(s) = \hat{\varphi}_{X,Y}(s, 0)$ ,  $\hat{\varphi}_Y(t) = \hat{\varphi}_{X,Y}(0, t)$  are the respective empirical characteristic functions. If we assume that

$\mu = \mu_1 \times \mu_2$  and is symmetric about the origin, then under the conditions where  $T(X, Y; \mu)$  exists,  $T_n(X, Y; \mu)$  also has the computable form

$$\begin{aligned} T_n(X, Y; \mu) = & \frac{1}{n^2} \sum_{i,j=1}^n \tilde{\mu}_1(X_i - X_j) \tilde{\mu}_2(Y_i - Y_j) \\ & + \frac{1}{n^4} \sum_{i,j,k,l=1}^n \tilde{\mu}_1(X_i - X_j) \tilde{\mu}_2(Y_k - Y_l) \\ & - \frac{2}{n^3} \sum_{i,j,k=1}^n \tilde{\mu}_1(X_i - X_j) \tilde{\mu}_2(Y_i - Y_k), \end{aligned}$$

where  $\tilde{\mu}(x) = (1 - \cos s, x) \mu(ds)$  (Davis et al. 2017).

The most popular choice of  $\mu$ , first mentioned by Feuerverger (1993) and then more extensively studied by Székely et al. (2007), is

$$\mu(ds, dt) = c_{p,q} |s|^{-\kappa-p} |t|^{-\kappa-q} ds dt. \quad (6)$$

where  $c_{p,q}$  is as defined in Lemma 1 of Székely et al. (2007). This choice of  $\mu$  gives  $\tilde{\mu}(x)\tilde{\mu}(y) = |x|^\kappa |y|^\kappa$ . Moreover, this is the only choice of  $\mu$  for which the distance covariance is invariant relative to scale and orthogonal transformations. Note that in order for the integral (5) to exist, it is required that

$$E[|X|^\kappa |Y|^\kappa + |X|^\kappa + |Y|^\kappa] < \infty.$$

We will utilize the described weight measure (6) with  $\kappa = 1$  in our simulations and data analyzes in Section 6, but applied to the log transformation on  $R$  to ensure that the moment condition is satisfied.

As detailed in Davis et al. (2017), if the sequence  $\{(X_i, Y_i)\}$  is stationary and ergodic, then

$$T_n(X, Y; \mu) \xrightarrow{\text{a.s.}} T(X, Y; \mu). \quad (7)$$

Further, if  $X$  and  $Y$  are independent, then under an  $\alpha$ -mixing condition,

$$n T_n(X, Y; \mu) \xrightarrow{d} \int_{\mathbb{R}^{p+q}} |G(s, t)|^2 \mu(s, t) \quad (8)$$

for some centered Gaussian field  $G$ . On the other hand, if  $X$  and  $Y$  are dependent, then

$$\sqrt{n}(T_n(X, Y; \mu) - T(X, Y; \mu)) \xrightarrow{d} G_\mu$$

for some non-trivial limit  $G_\mu$ , implying that  $n T_n(X, Y; \mu)$  diverges as  $n \rightarrow \infty$ . Naturally one can devise a test of independence between  $X$  and  $Y$  using the statistic  $n T_n(X, Y; \mu)$ : the null hypothesis of independence is rejected at level  $\chi$  if  $n T_n(X, Y; \mu) > c$ , where  $c$  is the upper  $\chi$ -quantile of  $\int_{\mathbb{R}^{p+q}} |G(s, t)|^2 \mu(s, t)$ .

In practice, the distribution  $\int_{\mathbb{R}^{p+q}} |G(s, t)|^2 \mu(s, t)$  is intractable and is typically approximated through bootstrap. Hence the main drawback of using distance covariance is the computation burden it brings for large sample size: the computation of a single distance covariance statistic requires  $O(n^2)$  operations, while finding the cut-off values via resampling requires much more additional computation. Our method,

however, overcomes this problem through subsampling the data, as will be described in Section 5.

## 4 Theoretical results

Let  $\{\mathbf{X}_i\}_{i=1}^n$  be iid observations in  $\mathbb{R}^d$  from a multivariate regularly varying distribution  $\mathbf{X}$  satisfying Eqs. 1 and 3, and  $\{(R_i, \Theta_i)\}_{i=1}^n$  be their polar coordinate transformations. Given a threshold  $r_n$ , we measure the dependence between  $R/r_n$  and  $\Theta$  conditional on  $R > r_n$  by the empirical distance covariance of  $(R/r_n, \Theta)$  on the set  $\{R > r_n\}$ :

$$T_n := \int_{\mathbb{R}^{d+1}} |C_n(s, t)|^2 \mu(ds, dt), \quad (9)$$

with

$$C_n(s, t) := \hat{\varphi}_{\frac{R}{r_n}, \Theta | r_n}(s, t) - \hat{\varphi}_{\frac{R}{r_n} | r_n}(s) \hat{\varphi}_{\Theta | r_n}(t),$$

where  $\hat{\varphi}_{\frac{R}{r_n}, \Theta | r_n}$  is the conditional empirical characteristic function of  $(R/r_n, \Theta)$ ,

$$\hat{\varphi}_{\frac{R}{r_n}, \Theta | r_n}(s, t) = \frac{1}{\sum_{j=1}^n \mathbf{1}_{\{R_j > r_n\}}} \sum_{j=1}^n e^{isR_j/r_n + it^T \Theta_j} \mathbf{1}_{\{R_j > r_n\}}, \quad s \in \mathbb{R}, \quad t = (t_1, \dots, t_d)^T \in \mathbb{R}^d,$$

and  $\hat{\varphi}_{\frac{R}{r_n} | r_n}, \hat{\varphi}_{\Theta | r_n}$  are the corresponding empirical conditional marginal characteristic functions,

$$\hat{\varphi}_{\frac{R}{r_n} | r_n}(s) = \hat{\varphi}_{\frac{R}{r_n}, \Theta | r_n}(s, 0), \quad \hat{\varphi}_{\Theta | r_n}(t) = \hat{\varphi}_{\frac{R}{r_n}, \Theta | r_n}(0, t).$$

In this section, we establish the limiting results (7) and (8) adapted to the conditional distance covariance. For ease of notation, let

$$p_n := \mathbb{P}(R > r_n), \quad \hat{p}_n := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{R_j > r_n\}}$$

be the theoretical and empirical probability of exceedance, and let

$$\varphi_{\frac{R}{r_n}, \Theta | r_n}(s, t) := \mathbb{E} e^{isR/r_n + it^T \Theta} | R > r_n = \frac{\mathbb{E} e^{isR/r_n + it^T \Theta} \mathbf{1}_{R > r_n}}{p_n},$$

and

$$\varphi_{\frac{R}{r_n} | r_n}(s) := \varphi_{\frac{R}{r_n}, \Theta | r_n}(s, 0), \quad \varphi_{\Theta | r_n}(t) := \varphi_{\frac{R}{r_n}, \Theta | r_n}(0, t),$$

be the theoretical conditional joint and marginal characteristic functions.

Recall from Eq. 3 that as  $n \rightarrow \infty$ ,  $R/r_n$  and  $\Theta$  become asymptotically independent and converge to  $v_\alpha$  and  $S$  respectively. Denote the characteristic functions of the corresponding limit distributions by

$$\varphi_R(s) := \int_1^\infty \exp(isr) \alpha r^{-\alpha-1} dr = \lim_{n \rightarrow \infty} \varphi_{\frac{R}{r_n} | r_n}(s), \quad (10)$$

$$\varphi_\Theta(t) := \int_{\mathbb{S}^{d-1}} \exp(it\theta) S(d\theta) = \lim_{n \rightarrow \infty} \varphi_{\Theta | r_n}(t). \quad (11)$$

We have the following results.

**Theorem 1** 1. Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid observations generated from  $\mathbf{X}$ , where  $\mathbf{X}$  is multivariate regularly varying with index  $\alpha > 1$ . Let  $T_n$  be the conditional empirical distance covariance between the angular and radial component defined in Eq. 9. Further assume that  $np_n \rightarrow \infty$  and the weight measure  $\mu$  satisfies

$$\int_{\mathbb{R}^{d+1}} (1 \wedge |s|^\beta)(1 \wedge |t|^2) \mu(ds, dt) < \infty, \quad (12)$$

for some  $1 < \beta < 2 \wedge \alpha$ . Then

$$T_n \xrightarrow{P} 0.$$

2. In addition, if  $\{r_n\}$  satisfies

$$np_n \int_{\mathbb{R}^{d+1}} |\varphi_{\frac{R}{r_n}, \Theta|_{r_n}}(s, t) - \varphi_{\frac{R}{r_n}}(s) \varphi_{\Theta|_{r_n}}(t)|^2 \mu(ds, dt) \rightarrow 0, \quad (13)$$

then

$$n \hat{p}_n T_n \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt), \quad (14)$$

where  $Q$  is a centered Gaussian process with covariance function

$$\text{cov}(Q(s, t), Q(s', t')) = (\varphi_R(s - s') - \varphi_R(s) \varphi_R(-s'))(\varphi_\Theta(t - t') - \varphi_\Theta(t) \varphi_\Theta(-t')) \quad (15)$$

with  $\varphi_R, \varphi_\Theta$  as defined in Eqs. 10 and 11.

**Remark 1** In the case where  $\mathbf{X}$  is regularly varying with index  $\alpha \leq 1$ , similar results hold if we replace  $R/r_n$  with  $\log(R/r_n)$  for which all moments exist.

The proof of the theorem is delayed to Appendix A. In the following remark, we discuss certain sufficient conditions for assumption (13).

**Remark 2** Assume that  $\mu = \mu_1 \times \mu_2$ , where  $\mu_1, \mu_2$  are measures on  $\mathbb{R}$  and  $\mathbb{R}^d$ , respectively, and symmetric about the origin. From Section 2.2 of Davis et al. (2017), condition (13) is equivalent to

$$\begin{aligned} np_n \left( \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R}{r_n}) \tilde{\mu}_2(\Theta - \Theta) | R, R > r_n] \right. \\ \left. + \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R}{r_n}) | R, R > r_n] \mathbb{E}[\tilde{\mu}_2(\Theta - \Theta) | R, R > r_n] \right. \\ \left. - 2 \mathbb{E}[\tilde{\mu}_1(\frac{R}{r_n} - \frac{R}{r_n}) \tilde{\mu}_2(\Theta - \Theta) | R, R > r_n] \right) \rightarrow 0, \end{aligned} \quad (16)$$

where

$$\tilde{\mu}_i(x) = \int (1 - \cos(x^T s)) \mu_i(ds), \quad i = 1, 2.$$

Let  $P_{\frac{R}{r_n}, \Theta | r_n}$  denote the conditional joint distribution of  $(R/r_n, \Theta)$  given  $R > r_n$  and  $P_{\frac{R}{r_n} | r_n}, P_{\Theta | r_n}$  be the respective conditional marginals. Then Eq. 16 can be expressed as

$$\begin{aligned}
 & np_n \int_{(1, \infty) \times \mathbb{S}^{d-1}} \int_{(1, \infty) \times \mathbb{S}^{d-1}} \tilde{\mu}_1(T - T) \tilde{\mu}_2(\Theta - \Theta) \\
 & P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) \\
 & + P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta) P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta) \\
 & - 2P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta) \Big\} \\
 = & \int_{(1, \infty) \times \mathbb{S}^{d-1}} \int_{(1, \infty) \times \mathbb{S}^{d-1}} \tilde{\mu}_1(T - T) \tilde{\mu}_2(\Theta - \Theta) \\
 & \sqrt{np_n} \left( P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) - P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta) \right) \\
 & \sqrt{np_n} \left( P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) - P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta) \right) \rightarrow 0, \quad (17)
 \end{aligned}$$

where  $(R, \Theta), (R, \Theta)$  are iid copies of  $(R, \Theta)$ . One way to verify (17) is to assume a second-order like condition on the distribution of  $(R, \Theta)$ . For example, assume that

$$\frac{P_{\frac{R}{r_n}, \Theta | r_n} - \nu_\alpha \times S}{A(r_n)} \xrightarrow{w} \chi,$$

where  $\chi$  is a signed measure such that  $\chi([r, \infty] \times B)$  is finite for all  $r \geq 1$  and  $B$  Borel set in  $\mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$ , and the scalar function  $A(t) \rightarrow 0$  as  $t \rightarrow \infty$ . When the components of  $\mathbf{X}$  are asymptotically independent, this is equivalent to the second order condition for multivariate regular variation (Resnick 2002). If we choose the sequence  $r_n$  such that  $\sqrt{np_n} \rightarrow \infty$  and  $\sqrt{np_n} A(r_n) \rightarrow 0$ , then

$$\begin{aligned}
 & \sqrt{np_n} A(r_n) \frac{P_{\frac{R}{r_n}, \Theta | r_n}((\cdot, \cdot)) - P_{\frac{R}{r_n} | r_n}(\cdot) \times P_{\Theta | r_n}(\cdot)}{A(r_n)} \\
 = & \sqrt{np_n} A(r_n) \left( \frac{P_{\frac{R}{r_n}, \Theta | r_n}((\cdot, \cdot)) - \nu_\alpha \times S((\cdot, \cdot))}{A(r_n)} - \frac{P_{\frac{R}{r_n} | r_n}(\cdot) - \nu_\alpha(\cdot)}{A(r_n)} \right) \times P_{\Theta | r_n}(\cdot) \\
 & - \frac{\nu_\alpha(\cdot) \times (P_{\Theta | r_n}(\cdot) - S(\cdot))}{A(r_n)} \Big\} \xrightarrow{w} 0
 \end{aligned}$$



on  $[1, \infty] \times \mathbb{S}^{d-1}$ . In the case where  $\mu_1, \mu_2$  are finite measures,  $\tilde{\mu}_1, \tilde{\mu}_2$  are bounded and Eq. 17 is satisfied since the integrand can be written as

$$\begin{aligned} & n p_n A^2(r_n) \int_{(1, \infty) \times \mathbb{S}^{d-1}} \left[ \left\{ \int_{(1, \infty) \times \mathbb{S}^{d-1}} \tilde{\mu}_1(T - T) \tilde{\mu}_2(\Theta - \Theta) \right. \right. \\ & \quad \left. \left. \frac{P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) - P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta)}{A(r_n)} \right\} \right. \\ & \quad \left. \frac{P_{\frac{R}{r_n}, \Theta | r_n}(dT, d\Theta) - P_{\frac{R}{r_n} | r_n}(dT) P_{\Theta | r_n}(d\Theta)}{A(r_n)} \right] \rightarrow 0. \end{aligned}$$

In the special case that  $|A| \in RV_\rho$  for  $\rho < 0$ , Eq. 17 is met provided  $r_n$  is chosen such that

$$O(n^{\frac{1}{\alpha+2|\rho|}+}) \leq r_n \leq o(n^{\frac{1}{\alpha}}), \quad \text{for some } > 0.$$

When the measures  $\mu_1, \mu_2$  are infinite, Eq. 13 can be verified in specific cases. This is illustrated in the following example.

*Example 1* Let  $\mathbf{X}$  follow a bivariate logistic distribution, i.e.,  $\mathbf{X}$  has cdf

$$\mathbb{P}(X_1 < x_1, X_2 < x_2) = \exp(-(x_1^{-1/\gamma} + x_2^{-1/\gamma})^\gamma), \quad \gamma \in (0, 1). \quad (18)$$

Then  $\mathbf{X}$  has asymptotically independent components if and only if  $\gamma = 1$ . It can be shown that  $\mathbf{X}$  is regularly varying with index  $\alpha = 1$ , i.e.,  $p_n = \mathbb{P}(R > r_n) \sim r_n^{-1}$  as  $r_n \rightarrow \infty$ . Using the  $L_1$ -norm,  $\|(x_1, x_2)\| = |x_1| + |x_2|$ , the pseudo-polar coordinate transform is  $(R, \Theta) = (X_1 + X_2, X_1/(X_1 + X_2)) \in (0, \infty) \times [0, 1]$  and the pdf of  $(R, \Theta)$  is

$$\begin{aligned} f_{R, \Theta}(r, \theta) &= r^{-2} (\theta(1-\theta))^{-\frac{\gamma+1}{\gamma}} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma-2} e^{-r^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^\gamma} \\ &\quad \left( r^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^\gamma - \frac{\gamma-1}{\gamma} \right) \Big\} \end{aligned}$$

We now consider the case of the infinite weight measure  $\mu$  given in Eq. 6 with  $\kappa = 1$  and derive the condition on the sequence  $\{r_k\}$  for which the conditions of Theorem 1 hold. First observe that

$$\begin{aligned} f_{\frac{R}{r_n}, \Theta | r_n}(t, \theta) &= t^{-2} (\theta(1-\theta))^{-\frac{\gamma+1}{\gamma}} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma-2} e^{-r_n^{-1} t^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^\gamma} \\ &\quad \left( r_n^{-1} t^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^\gamma - \frac{\gamma-1}{\gamma} \right) \Big\} \\ &\rightarrow t^{-2} \frac{1-\gamma}{\gamma} (\theta(1-\theta))^{-\frac{\gamma+1}{\gamma}} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma-2}, \quad \text{as } n \rightarrow \infty, \quad (19) \\ &=: f_T(t) f_\Theta(\theta), \end{aligned}$$

and

$$\begin{aligned}
 & r_n \left| \frac{f_{\frac{R}{r_n}, \Theta|_{r_n}}(t, \theta) - f_T(t)f_{\Theta}(\theta)}{r_n} \right| \left\{ \right. \\
 & \leq f_T(t)f_{\Theta}(\theta) \left( r_n \left| \left\{ \left[ -r_n^{-1}t^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma} - 1 \right] \right. \right. \right. \\
 & \quad \left. \left. \left. + e^{-r_n^{-1}t^{-1} \left( \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma}} t^{-1} \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma} \frac{\gamma}{1-\gamma} \right) \right\} \\
 & \leq f_T(t)f_{\Theta}(\theta) \left( t^{-1} \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma} + t^{-1} \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \right)^{\gamma} \frac{\gamma}{1-\gamma} \left\{ \right. \\
 & \leq t^{-3} \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \left. \right)^{2\gamma-2} \frac{1}{1-\gamma} \\
 & \leq ct^{-3}, \quad \text{for } t \geq 1 \text{ and } \theta \in [0, 1],
 \end{aligned}$$

where  $c$  denotes a generic a constant whose value may change from line to line throughout the paper, and the last inequality comes from the facts that

$$\theta(1-\theta) \leq \frac{1}{4} \quad \text{and} \quad \theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}} \left. \right)^{2\gamma-2} \leq \left( \frac{1}{2} \right)^{\frac{2-2\gamma}{\gamma}} < \infty.$$

Letting

$$h_n(t, \theta) := \frac{f_{\frac{R}{r_n}, \Theta|_{r_n}}(t, \theta) - f_{\frac{R}{r_n}|_{r_n}}(t)f_{\Theta|_{r_n}}(\theta)}{r_n^{-1}},$$

we have

$$\begin{aligned}
 & \max \left( \int_0^1 \int_1^{\infty} |h_n(t, \theta)| dt d\theta, \int_0^1 \int_1^{\infty} |\log(t)h_n(t, \theta)| dt d\theta \right) \left\{ \right. \\
 & \leq \int_0^1 \int_1^{\infty} |th_n(t, \theta)| dt d\theta \\
 & \leq \int_0^1 \int_1^{\infty} \left| \frac{f_{\frac{R}{r_n}, \Theta|_{r_n}}(t, \theta) - f_T(t)f_{\Theta}(\theta)}{t^{-1}r_n^{-1}} \right| dt d\theta \\
 & \quad + \int_0^1 \int_1^{\infty} \left| f_T(t) \frac{f_{\Theta|_{r_n}}(\theta) - f_{\Theta}(\theta)}{t^{-1}r_n^{-1}} \right| dt d\theta \\
 & \quad + \int_0^1 \int_1^{\infty} \left| f_{\Theta|_{r_n}}(\theta) \frac{f_{\frac{R}{r_n}|_{r_n}}(t) - f_T(t)}{t^{-1}r_n^{-1}} \right| dt d\theta,
 \end{aligned}$$

where the first term can be bounded by

$$\int_0^1 \int_1^{\infty} ct^{-2} dt d\theta < \infty,$$

and the other terms can be bounded in the same way. Since  $R$  has infinite first moment, we apply the distance correlation to  $\log R$  and  $\Theta$ . The integral in Eq. 17 is bounded by

$$\begin{aligned} & \frac{np_n}{r_n^2} \int_0^1 \int_1^\infty \int_0^1 \int_1^\infty |\log t - \log t| |\theta - \theta| |h_n(t, \theta)| |h_n(t, \theta)| dt d\theta dt d\theta \\ & \leq c \frac{n}{r_n^3} \int_0^1 \int_1^\infty \int_0^1 \int_1^\infty (|\log t| + |\log t|) |h_n(t, \theta)| |h_n(t, \theta)| dt d\theta dt d\theta \\ & \leq c \frac{n}{r_n^3} \left( \int_0^1 \int_1^\infty |\log(t) h_n(t, \theta)| dt d\theta \right) \left( \int_0^1 \int_1^\infty |h_n(t, \theta)| dt d\theta \right) \leq c \frac{n}{r_n^3}, \end{aligned}$$

which converges to zero if  $n = o(r_n^3)$ . Therefore if  $\{r_n\}$  is chosen such that  $r_n = o(n)$  and  $n = o(r_n^3)$ , then Theorem 1 holds.

The result in Theorem 1 can be generalized from iid to a regularly varying time series setting, which we present in the next theorem. For a multivariate stationary time series  $\{\mathbf{X}_t\}$  and  $h \geq 1$ , set  $\mathbf{Y}_h = (\mathbf{X}_0, \dots, \mathbf{X}_h)$ . Then  $\{\mathbf{X}_t\}$  is regularly varying if

$$\frac{\mathbb{P}(x^{-1} \mathbf{Y}_h \in \cdot)}{\mathbb{P}(x^{-1} \|\mathbf{X}_0\| > 1)} \xrightarrow{v} \mu_h^*(\cdot), \quad x \rightarrow \infty,$$

for some non-null measure  $\mu_h^*$  on  $\overline{\mathbb{R}}_0^{(h+1)d} = \overline{\mathbb{R}}^{(h+1)d} \setminus \{\mathbf{0}\}$ ,  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ , with the property that  $\mu_h^*(tC) = t^{-\alpha} \mu_h^*(C)$  for any  $t > 0$  and Borel set  $C \subset \overline{\mathbb{R}}_0^{(h+1)d}$ . See, for example, page 979 of Davis and Mikosch (2009). It follows easily that

$$\frac{\mathbb{P}(x^{-1}(\mathbf{X}_0, \mathbf{X}_h) \in \cdot)}{\mathbb{P}(\|\mathbf{X}_0\| > x)} \xrightarrow{v} \mu_h(\cdot), \quad (20)$$

where

$$\mu_h(D) = C \cdot \mu_h^*({\mathbf{s}} \in \overline{\mathbb{R}}^{(h+1)d} : ({\mathbf{s}}_1, {\mathbf{s}}_h) \in D).$$

Assume that  $\{\mathbf{X}_t\}$  is  $\alpha$ -mixing. We assume the following conditions between  $\{\mathbf{X}_t\}$  and the sequence of threshold  $\{r_n\}$ , which can be verified for various time series models (Davis and Mikosch 2009).

(M) Assume  $p_n^{-1} = \mathbb{P}^{-1}(\|\mathbf{X}_1\| > r_n) = o(n^{1/3})$  and that there exists a sequence  $\{l_n\}$  such that  $l_n \rightarrow \infty$ ,  $l_n p_n \rightarrow 0$ , and

i)

$$\left( \frac{1}{p_n} \right) \Big|_{h=l_n}^{\infty} \alpha_h^\delta \rightarrow 0 \text{ for some } \delta \in (0, 1); \quad (21)$$

ii)

$$\lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{p_n} \prod_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) = 0; \quad (22)$$

iii)

$$np_n \alpha_{l_n} \rightarrow 0. \quad (23)$$

**Theorem 2** Let  $\{\mathbf{X}_t\}$  be a multivariate regularly varying time series with tail index  $\alpha > 1$  and  $\alpha$ -mixing with coefficients  $\{\alpha_h\}_{h \geq 0}$ . Assume the same conditions for the weight measure  $\mu$  and the sequence of thresholds  $\{r_n\}$  in Theorem 1, i.e., Eqs. 12, 13 hold, and that condition (M) holds. Then

$$n \hat{p}_n T_n \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt),$$

where  $Q$  is a centered Gaussian process. In particular,

$$T_n \xrightarrow{p} 0.$$

The proof of Theorem 2 is given in Appendix B.

Note that the limiting distributions  $Q$  in Theorem 1 and  $Q$  in Theorem 2 are both intractable. In practice, quantiles of the distributions are calculated using resampling methods. While in the iid case this can be done straightforwardly, in the weakly dependent case one needs to apply the block bootstrap or stationary bootstrap to obtain the desired result (see Davis et al. (2012)). In the following section, we present a threshold selection framework with a subsampling scheme that does not require independence between the observations.

## 5 Threshold selection

In this section, we propose a procedure to select the threshold for estimating the spectral measure  $S$  from observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . Let us first consider the case where a specific threshold  $r_n$  is given. Then Eq. 9 specifies the empirical distance covariance between  $R/r_n$  and  $\Theta$  conditional on  $R > r_n$ . Under the assumption (13), we have from Theorem 1,

$$n \hat{p}_n T_n \rightarrow \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt),$$

where  $n \hat{p}_n$  is the number of observations such that  $R_i > r_n$ . In practice, the limit distribution  $|Q|^2 \mu(s, t)$  is intractable, but one can resort to bootstrapping. Consider the hypothesis testing framework:

$$H_0 : R/r_n \text{ and } \Theta \text{ are independent given } R > r_n$$

$$H_1 : R/r_n \text{ and } \Theta \text{ are not independent given } R > r_n.$$

Define the  $p$ -value for testing  $H_0$  versus  $H_1$  to be

$$pv = \mathbb{P} \left( \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt) > u \right) \Big|_{u=n \hat{p}_n T_n}. \quad (24)$$

Then under  $H_0$ ,  $pv$  follows  $U(0, 1)$ , while under  $H_1$ ,  $pv$  should be sufficiently small.

Now consider a decreasing sequence of candidate thresholds  $\{r_k\}$ . From Eq. 24, a sequence of  $p$ -values  $\{pv_k\}$ , each corresponding to a threshold  $r_k$ , can be obtained. Our goal is to find the smallest threshold  $r^*$  such that conditional on  $R > r^*$ ,  $\Theta$  can reasonably be considered independent of  $R$ . Note that the  $pv_k$ 's are not independent for each  $k$  since they are computed from the same set of data. Conventional multiple

testing procedures, such as Bonferroni correction, are problematic to implement for dependent  $p$ -values. To counter these limitations, we propose an intuitive and direct method based on subsampling.

The idea is outlined as follows: For a fixed level  $r_k$ , we choose a subsample of size  $n_k$  from the conditional empirical cdf  $\hat{F}_{r_n, \Theta|r_k}^R$  of  $(R_i/r_k, \Theta_i)$  with  $R_i > r_k, i = 1, \dots, n$ . For this subsample, we compute the distance covariance  $T_{n,k}$ . To compute a  $p$ -value of  $T_{n,k}$  under the assumption that the conditional empirical distribution is a product of the conditional marginals, we take a large number ( $L$ ) of subsamples of size  $n_k$  from

$$\tilde{F}_{r_n, \Theta|r_k}^R(d\theta, dr) = \hat{F}_{\Theta|r_k}(d\theta) \hat{F}_{r_n}^R(dr),$$

and calculate the value  $\tilde{T}_{n,k}^{(l)}, l = 1, \dots, L$  for each subsample. The  $p$ -value of  $T_{n,k}$ ,  $pv_k$ , is then the empirical  $p$ -value of  $T_{n,k}$  relative the  $\{\tilde{T}_{n,k}^{(l)}\}_{l=1, \dots, L}$ . This process, starting with an initial subsample of  $n_k$  from  $\hat{F}_{r_n, \Theta|r_k}^R$  is repeated  $m$  times, which produces  $m$  estimates  $\{pv_k^{(j)}\}_{j=1, \dots, m}$  of the  $pv_k$ , which are independent conditional on the original sample. These are then averaged

$$\overline{pv}_k = \frac{1}{m} \sum_{j=1}^m pv_k^{(j)}.$$

So for the sequence of levels  $\{r_k\}$ , we produce a sequence of independent  $p$ -values  $\{\overline{pv}_k\}$ .

Our choice of threshold  $r$  at which  $(\Theta, R)|R > r$  are independent (and dependent otherwise) will be based on an examination of the path of the mean  $p$ -values,  $\{\overline{pv}_k\}$ . Note the following two observations:

- If  $R$  and  $\Theta$  are independent given  $R > r_k$ , then the  $pv_k^{(1)}, \dots, pv_k^{(m)}$  will be iid and approximately  $U(0, 1)$ -distributed, so that  $\overline{pv}_k$  should center around 0.5.
- If  $R$  and  $\Theta$  are dependent given  $R > r_k$ , then the  $pv_k^{(j)}$ 's will be well below 0.5 (closer to 0), and so will  $\overline{pv}_k$ .

By studying the sequence  $\{\overline{pv}_k\}$ , which we call the mean  $p$ -value path, we choose the threshold to be the smallest  $r_k$  such that  $\overline{pv}_l$  is around 0.5 for  $l < k$ . A well-suited change-point method for our situation is the CUSUM algorithm, by Page (1954), which detects the changes in mean in a sequence by looking at mean-corrected partial sums. In our algorithm, we use a spline fitting method that is based on the CUSUM approach called wild binary segmentation (WBS), proposed by Fryzlewicz (2014). The WBS procedure uses the CUSUM statistics of subsamples and fits a piecewise constant spline to  $\{\overline{pv}_k\}$ . In our setting, we may choose  $r_k$  to be the knot of the spline after which the fitted value is comfortably below 0.5.

There are several advantages to using the subsampling scheme. First, recall that the  $p$ -value path  $\{pv_k\}$ , which is obtained from the whole data set, has complicated serial structure and varies greatly from each realization. In contrast, the mean  $p$ -values  $\overline{pv}_k$  from subsampling are conditionally independent and will center around 0.5 with small variance when the total sample size  $n$  and the number of subsample  $m$  is large. This, in turns, helps to present a justifiable estimation for the threshold.

Second, the calculation of distance covariance can be extremely slow for moderate sample size. Using smaller sample sizes for the subsamples, our computational burden is greatly reduced. In addition, this procedure is amenable to parallel computing, reducing the computation time even further. Third, the subsampling makes it possible to accommodate stationary but dependent data, waiving the stringent independent assumption.

The idea of looking at the mean  $p$ -value path is inspired by Mallik et al. (2011), which used the mean of  $p$ -values from multiple independent tests to detect change points in population means.

## 6 Data illustration

In this section, we demonstrate our threshold selection method through simulated and real data examples.

In practice, we set the sequence of thresholds  $\{r_k\}$  to be the corresponding upper quantiles to  $\{q_k\}$ , a pre-specified sequence of quantile levels. The subsample size  $n_k$  at each threshold  $r_k$  is set as  $n_k = n_0 \cdot q_k$  for some  $n_0 \ll n$ . This is designed such that for any  $r_k$ , each subsample is a  $n_0/n$  fraction of all the eligible data points with  $R > r_k$ . Then the choice of  $\{n_k\}$  boils down to the choice of  $n_0$ , which should reflect the following considerations: i)  $n_0$  should be large enough to ensure good resolution of  $p$ -values at all levels; ii)  $n_0/n$  should be sufficiently small such that the subsamples do not contain too much overlap in observations; iii) larger  $n_0$  requires heavier computation for the distance correlation. In our examples, where the total sample size  $n$  ranges from 3000 to 20000, we find  $n_0$  between 500 and 1000 to be a suitable choice. The number of subsamples  $m$  can be set as large as computation capacity allows. In our examples, we take  $m = 60$ .

For all the examples, we choose the weight function  $\mu$  for distance covariance to be Eq. 6 with  $\kappa = 1$ , and the number of replications used to calculate each  $p$ -value is  $L = 200$ . To ensure that the moment conditions are met, the distance correlation is applied to the log of the radial part  $R$  in all examples.

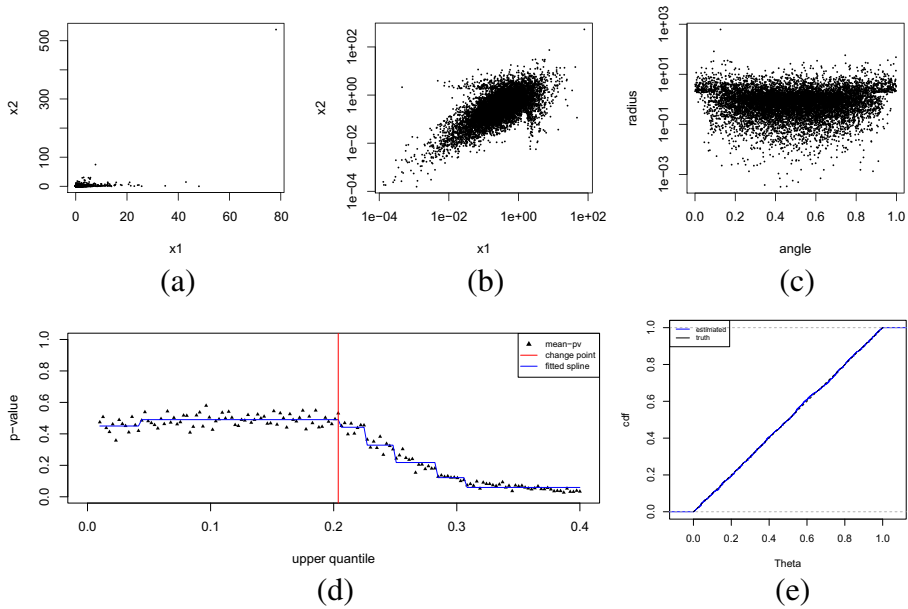
### 6.1 Simulated data with known threshold

To illustrate our methodology, we simulate observations from a distribution with a known threshold for which  $R$  and  $\Theta$  become independent.

Let  $R$  be the absolute value of a  $t$ -distribution with 2 degrees of freedom and  $\Theta_1, \Theta_2$  be independent random variables such that  $\Theta_1 \sim U(0, 1)$ ,  $\Theta_2 \sim \text{Beta}(3, 3)$ . Set

$$\Theta|R = \begin{cases} \Theta_1, & \text{if } R > r_{0.2}, \\ \Theta_2, & \text{if } R \leq r_{0.2}, \end{cases}$$

where  $r_{0.2}$  is the upper 20%-quantile of  $R$ . Then  $R$  and  $\Theta$  are independent given  $R > r$  if and only if  $r \geq r_{0.2}$ . Let  $(X_{i1}, X_{i2}) = (R_i \Theta_i, R_i(1 - \Theta_i))$ ,  $i = 1, \dots, n$ , be the simulated observations. We generate  $n = 10000$  iid observations from this distribution. Figure 1a, b and c show the data in Cartesian and polar coordinates. Our goal is to recover the tail angular distribution by choosing the appropriate threshold.



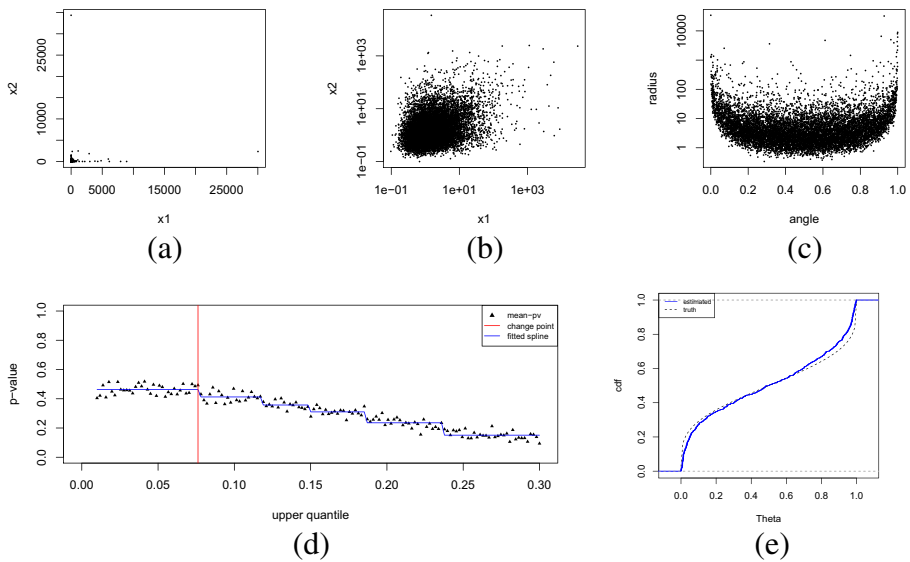
**Fig. 1** Example 6.1. **a** scatterplot of  $(X_{i1}, X_{i2})$ ; **b** scatterplot of  $(X_{i1}, X_{i2})$  in log-log scale; **c** scatterplot of  $(R_i, \Theta_i)$ ; **d** mean  $p$ -value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); **e** estimated cdf of  $\Theta$  using the threshold chosen, compared with the truth (black dotted)

A sequence of candidate thresholds  $\{r_k\}$  is selected to be the empirical upper quantiles of  $R$  corresponding to  $\{q_k\}$ , 150 equidistant points between 0.01 and 0.4. We apply the procedure described in Section 5 to the data. For each  $r_k$ , the mean  $p$ -value  $\overline{pv}_k$  is calculated using  $m = 60$  random subsamples, each of size  $n_k = 500 \cdot q_k$ , from the observations with  $R_i > r_k$ . Figure 1d shows the mean  $p$ -value path. For the WBS algorithm, we set the threshold to be the largest  $r_k$  such that for all thresholds  $r$  (quantile level  $q$ ) such that  $r < r_k$  ( $q > q_k$ ), the fitted spline of the  $p$ -value stays below 0.45<sup>1</sup>. The threshold levels chosen is 20.4%, which are in good agreement with the true independence level 0.2. The empirical cdfs of the truncated  $\Theta_i$ 's corresponding to the chosen thresholds is shown in Fig. 1e. We can see that the true tail angular cdf (i.e.,  $U(0, 1)$ ) is accurately recovered.

## 6.2 Simulated logistic data

We simulate data from a bivariate logistic distribution, which is bivariate regularly varying. Recall from Example 1 that  $(X_1, X_2)$  follows a bivariate logistic distribution

<sup>1</sup>Of course, other selection rules can be used. For example, a more conservative approach would be choosing the threshold as the largest  $r_k$  such that for  $r > r_k$ , the fitted spline of the  $p$ -value stays above 0.45.



**Fig. 2** Example 6.2. **a** scatterplot of  $(X_{i1}, X_{i2})$ ; **b** scatterplot of  $(X_{i1}, X_{i2})$  in log-log scale; **c** scatterplot of  $(R_i, \Theta_i)$ ; **d** mean  $p$ -value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); **e** estimated cdf of  $\Theta$  using the threshold chosen, compared with the theoretical limiting cdf (black dotted)

if it has cdf (18). In this example, we set  $\gamma = 0.8$  and generate  $n = 10000$  iid observations from this distribution. Similar to the previous example, for each threshold  $r_k$  corresponding to the upper  $q_k$  quantile, where  $\{q_k\}$  is chosen to be the 150 equidistant points between 0.01 and 0.3. The mean  $p$ -value  $\bar{p}_{v_k}$  is calculated using  $m = 60$  random subsamples of size  $n_k = 500 \cdot q_k$  from the observations with  $R_i > r_k$ .

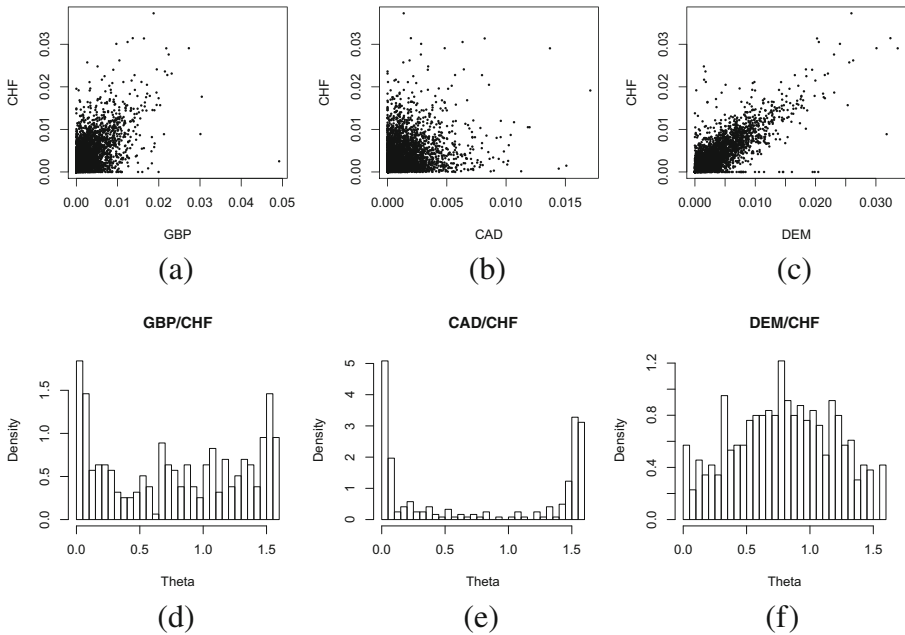
Figure 2a, b and c show the scatterplots of the data. Here the  $L_1$ -norm is used to transform the data into polar coordinates. Our algorithms suggests using 7.4% of the data to estimate the angular distribution. The estimated cdf of the angular distribution is shown with the theoretical limiting cdf, derived from Eq. 19, in Fig. 2e. So even though  $R$  and  $\Theta$  are not independent for any threshold  $r_k$ , our procedures produce good estimates of the limiting distribution of  $\Theta$ .

### 6.3 Real data

In this example, we look at the following exchange rate returns relative to the US dollar: Deutsche mark (DEM), British pound (GBP), Canadian dollar (CAD), and Swiss franc (CHF). The time spans for the data are 1990-01-01 to 1998-12-31 with a total of 3287 days of observations. We examine the pairs GBP/CHF, CAD/CHF, DEM/CHF and estimate the angular density in the tail for each pair. Figure 3a–c present the scatter plots of the data. The marginals of the observations are standardized using the rank transformation proposed in Joe et al. (1992):

$$Z_i = 1/\log\{n/(\text{Rank}(X_i) - .5)\}, \quad i = 1, \dots, n.$$





**Fig. 3** Example 6.3. Analysis of the paired exchange rate returns: CHF/DEM, CHF/GBP, CHF/CAD with respect to USD between 1990-01-01 to 1998-12-31. **a–c**: Scatter plots of the standardized paired exchange rate returns; **d–f**: Estimated angular densities using the estimated thresholds chosen

Again  $\{q_k\}$  is chosen to be the 150 equidistant points between 0.01 and 0.3, and the mean  $p$ -value  $\bar{p}v_k$  is calculated using  $m = 60$  random subsamples of size  $n_k = 500 \cdot q_k$  from the observations with  $R_i > r_k$ . Note that while it may not be reasonable to view the observations as iid, the subsampling scheme can still be applied to choose the threshold of independence between  $R$  and  $\Theta$ .

The mean  $p$ -value paths are shown in Fig. 4a–c. The threshold levels selected for the three pairs are 9.6%, 7.4%, 16%, respectively. Figure 3d–f show the shape of the estimated angular densities for each pairs. As expected, the tails of the two central European exchange rates, DEM and CHF, are highly dependent. In contrast, that of CAD and CHF are almost independent.

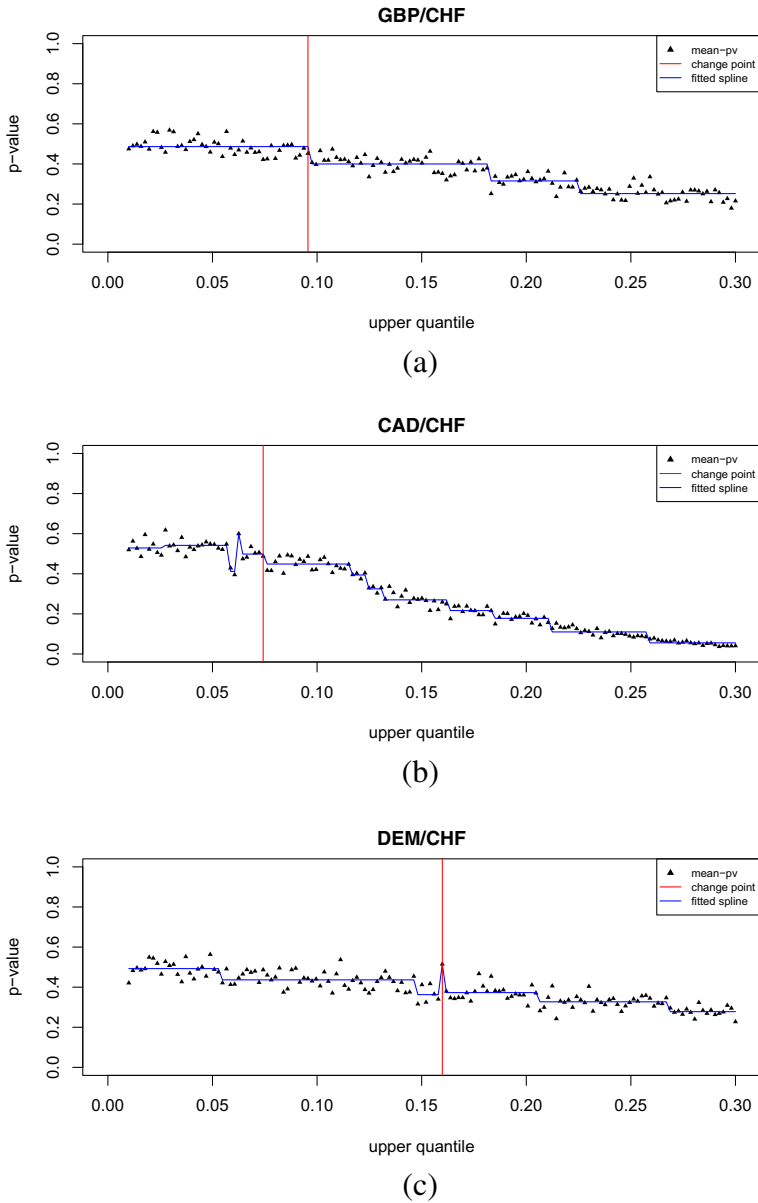
## 6.4 Simulated non-regularly varying data

In this example, we generate data from a model which is not regularly varying. Let  $R$  be a random variable from the standard Pareto distribution:

$$\mathbb{P}(R > r) = r^{-1}, \quad r \geq 1.$$

Let  $\Theta_1, \Theta_2$  be independent random variables such that  $\Theta_1 \sim U(0, 0.5)$ ,  $\Theta_2 \sim U(0.5, 1)$ . Set

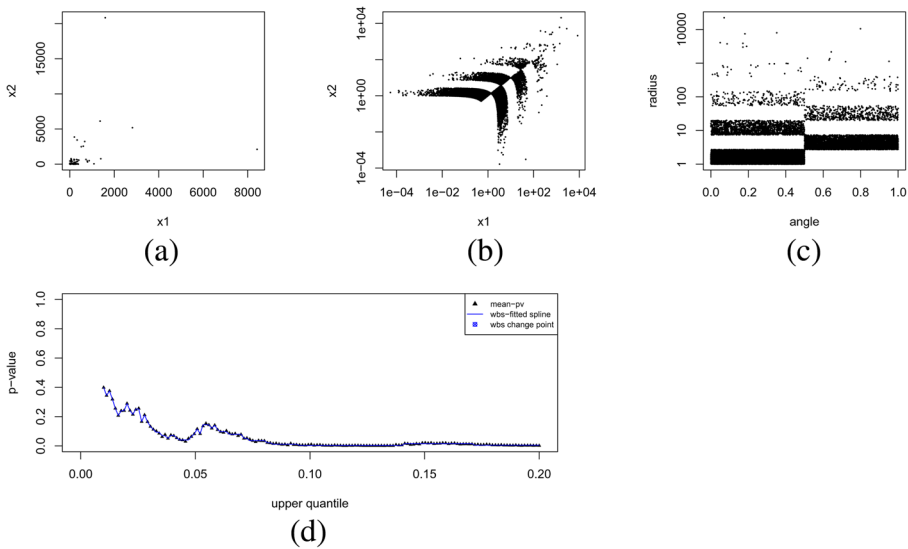
$$\Theta|R \sim \begin{cases} \Theta_1, & \text{if } \log R \in (2k, 2k+1] \text{ for some integer } k, \\ \Theta_2, & \text{if } \log R \in (2k+1, 2k+2] \text{ for some integer } k. \end{cases}$$



**Fig. 4** Example 6.3 (cont.). Analysis of the paired exchange rate returns: CHF/DEM, CHF/GBP, CHF/CAD with respect to USD between 1990-01-01 to 1998-12-31. **a–c**: mean  $p$ -value paths (black triangles), fitted WBS splines (blue lines) and the chosen threshold quantiles (red vertical line)

For any positive integer  $k$ , it can be verify that

$$\mathbb{P}(\Theta \in (0, 0.5) | R > e^{2k}) = \frac{1 - e^{-1}}{1 - e^{-2}},$$



**Fig. 5** Example 6.4. **a** scatterplot of  $(X_{i1}, X_{i2})$ ; **b** scatterplot of  $(X_{i1}, X_{i2})$  in log-log scale; **c** scatterplot of  $(R_i, \Theta_i)$ ; **d** mean  $p$ -value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line)

while

$$\mathbb{P}(\Theta \in (0, 0.5) | R > e^{2k+1}) = \frac{e^{-1} - e^{-2}}{1 - e^{-2}}.$$

Hence  $\mathbb{P}(\Theta \in \cdot | R > r)$  does not convergence as  $r \rightarrow \infty$  and  $\mathbf{X} = (R\Theta, R(1 - \Theta))$  is not regularly varying.

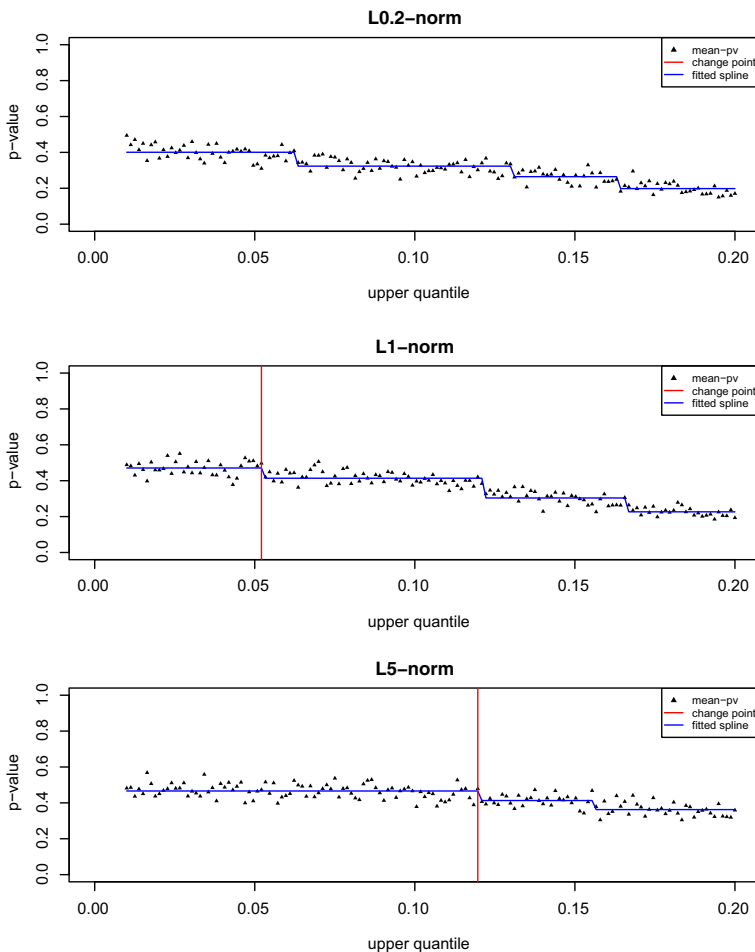
Let  $(X_{i1}, X_{i2}) = (R_i\Theta_i, R_i(1 - \Theta_i))$ ,  $i = 1, \dots, n$ , be iid observations from this distribution, where  $n = 20000$ . Figure 5a, b and c show the data in Cartesian and polar coordinates. We apply our threshold selection algorithm to the data, with the threshold upper quantile levels  $q_k$  chosen as the 150 equidistant points between 0.01 and 0.2. The mean  $p$ -value  $\overline{p}v_k$  is calculated using  $m = 60$  random subsamples of size  $n_k = 500 \cdot q_k$  from the observations with  $R_i > r_k$ . This is shown in Fig. 5d.

In this model, the radial part  $R$  is regularly varying, but  $\Theta$  and  $R$  are dependent given  $R > r$  for any  $r$ . We expect the mean  $p$ -values to be well below 0.5, as are observed. No threshold is selected by the algorithm. This suggests that our technique can potentially be used to detect misspecified models from the regular variation assumption, especially in the scenario where the heavy-tailedness of  $R$  is observed but dependence between  $R$  and  $\Theta$  is suspected.

## 7 Discussion

In this paper, we propose a threshold selection procedure for multivariate regular variation, for which  $R$  and  $\Theta$  are approximately independent for  $R$  beyond the threshold.

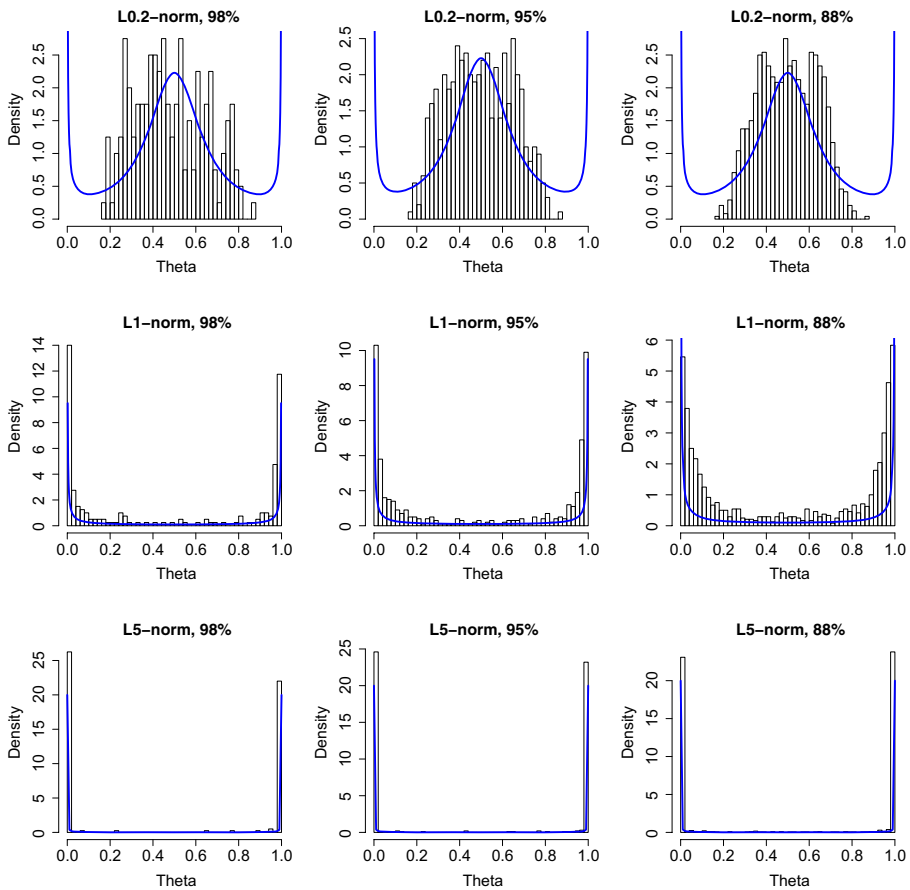
While our problem is set in the multivariate heavy-tailed setting and we utilize distance covariance as our measure of dependence, our algorithm is essentially a change point detection method based on  $p$ -values generated through subsampling schemes. Hence this may be generalized to other problem settings and potentially incorporates other dependence measures. Though we have proposed an automatic selection for the threshold based on the fitted mean  $p$ -value path, we would like to emphasize that, like the Hill plot, this should be viewed as a visual tool rather than an optimal selection criterion. The final threshold should be based on the automatic procedure in conjunction with visual inspection of the  $p$ -value path.



**Fig. 6** Simulated logistic data of sample size  $n = 10000$  with  $\gamma = 0.95$ . Threshold selection algorithm applied under the  $L_{0.2}$ -,  $L_1$ - and  $L_5$ -norms: mean  $p$ -value paths (black triangles), fitted WBS splines (blue lines) and the chosen threshold quantiles (red vertical line)

We note that the choice of norm in the polar coordinate transformation (1) may result in significant differences in the choice of thresholds, which indicates the rate of convergence to the limit spectral density. This is especially evident in the near ‘asymptotic independence’ case, where the mass of the angular distribution concentrates on the axes.

As an illustration, we simulated iid observations  $\{(X_{i1}, X_{i2})\}_{i=1,\dots,n}$  from the bivariate logistic distribution, where the cdf is given in Eq. 18, with  $\gamma = 0.95$  and  $n = 10000$ . We apply the polar coordinate transformation with respect to the  $L_p$ -norm for  $p = 0.2, 1, 5$ . Note that in the case of  $p = 0.2$ ,  $L_p$  is only a quasi-norm as it does not satisfy the triangular inequality. However, it can be shown that Eq. 4 holds and the limiting angular distribution exists for bivariate logistic distribution. We compare the threshold selection results in Fig. 6. Note that in the cases of the



**Fig. 7** Simulated logistic data of sample size  $n = 10000$  with  $\gamma = 0.95$ . Histogram of  $X_1^p / (X_1^p + X_2^p)$  for truncated levels 2%, 5% and 12% for  $p = 0.2, 1, 5$

$L_1$  and  $L_5$ -norms, the threshold levels are chosen to be upper 5% and 12%, respectively, while in the case of the  $L_{0.2}$ -norm, it is not possible to select the threshold as the dependence between  $R$  and  $\Theta$  at all levels were shown to be significant. Indeed, this can be seen in Fig. 7, where we compare the histogram of  $X_1^p/(X_1^p + X_2^p)$  given  $\|X\|_p$  is large across three levels of truncations, 2%, 5% and 12%, together with the theoretical limiting density curve. For the  $L_{0.2}$ -norm, the limiting angular density is poorly approximated by the truncated data for all levels. For the other two norms, the truncated observations according to the selected threshold provide decent approximations to the true limiting density of the angular component. One possible explanation for this is that under the  $L_{0.2}$ -norm, the threshold is concave and hence observations on the diagonal are much easier to be classified as “extremes” than those near the axis. As a result, the estimator of the angular density uses more observations near the diagonal, which may not be, in fact, close enough to the limit. This choice of norm is an interesting topic and is the subject of ongoing research.

**Acknowledgements** The foreign exchange rate data are obtained from OANDA from through R-package ‘qrmtools’. We would like to thank Bodhisattva Sen for helpful discussions. We would also like to thank the editor and referees for their many constructive and insightful comments.

## Appendix A: Proof of Theorem 1

Note from the definition of the empirical distance covariance in Eq. 9, the integrand can be expressed as

$$\begin{aligned} C_n(s, t) &= \frac{1}{n\hat{p}_n} \sum_{j=1}^n e^{isR_j/r_n + it^T\Theta_j} \mathbf{1}_{\{R_j > r_n\}} \\ &\quad - \frac{1}{n\hat{p}_n} \sum_{j=1}^n e^{isR_j/r_n} \mathbf{1}_{\{R_j > r_n\}} \frac{1}{n\hat{p}_n} \sum_{k=1}^n e^{it^T\Theta_k} \mathbf{1}_{\{R_k > r_n\}} \\ &= \frac{1}{n\hat{p}_n} \sum_{j=1}^n e^{isR_j/r_n - \varphi_{\frac{R}{r_n}|r_n}(s)} \left( e^{it^T\Theta_j} - \varphi_{\Theta|r_n}(t) \right) \mathbf{1}_{\{R_j > r_n\}} \\ &\quad - \frac{1}{n\hat{p}_n} \sum_{j=1}^n e^{isR_j/r_n - \varphi_{\frac{R}{r_n}|r_n}(s)} \left\{ \mathbf{1}_{\{R_j > r_n\}} \frac{1}{n\hat{p}_n} \sum_{k=1}^n e^{it^T\Theta_k} - \varphi_{\Theta|r_n}(t) \right\} \mathbf{1}_{\{R_k > r_n\}}. \end{aligned}$$

Writing  $U_{jn} = e^{isR_j/r_n - \varphi_{\frac{R}{r_n}|r_n}(s)} \mathbf{1}_{\{R_j > r_n\}}$ ,  $V_{jn} = e^{it^T\Theta_j} - \varphi_{\Theta|r_n}(t) \mathbf{1}_{\{R_j > r_n\}}$ , we have

$$C_n(s, t) = \frac{p_n}{\hat{p}_n} \frac{1}{n} \sum_{j=1}^n \frac{U_{jn} V_{jn}}{p_n} - \left( \frac{p_n}{\hat{p}_n} \right)^2 \frac{1}{n} \sum_{j=1}^n \frac{U_{jn}}{p_n} \frac{1}{n} \sum_{k=1}^n \frac{V_{kn}}{p_n}.$$

Since  $\mathbb{E}U_{jn} = \mathbb{E}V_{jn} = 0$  and  $\mathbb{E}U_{jn}V_{jn}/p_n = \varphi_{\frac{R}{r_n}, \Theta|r_n}(s, t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)$ , it is convenient to mean correct the summands and obtain

$$\begin{aligned} C_n(s, t) &= \frac{p_n}{\hat{p}_n} \frac{1}{n} \sum_{j=1}^n \left( \frac{U_{jn}V_{jn}}{p_n} - \varphi_{\frac{R}{r_n}, \Theta|r_n}(s, t) + \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t) \right) \left\{ \right. \\ &\quad \left. - \left( \frac{p_n}{\hat{p}_n} \right)^2 \frac{1}{n} \sum_{j=1}^n \frac{U_{jn}}{p_n} \frac{1}{n} \sum_{k=1}^n \frac{V_{kn}}{p_n} \right. \\ &\quad \left. + \frac{p_n}{\hat{p}_n} (\varphi_{\frac{R}{r_n}, \Theta|r_n}(s, t) - \varphi_{\frac{R}{r_n}|r_n}(s)\varphi_{\Theta|r_n}(t)) \right\} \\ &=: \left( \frac{p_n}{\hat{p}_n} \right) \left\{ \tilde{E}_1 - \left( \frac{p_n}{\hat{p}_n} \right)^2 \tilde{E}_{21} \tilde{E}_{22} + \left( \frac{p_n}{\hat{p}_n} \right) \tilde{E}_3 \right\} \\ &=: \left( \frac{p_n}{\hat{p}_n} \right) \left\{ \tilde{E}_1 - \left( \frac{p_n}{\hat{p}_n} \right)^2 \tilde{E}_2 + \left( \frac{p_n}{\hat{p}_n} \right) \tilde{E}_3 \right\} \end{aligned}$$

Note that  $\tilde{E}_1, \tilde{E}_{21}, \tilde{E}_{22}$  are averages of iid zero-mean random variables and  $\tilde{E}_3$  is non-random. We first prove the second part of Theorem 1. The first part of Theorem 1 follows easily in a similar fashion.

*Proof (Proof of Theorem 1(2))*

In order to show (14), it suffices to establish that

$$n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left( \frac{p_n}{\hat{p}_n} \right)^2 |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt), \quad (25)$$

and

$$\left| n\hat{p}_n T_n - n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left( \frac{p_n}{\hat{p}_n} \right)^2 |\tilde{E}_1|^2 \mu(ds, dt) \right| \xrightarrow{p} 0, \quad (26)$$

where Eq. 26 can be implied by

$$n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left( \frac{p_n}{\hat{p}_n} \right)^2 |\tilde{E}_2|^2 \mu(ds, dt) + n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left( \frac{p_n}{\hat{p}_n} \right)^2 |\tilde{E}_3|^2 \mu(ds, dt) \xrightarrow{p} 0. \quad (27)$$

Notice that

$$\begin{aligned} \mathbb{E} \left| \frac{\hat{p}_n}{p_n} - 1 \right|^2 &= \mathbb{E} \left| \frac{1}{n} \sum_{j=1}^n \left( \frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} - 1 \right) \right|^2 = \frac{1}{n} \mathbb{E} \left| \frac{\mathbf{1}_{\{R_1 > r_n\}}}{p_n} - 1 \right|^2 \\ &\leq \frac{1}{np_n} O(1) + \frac{1}{n} O(1) \rightarrow 0. \end{aligned}$$

Hence  $\hat{p}_n/p_n \xrightarrow{p} 1$  and for Eqs. 25 and 27, it is equivalent to prove that

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt) \quad (28)$$

and

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds, dt) + np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_3|^2 \mu(ds, dt) \xrightarrow{p} 0. \quad (29)$$

We will show the convergence (28) in Proposition 1. By Eq. 13,

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_3|^2 \mu(ds, dt) \rightarrow 0.$$

So that Eq. 29 holds provided

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds, dt) \xrightarrow{p} 0, \quad (30)$$

which follows in a similar fashion as Proposition 1.  $\square$

**Proposition 1** Assume  $\mu$  satisfies

$$\int_{\mathbb{R}^{d+1}} (1 \wedge |s|^\beta)(1 \wedge |t|^2) \mu(ds, dt) < \infty,$$

and that  $np_n \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt),$$

where  $Q$  is a centered Gaussian process with covariance function (15).

*Proof (Proof of Proposition 1)*

We first show that

$$\sqrt{np_n} \tilde{E}_1 \xrightarrow{d} Q(s, t), \quad \text{on } \mathcal{C}(\mathbb{R}^{d+1}) \quad (31)$$

which can be implied by the finite distributional convergence of  $\sqrt{np_n} \tilde{E}_1(s, t)$  and its tightness on  $\mathcal{C}(\mathbb{R}^{d+1})$ .

Write

$$\sqrt{np_n} \tilde{E}_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left( \frac{U_{jn} V_{jn}}{\sqrt{p_n}} - \sqrt{p_n} (\varphi_{\frac{R}{r_n}, \Theta|_{r_n}}(s, t) - \varphi_{\frac{R}{r_n}|_{r_n}}(s) \varphi_{\Theta|_{r_n}}(t)) \right) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_{jn},$$

where  $Y_{jn}$ 's are iid random variables with mean 0. For fixed  $(s, t)$ , note that

$$\text{Var}(Y_{1n}) = \mathbb{E}|Y_{1n}|^2 = \frac{\mathbb{E}|U_{1n} V_{1n}|^2}{p_n} (1 + o(1)) = \frac{\mathbb{E} \mathbf{1}_{\{R_1 > r_n\}}}{p_n} O(1) < \infty.$$

On the other hand, any  $\delta > 0$ ,

$$\mathbb{E}|Y_{1n}|^{2+\delta} = \frac{\mathbb{E}|U_{1n} V_{1n}|^{2+\delta}}{p_n^{1+\delta/2}} (1 + o(1)) \leq c \frac{\mathbb{E} \mathbf{1}_{\{R_1 > r_n\}}}{p_n^{1+\delta/2}} (1 + o(1)) = O(p_n^{-\delta/2})$$

Then we can apply the central limit theorem for triangular arrays by checking the Lyapounov condition (see, e.g., Billingsley (1995)) for the  $Y_{jn}$ 's:

$$\frac{\sum_{j=1}^n \mathbb{E}|Y_{jn}|^{2+\delta}}{\left( \text{Var} \sum_{j=1}^n Y_{jn} \right)^{\frac{2+\delta}{2}}} = \frac{O(np_n^{-\frac{\delta}{2}})}{n^{1+\frac{\delta}{2}} \text{Var}(Y_{1n})^{1+\frac{\delta}{2}}} = O((np_n)^{-\frac{\delta}{2}}) \rightarrow 0.$$



It follows easily that for fixed  $(s, t)$ ,

$$\sqrt{np_n} \tilde{E}_1 \xrightarrow{d} Q(s, t).$$

The finite-dimensional distribution can be obtained using the Cramér-Wold device and the covariance function can be verified through calculations.

We now show the tightness of  $\sqrt{np_n} \tilde{E}_1$ . Note that

$$\begin{aligned} \tilde{E}_1(s, t) &= \frac{1}{n} \sum_{j=1}^n \frac{e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s)}{p_n} \left( \left( e^{it^T \Theta_j} - \varphi_{\Theta|r_n}(t) \right) \mathbf{1}_{\{R_j > r_n\}} \right. \\ &\quad \left. - \varphi_{\frac{R}{r_n}, \Theta|r_n}(s, t) - \varphi_{\frac{R}{r_n}|r_n}(s) \varphi_{\Theta|r_n}(t) \right) \Bigg\} \\ &= \left( \left( \frac{1}{n} \sum_{j=1}^n \frac{e^{isR_j/r_n + it^T \Theta_j} \mathbf{1}_{\{R_j > r_n\}}}{p_n} - \varphi_{\frac{R}{r_n}, \Theta|r_n}(s, t) \right) \right. \\ &\quad \left. - \left( \frac{1}{n} \sum_{j=1}^n \frac{e^{isR_j/r_n} \mathbf{1}_{\{R_j > r_n\}}}{p_n} - \varphi_{\frac{R}{r_n}|r_n}(s) \right) \varphi_{\Theta|r_n}(t) \right. \\ &\quad \left. - \left( \frac{1}{n} \sum_{j=1}^n \frac{e^{it^T \Theta_j} \mathbf{1}_{\{R_j > r_n\}}}{p_n} - \varphi_{\Theta|r_n}(t) \frac{\hat{p}_n}{p_n} \right) \varphi_{\frac{R}{r_n}|r_n}(s) \right) \Bigg\} \\ &=: \tilde{E}_{11} + \tilde{E}_{12} + \tilde{E}_{13}. \end{aligned}$$

Without loss of generality, we show the tightness for  $\sqrt{np_n} \tilde{E}_{11}$  and that of  $\sqrt{np_n} \tilde{E}_{12}$  and  $\sqrt{np_n} \tilde{E}_{13}$  follows from the same argument.

First we introduce some notation following that from Bickel and Wichura (1971). Fix  $(s, t), (s, t) \in \mathbb{R}^{d+1}$  where  $s < s$  and  $t < t$ . Let  $B$  be the subset of  $\mathbb{R}^{d+1}$  of the form

$$B := ((s, t), (s, t)) \Bigg\} = (s, s] \times \prod_{k=1}^d [t_k, t_k] \subset \mathbb{R}^{d+1}.$$

For ease of notation, we suppress the dependence of  $B$  on  $(s, t), (s, t)$ . Define the increment of the stochastic process  $\tilde{E}_{11}$  on  $B$  to be

$$\begin{aligned} \tilde{E}_{11}(B) &:= \frac{1}{n} \sum_{j=1}^n \dots (-1)^{d+1 - \sum_j z_j} \\ &\quad \tilde{E}_{11}(s + z_0(s - s), t_1 + z_1(t_1 - t_1), \dots, t_d + z_d(t_d - t_d)) \Bigg\} \end{aligned}$$

From a sufficient condition of Theorem 3 of Bickel and Wichura (1971), the tightness of  $\sqrt{np_n} \tilde{E}_1$  is implied if the following statement holds for any  $(s, t), (s, t)$  and corresponding  $B$ ,

$$\mathbb{E} |\sqrt{np_n} \tilde{E}_{11}(B)|^2 \leq c |s - s|^\beta \prod_{k=1}^d |t_k - t_k|^\beta, \quad \text{for some } \beta > 1.$$

It follows that

$$\begin{aligned}
 & \mathbb{E} \left| \sqrt{np_n} \tilde{E}_{11}(B) \right|^2 \\
 &= np_n \mathbb{E} \left[ \left( \sum_{j=0,1} \dots (-1)^{d+1-\sum_j z_j} \frac{1}{n} \prod_{j=1}^n e^{i(s+z_0(s-s))R_j/r} \prod_{k=1}^d \left\{ e^{i(t_k+z_k(t_k-t_k))\Theta_{jk}} \frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} \right\} \right. \right. \\
 & \quad \left. \left. - \sum_{z_0=0,1} \dots (-1)^{d+1-\sum_j z_j} \mathbb{E} \left[ e^{i(s+z_0(s-s))R/r} \prod_{k=1}^d \left\{ e^{i(t_k+z_k(t_k-t_k))\Theta_k} \frac{\mathbf{1}_{\{R > r_n\}}}{p_n} \right\} \right] \right)^2 \right] \\
 &= np_n \mathbb{E} \left[ \left( \sum_{j=1}^n (e^{isR_j/r_j} - e^{isR/r_j}) \prod_{k=1}^d \left\{ e^{it_k\Theta_{jk}} - e^{it_k\Theta_k} \right\} \frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} \right. \right. \\
 & \quad \left. \left. - \mathbb{E} \left[ \left( e^{isR/r} - e^{isR/r} \right) \prod_{k=1}^d \left\{ e^{it_k\Theta_k} - e^{it_k\Theta_k} \right\} \frac{\mathbf{1}_{\{R > r_n\}}}{p_n} \right] \right)^2 \right] \\
 &= p_n \text{Var} \left( (e^{isR/r_j} - e^{isR/r_j}) \prod_{k=1}^d \left\{ e^{it_k\Theta_k} - e^{it_k\Theta_k} \right\} \frac{\mathbf{1}_{\{R > r_n\}}}{p_n} \right) \\
 &\leq \mathbb{E} \left[ \left( e^{isR/r} - e^{isR/r} \right) \prod_{k=1}^d \left\{ e^{it_k\Theta_k} - e^{it_k\Theta_k} \right\} \frac{\mathbf{1}_{\{R > r_n\}}}{p_n} \right]^2
 \end{aligned} \tag{32}$$

From a Taylor series argument,

$$|e^{ix} - e^{ix}|^2 \leq c1 \wedge |x - x|^2 \leq c1 \wedge |x - x|^\beta \leq c|x - x|^\beta, \quad \text{for any } \beta \in (0, 2].$$

Hence for any  $\beta \in (1, 2 \wedge \alpha)$ ,

$$\begin{aligned}
 \mathbb{E} \left| \sqrt{np_n} \tilde{E}_{11}(B) \right|^2 &\leq c|s - s|^\beta \prod_{k=1}^d |t_i - t_i|^\beta \mathbb{E} \left[ (R/r_n)^\beta \prod_{k=1}^d \left\{ |\Theta_k|^\beta |R > r_n| \right\} \right] \\
 &< c|s - s|^\beta \prod_{k=1}^d |t_i - t_i|^\beta,
 \end{aligned}$$

since  $|\Theta_k|^\beta$ 's are bounded and  $\sup_n \mathbb{E}[(R/r_n)^\beta |R > r_n] < \infty$  by the regular variation assumption. This proves the tightness.

Define the bounded set

$$K_\delta = \{(s, t) \mid \delta < |s| < 1/\delta, \delta < |t| < 1/\delta\}, \quad \text{for } \delta < .5.$$

Then, using (31), we have from the continuous mapping theorem,

$$np_n \int_{K_\delta} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{K_\delta} |Q(s, t)|^2 \mu(ds, dt). \tag{33}$$

On the other hand, for any  $\beta < 2 \wedge \alpha$ , we have

$$\begin{aligned}
 & \mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2 \\
 &= np_n \mathbb{E} \left| \frac{1}{n} \sum_{j=1}^n \left( \frac{U_{jn}V_{jn}}{p_n} - \mathbb{E} \left[ \frac{U_{jn}V_{jn}}{p_n} \right] \right) \right|^2 \\
 &\leq \frac{\mathbb{E}|U_{jn}V_{jn} - \mathbb{E}U_{jn}V_{jn}|^2}{p_n} \\
 &\leq \frac{c \mathbb{E}|U_{jn}V_{jn}|^2}{p_n} \tag{34} \\
 &= \frac{c \mathbb{E} \left[ \left| e^{isR_j/r_n} - \varphi_{\frac{R}{r_n}|r_n}(s) \right|^2 \left| e^{it^T\Theta_j} - \varphi_{\Theta|r_n}(t) \right|^2 \mathbf{1}_{\{R_j > r_n\}} \right]}{p_n} \\
 &\leq \frac{c \mathbb{E} \left[ \left( \left| e^{isR_j/r_n} - 1 \right|^2 + \left| \varphi_{\frac{R}{r_n}|r_n}(s) - 1 \right|^2 \right) \left( \left| e^{it^T\Theta_j} - 1 \right|^2 + \left| \varphi_{\Theta|r_n}(t) - 1 \right|^2 \right) \mathbf{1}_{\{R_j > r_n\}} \right]}{p_n} \\
 &\leq \frac{c \mathbb{E} \left[ \left( 1 \wedge \left| \frac{sR_j}{r_n} \right|^2 + \mathbb{E} \left[ 1 \wedge \left| \frac{sR_j}{r_n} \right|^2 \mid \left| \frac{R}{r_n} \right| > 1 \right] \right) \left( 1 \wedge |t\Theta_j|^2 + \mathbb{E} \left[ 1 \wedge |t\Theta_j|^2 \mid \left| \frac{R}{r_n} \right| > 1 \right] \right) \mathbf{1}_{\{R_j > r_n\}} \right]}{p_n} \\
 &\leq \frac{c \mathbb{E} \left[ \left( 1 \wedge \left| \frac{sR_j}{r_n} \right|^2 + \mathbb{E} \left[ 1 \wedge \left| \frac{sR_j}{r_n} \right|^2 \mid \left| \frac{R}{r_n} \right| > 1 \right] \right) \left( 1 \wedge |t\Theta_j|^2 + \mathbb{E} \left[ 1 \wedge |t\Theta_j|^2 \mid \left| \frac{R}{r_n} \right| > 1 \right] \right) \mathbf{1}_{\{R_j > r_n\}} \right]}{p_n} \\
 &\leq \frac{c \mathbb{E} \left[ (1 \wedge |s|^\beta) \left( \left| \frac{R_j}{r_n} \right|^2 + \mathbb{E} \left[ \left| \frac{R_j}{r_n} \right|^2 \mid \left| \frac{R}{r_n} \right| > 1 \right] \right) (1 \wedge |t|^2) \mathbf{1}_{\{R_j > r_n\}} \right]}{p_n} \\
 &\leq c \mathbb{E} \left( (1 \wedge |s|^\beta) (|R_j/r_n|^\beta + \mathbb{E}[|R/r_n|^\beta | R > r_n]) (1 \wedge |t|^2) \right) \mathbf{1}_{\{R > r_n\}} \\
 &\leq c(1 \wedge |s|^\beta)(1 \wedge |t|^2).
 \end{aligned}$$

Therefore for any  $\delta > 0$ ,

$$\begin{aligned}
 \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left[ np_n \int_{K_\delta^c} |\tilde{E}_1|^2 \mu(ds, dt) > \delta \right] &\leq \frac{1}{\delta} \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \int_{K_\delta^c} \mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2 \mu(ds, dt) \\
 &\leq \frac{1}{\delta} \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \int_{K_\delta^c} c(1 \wedge |s|^\beta)(1 \wedge |t|^2) \mu(ds, dt) \rightarrow 0
 \end{aligned}$$

by the dominated convergence theorem. This combined with Eq. 33 shows the convergence of  $np_n |\tilde{E}_1|^2 \mu(ds, dt)$  to  $|Q(s, t)|^2 \mu(ds, dt)$ , and hence completes the proof of the proposition.  $\square$

*Proof (Proof of Theorem 1(2) (cont.))*

Now it remains to show (30). Similar to the proof of Proposition 1, we can show that

$$\sqrt{np_n}\tilde{E}_{21} \xrightarrow{d} Q$$

for a centered Gaussian process  $Q$ , and

$$\tilde{E}_{22} \xrightarrow{p} 0.$$

Hence

$$\sqrt{np_n}\tilde{E}_2 = \sqrt{np_n}\tilde{E}_{21}\tilde{E}_{22} \xrightarrow{p} 0.$$

The argument then follows similarly from the continuous mapping theorem and bounding the tail integrals.  $\square$

*Proof (Proof of Theorem 1(1))*

Similar to the proof of Theorem 1(2), it suffices to show that

$$\iint |\tilde{E}_i|^2 \mu(ds, dt) \xrightarrow{P} 0, \quad i = 1, 2, 3. \quad (35)$$

The convergence (35) for  $i = 1, 2$  follows trivially from the more general results (28) and (30) in the proof of Theorem 1(2). Hence it suffices to show

$$\iint |\tilde{E}_3|^2 \mu(ds, dt) \rightarrow 0, \quad (36)$$

where we recall that  $\tilde{E}_3 := \varphi_{\frac{R}{r_n}, \Theta|_{r_n}}(s, t) - \varphi_{\frac{R}{r_n}|_{r_n}}(s) \varphi_{\Theta|_{r_n}}(t)$  is non-random.

Let  $P_{\frac{R}{r_n}, \Theta|_{r_n}}(\cdot) = P_{\left(\frac{R}{r_n}, \Theta\right) \left\{ \left| \frac{R}{r_n} > 1 \right. \right\}}$  and  $P_{\frac{R}{r_n}|_{r_n}}, P_{\Theta|_{r_n}}$  be the corresponding marginal measures. Then from Eq. 3,

$$P_{\frac{R}{r_n}, \Theta|_{r_n}} - P_{\frac{R}{r_n}|_{r_n}} P_{\Theta|_{r_n}} \xrightarrow{v} \nu_\alpha \times S - \nu_\alpha \times S = 0,$$

and hence for fixed  $(s, t)$ ,

$$\tilde{E}_3(s, t) = \iint e^{isT + itT} (P_{\frac{R}{r_n}, \Theta|_{r_n}} - P_{\frac{R}{r_n}|_{r_n}} P_{\Theta|_{r_n}})(dT, d\Theta) \rightarrow 0.$$

For any  $\beta < 2 \wedge \alpha$ , using the same argument in Eq. 34,

$$|\tilde{E}_3|^2 = \left( \frac{\mathbb{E}|U_{jn} V_{jn}|}{p_n} \right)^2 \leq c(1 \wedge |s|^\beta)(1 \wedge |t|^2).$$

Then Eq. 36 follows from Eq. 12 and dominated convergence. This concludes the proof.  $\square$

## Appendix B: Proof of Theorem 2

Following the same notation and steps as the proof of Theorem 1 in Appendix, it suffices to prove the following convergences for the mixing case:

$$\frac{\hat{p}_n}{p_n} \xrightarrow{P} 1, \quad (37)$$

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \iint_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt) \quad (38)$$

and

$$np_n \iint_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds, dt) \xrightarrow{P} 0. \quad (39)$$

We prove (37) and (38) in Propositions 2 and 3, respectively. The proof of Eq. 39 follows in a similar fashion. The proofs of both propositions rely on the following lemma.

Throughout this proof we make use of the results that if  $\{Z_t\}$  is stationary and  $\alpha$ -mixing with coefficient  $\{\alpha_h\}$ , then

$$|\text{cov}(Z_0, Z_h)| \leq c\alpha_h^\delta \mathbb{E}|Z_0|^{2/(1-\delta)}^{1-\delta}, \quad \text{for any } \delta \in (0, 1), \quad (40)$$

see Section 1.2.2, Theorem 3(a) of Doukhan (1994).

**Lemma 1** *Let  $\{X_t\}$  be a multivariate stationary time series that is regularly varying and  $\alpha$ -mixing with mixing coefficient  $\{\alpha_h\}$ . For a sequence  $r_n \rightarrow \infty$ , set  $p_n = \mathbb{P}(\|X_0\| > r_n)$ . Let  $f_1, f_2$  be bounded functions which vanish outside  $\mathbb{R}^d \setminus B_1(\mathbf{0})$ , where  $B_1(\mathbf{0})$  is the unit open ball  $\{\mathbf{x} | \|\mathbf{x}\| < 1\}$ , with sets of discontinuity of measure zero. Set,*

$$S_n^{(i)} = \frac{1}{n} \sum_{t=1}^n \left( f_i \left( \frac{X_t}{r_n} \right) - \mathbb{E} f_i \left( \frac{X_0}{r_n} \right) \right) \quad i = 1, 2.$$

Assume that condition (M) holds for  $\{\alpha_h\}$  and  $\{r_n\}$ . Then

$$\frac{1}{\sqrt{np_n}} (S_n^{(1)}, S_n^{(2)})^T \xrightarrow{d} N(\mathbf{0}, \Sigma), \quad (41)$$

where the covariance matrix  $[\Sigma_{ij}]_{i,j=1,2} = [\sigma^2(f_i, f_j)]_{i,j=1,2}$  with

$$\sigma^2(f_1, f_2) := \sigma_0^2(f_1, f_2) + 2 \sum_{h=1}^{\infty} \sigma_h^2(f_1, f_2) \quad (42)$$

and

$$\sigma_h^2(f_1, f_2) = \int f_1 f_2 d\mu_h, \quad h \geq 0. \quad (43)$$

In particular,

$$\frac{1}{np_n} (S_n^{(1)}, S_n^{(2)})^T \xrightarrow{p} \mathbf{0}.$$

The proof of Lemma 1 is provided after the proofs of the propositions.

**Proposition 2** *Assume that condition (M) holds, then*

$$\frac{\hat{p}_n}{p_n} \xrightarrow{p} 1,$$

*Proof* We have

$$\frac{\hat{p}_n}{p_n} - 1 = \frac{1}{n} \sum_{j=1}^n \left( \frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} - 1 \right) = \frac{1}{np_n} \sum_{j=1}^n (\mathbf{1}_{\{R_j > r_n\}} - p_n).$$

Apply Lemma 1 to  $f(\mathbf{x}) = \mathbf{1}_{\{\|\mathbf{x}\| > 1\}}$  and the result follows.  $\square$

**Proposition 3** Assume that condition (M) holds, and that  $\mu$  and  $\{r_n\}$  satisfies (12) and (13), respectively, then

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s, t)|^2 \mu(ds, dt),$$

where  $Q$  is a centered Gaussian process.

*Proof* Let us first establish the convergence of  $\sqrt{np_n} \tilde{E}_1(s, t)$  for fixed  $(s, t)$ . Take

$$\begin{aligned} f_1(\mathbf{x}) &= \operatorname{Re} \left( e^{is\|\mathbf{x}\|} - \mathbb{E}[e^{is\|\mathbf{x}\|} | \|\mathbf{x}\| > 1] \right) \left( \left( e^{it\mathbf{x}/\|\mathbf{x}\|} - \mathbb{E}[e^{it\mathbf{x}/\|\mathbf{x}\|} | \|\mathbf{x}\| > 1] \right) \mathbf{1}_{\|\mathbf{x}\| > 1} \right) \\ f_2(\mathbf{x}) &= \operatorname{Im} \left( e^{is\|\mathbf{x}\|} - \mathbb{E}[e^{is\|\mathbf{x}\|} | \|\mathbf{x}\| > 1] \right) \left( \left( e^{it\mathbf{x}/\|\mathbf{x}\|} - \mathbb{E}[e^{it\mathbf{x}/\|\mathbf{x}\|} | \|\mathbf{x}\| > 1] \right) \mathbf{1}_{\|\mathbf{x}\| > 1} \right) \end{aligned}$$

Then from Lemma 1,

$$\frac{1}{\sqrt{np_n}} (S_n^{(1)}, S_n^{(2)})^T = \sqrt{np_n} (\operatorname{Re}\{\tilde{E}_1(s, t)\}, \operatorname{Im}\{\tilde{E}_1(s, t)\}) \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

where the covariance structure  $\Sigma$  can be derived from Eqs. 42 and 43. This implies that

$$\sqrt{np_n} \tilde{E}_1(s, t) \xrightarrow{d} Q(s, t),$$

where  $Q(s, t)$  is a zero-mean complex normal process with covariance matrix  $\Sigma_{11} + \Sigma_{22}$  and relation matrix  $\Sigma_{11} - \Sigma_{22} + i(\Sigma_{12} + \Sigma_{21})$ .

The finite-dimensional distributional convergence of  $\sqrt{np_n} \tilde{E}_1$  to a  $Q(s, t)$  can be generalized using the Cramér-Wold device and we omit the calculation of the covariance structure. The tightness condition for the functional convergence follows the same arguments in the proof of Proposition 1 from Bickel and Wichura (1971), with equality (32) replaced by a variance calculation of the sum of  $\alpha$ -mixing components using the inequality (40) and condition (33) is verified through the same argument. This completes the proof of Proposition 3.  $\square$

*Proof (of Lemma 1)* The proof follows from that of Theorem 3.2 in Davis and Mikosch (2009). Here we outline the sketch of the proof and detail only the parts that differ from their proof.

By the vague convergence in Eq. 20, we have

i)

$$\frac{1}{p_n} \mathbb{E} \left[ f_i \left( \frac{\mathbf{X}_0}{r_n} \right) \right] \left\{ \rightarrow \int f_i d\mu_0 \quad \text{and} \quad \frac{1}{p_n} \mathbb{E} \left[ f_i^2 \left( \frac{\mathbf{X}_0}{r_n} \right) \right] \left\{ \rightarrow \int f_i^2 d\mu_0; \right. \right.$$

ii)

$$\frac{1}{p_n} \operatorname{Var} \left[ f_i \left( \frac{\mathbf{X}_0}{r_n} \right) \right] = \frac{1}{p_n} \mathbb{E} \left[ f_i^2 \left( \frac{\mathbf{X}_0}{r_n} \right) \right] - p_n \left( \frac{1}{p_n} \mathbb{E} \left[ f_i \left( \frac{\mathbf{X}_0}{r_n} \right) \right] \right)^2 \rightarrow \int f_i^2 d\mu_0 = \sigma_0^2(f_i, f_i);$$

iii)

$$\frac{1}{p_n} \operatorname{cov} \left[ f_i \left( \frac{\mathbf{X}_0}{r_n} \right), f_j \left( \frac{\mathbf{X}_h}{r_n} \right) \right] \left\{ \rightarrow \int f_i f_j d\mu_h = \sigma_h^2(f_i, f_j).$$

Let us first consider the marginal convergence of  $\frac{1}{\sqrt{n}p_n}S_n^{(i)}$  for  $i = 1, 2$ . Without loss of generality, we suppress the dependency on  $i$  and set

$$Y_{1n} := f\left(\frac{\mathbf{X}_t}{r_n}\right) \Big\{ - \mathbb{E} f\left(\frac{\mathbf{X}_0}{r_n}\right).$$

Then

$$\frac{1}{p_n} \text{Var}[Y_{1n}] \rightarrow \sigma_0^2(f, f) \quad \text{and} \quad \frac{1}{p_n} \text{cov}(Y_{1n}, Y_{(h+1)n}) \Big\{ \rightarrow \sigma_h^2(f, f).$$

We also have the following two results for  $|\text{cov}(Y_{1n}, Y_{(h+1)n})|$ :

$$\begin{aligned} & \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{p_n} |\text{cov}(Y_{1n}, Y_{(j+1)n})| \\ & \leq \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{p_n} \mathbb{E} \left| f\left(\frac{\mathbf{X}_0}{r_n}\right) f\left(\frac{\mathbf{X}_j}{r_n}\right) \right| \Big\{ + \frac{1}{p_n} \mathbb{E} \left( \left| f\left(\frac{\mathbf{X}_0}{r_n}\right) \right| \right)^2 \\ & \leq \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{c}{p_n} \mathbb{E} (\mathbf{1}_{\{\|\mathbf{X}_0\| > r_n\}}) (\mathbf{1}_{\{\|\mathbf{X}_j\| > r_n\}}) \Big\{ + \frac{c}{p_n} (\mathbb{E} \mathbf{1}_{\{\|\mathbf{X}_0\| > r_n\}})^2 \\ & \leq \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{c}{p_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) \Big\{ + cl_n p_n \\ & = 0 \end{aligned} \tag{44}$$

from condition (22), and

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{p_n} |\text{cov}(Y_{1n}, Y_{(j+1)n})| & \leq \lim_{n \rightarrow \infty} \frac{1}{p_n} \left| \text{cov} \left[ f\left(\frac{\mathbf{X}_0}{r_n}\right) \Big\{ f\left(\frac{\mathbf{X}_j}{r_n}\right) \right] \right| \Big\{ \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{p_n} \alpha_j^\delta \left( \mathbb{E} \left| f\left(\frac{\mathbf{X}_0}{r_n}\right) \right|^{2/(1-\delta)} \right)^{1-\delta} \\ & \leq \lim_{n \rightarrow \infty} \frac{c}{p_n} \alpha_j^\delta \mathbb{E} (\mathbf{1}_{\{\|\mathbf{X}_0\| > r_n\}})^{2/(1-\delta)}^{1-\delta} \\ & \leq \lim_{n \rightarrow \infty} c \alpha_j^\delta p_n^{-\delta} \\ & = 0 \end{aligned} \tag{45}$$

from condition (21).

We apply the same technique of small/large blocks as used in Davis and Mikosch (2009). Let  $m_n$  and  $l_n$  be the sizes of big and small blocks, respectively, where  $l_n \ll m_n \ll n$ . Let  $I_{kn} = \{(k-1)m_n + 1, \dots, km_n\}$  and  $J_{kn} = \{(k-1)m_n + 1, \dots, (k-1)m_n + l_n\}$ ,  $k = 1, \dots, n/m_n$ , be the index sets of big and small blocks respectively.

Set  $\tilde{I}_{kn} = I_{kn} \setminus J_{kn}$ , i.e.,  $\tilde{I}_{kn}$  are the big blocks with the first  $l_n$  observations removed. For simplicity, we set  $m_n := 1/p_n$  and assume that the number of big blocks  $n/m_n = np_n$  is integer-valued. The non-integer case can be generalized without additional difficulties.

Denote

$$S_n(B) := \sum_{t \in B} Y_{tn},$$

then

$$Y_{tn} = S_n(1:n) = \sum_{k=1}^{np_n} S_n(I_{kn}) = \sum_{k=1}^{np_n} S_n(\tilde{I}_{kn}) + \sum_{k=1}^{np_n} S_n(J_{kn}).$$

Let  $\{\tilde{S}_n(\tilde{I}_{kn})\}_{k=1, \dots, np_n}$  be iid copies of  $\tilde{S}_n(\tilde{I}_{1n})$ . To prove the convergence of  $\frac{1}{\sqrt{np_n}} S_n(1:n)$ , it suffices to show the following:

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} \tilde{S}_n(\tilde{I}_{kn}) \text{ and } \frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} S_n(\tilde{I}_{kn}) \text{ has the same limiting distribution,} \quad (46)$$

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} S_n(J_{kn}) \xrightarrow{p} 0, \quad (47)$$

and

$$\frac{1}{\sqrt{np_n}} \sum_{k=1}^{np_n} \tilde{S}_n(\tilde{I}_{kn}) \xrightarrow{d} N(0, \sigma^2(f, f)). \quad (48)$$

The statement (46) holds if

$$np_n \alpha_{l_n} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (49)$$

This follows from the same argument in equation (6.2) in Davis and Mikosch (2009).

For condition (47), it suffices to show that

$$\frac{1}{np_n} \text{Var} \left( \sum_{k=1}^{np_n} S_n(J_{kn}) \right) \rightarrow 0.$$

Note that

$$\frac{1}{np_n} \text{Var} \left( \sum_{k=1}^{np_n} S_n(J_{kn}) \right) \leq \text{Var}(S_n(J_{1n})) + 2 \sum_{h=1}^{np_n-1} \left(1 - \frac{h}{np_n}\right) |\text{cov}(S_n(J_{1n}), S_n(J_{(h+1)n}))| =: P_1 + P_2.$$



We have

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} P_1 &= \limsup_{n \rightarrow \infty} \text{Var} \left( \begin{matrix} l_n \\ Y_{jn} \end{matrix} \right) \Bigg\{ \\
 &\leq \limsup_{n \rightarrow \infty} l_n p_n \left( \frac{\text{Var}(Y_{1n})}{p_n} + 2 \sum_{j=1}^{l_n-1} \left(1 - \frac{j}{l_n}\right) \frac{|\text{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \right) \Bigg\} \\
 &\leq \limsup_{n \rightarrow \infty} l_n p_n \frac{\text{Var}(Y_{1n})}{p_n} + \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} 2 l_n p_n \sum_{j=1}^{h-1} \frac{|\text{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \\
 &\quad + \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} 2 l_n p_n \sum_{j=h}^{l_n-1} \frac{|\text{cov}(Y_{1n}, Y_{(j+1)n})|}{p_n} \\
 &= 0
 \end{aligned}$$

where the last step follows from dominated convergence and Eq. 44. And for the other term,

$$\begin{aligned}
 P_2 &\leq 2 \sum_{h=1}^{np_n-1} \sum_{s \in J_{1n}} \sum_{t \in J_{(h+1)n}} |\text{cov}(Y_{sn}, Y_{tn})| \\
 &\leq 2 \sum_{h=1}^{np_n-1} \sum_{k=h/p_n-l_n+1}^{h/p_n} l_n |\text{cov}(Y_{1n}, Y_{(k+1)n})| \Bigg\{ \\
 &\leq 2 l_n p_n \sum_{k=1/p_n-l_n+1}^{\infty} \frac{|\text{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} \\
 &\leq 2 l_n p_n \sum_{k=l_n+1}^{\infty} \frac{|\text{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} \rightarrow 0.
 \end{aligned}$$

Note that  $1/p_n = m_n$  is the size of big blocks  $I_{kn}$ 's and  $1/p_n - l_n + 1 = m_n - l_n + 1$  is the distance between consecutive small blocks  $(J_{kn}, J_{(k+1)n})$ 's. The last limit follows from Eq. 45.

To finish the proof, we need to establish the central limit theorem in Eq. 48. Note the  $\tilde{S}_n(\tilde{I}_{1n})$ 's are iid with  $\mathbb{E}\tilde{S}_n(\tilde{I}_{1n}) = 0$ . We now calculate its variance. Recall that  $1/p_n - l_n$  is the size of  $\tilde{I}_{1n}$ , the big block with small block removed. Then

$$\begin{aligned}
 \text{Var} \tilde{S}_n(\tilde{I}_{1n}) &= \text{Var} \left( \sum_{j=1}^{1/p_n-l_n} Y_{jn} \right) \Bigg\{ \\
 &= \left( \frac{1}{p_n} - l_n \right) \text{Var}(Y_{jn}) + 2 \sum_{k=1}^{1/p_n-l_n-1} (1/p_n - l_n - k) \text{cov}(Y_{1n}, Y_{(k+1)n}) \\
 &= \left( \frac{1}{p_n} - l_n \right) \text{Var}(Y_{jn}) + 2 \left( \sum_{k=1}^h + \sum_{k=h+1}^{l_n} + \sum_{k=l_n+1}^{1/p_n-l_n-1} \right) \left( 1 - \frac{l_n+k}{1/p_n} \right) \frac{1}{p_n} \text{cov}(Y_{1n}, Y_{(k+1)n}) \\
 &:= I_0 + I_1 + I_2 + I_3.
 \end{aligned}$$

Here

$$\lim_{n \rightarrow \infty} I_0 = \lim_{n \rightarrow \infty} (1 - l_n p_n) \frac{1}{p_n} \text{Var}(Y_{jn}) = \sigma_0^2(f, f),$$

and

$$\lim_{n \rightarrow \infty} I_1 = \lim_{n \rightarrow \infty} 2 \sum_{k=1}^h (1 - p_n(l_n + k)) \frac{\text{cov}(Y_{1n}, Y_{(k+1)n})}{p_n} = 2 \sum_{k=1}^h \sigma_k^2(f, f).$$

We also have

$$\lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} |I_2| \leq \lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{k=h+1}^{l_n} \frac{|\text{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} = 0$$

from Eq. 44, and

$$\lim_{n \rightarrow \infty} |I_3| \leq \lim_{n \rightarrow \infty} \sum_{k=l_n}^{\infty} \frac{|\text{cov}(Y_{1n}, Y_{(k+1)n})|}{p_n} = 0$$

from Eq. 45. Therefore

$$\lim_{n \rightarrow \infty} \text{Var} \tilde{S}_n(\tilde{I}_{1n}) \Big| = \lim_{n \rightarrow \infty} I_0 + \lim_{h \rightarrow \infty} \lim_{n \rightarrow \infty} I_1 = \sigma_0^2(f, f) + 2 \sum_{k=1}^{\infty} \sigma_k^2(f, f) =: \sigma^2(f, f)$$

as defined. To show that this infinite sum converges, it suffices to show that

$$\sum_{h=1}^{\infty} \mu_h(\{(\mathbf{x}, \mathbf{x}) | \|\mathbf{x}\| > 1, \|\mathbf{x}\| > 1\}) < \infty.$$

This follows from Eq. 22 in condition **(M)**, for if

$$\sum_{h=1}^{\infty} \mu_h(\{(\mathbf{x}, \mathbf{x}) | \|\mathbf{x}\| > 1, \|\mathbf{x}\| > 1\}) = \infty,$$

then

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n) &\geq \liminf_{n \rightarrow \infty} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > r_n, \|\mathbf{X}_j\| > r_n | \|\mathbf{X}_0\| > r_n) \\ &\geq \sum_{j=h}^{\infty} \mu_j(\{(\mathbf{x}, \mathbf{x}) | \|\mathbf{x}\| > 1, \|\mathbf{x}\| > 1\}) = \infty, \end{aligned}$$

which leads to a contradiction.

To apply the central limit theorem, we verify the Lindeberg's condition,

$$\begin{aligned} \mathbb{E} (\tilde{S}_n(\tilde{I}_{1n}))^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon \sqrt{np_n}\}} &\leq \mathbb{E} \left[ \left( \sum_{j=1}^{1/p_n - l_n} Y_{jn} \right)^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon \sqrt{np_n}\}} \right] \\ &\leq \mathbb{E} c(1/p_n - l_n)^2 \mathbf{1}_{\{|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon \sqrt{np_n}\}} \\ &\leq c \frac{1}{p_n^2} \mathbb{P}(|\tilde{S}_n(\tilde{I}_{1n})| > \epsilon \sqrt{np_n}) \\ &\leq c \frac{1}{p_n^2} \frac{\text{Var } \tilde{S}_n(\tilde{I}_{1n})}{2np_n} = O\left(\frac{1}{np_n^3}\right) \rightarrow 0. \end{aligned}$$

This completes the proof for the convergence of  $\frac{1}{\sqrt{np_n}} S_n(1:n)$ .

The joint convergence of  $\frac{1}{\sqrt{np_n}} (S_n^{(1)}, S_n^{(2)})^T$  follows from the same line of argument together with the Crámer-Wold device. In particular,

$$\frac{1}{np_n} \text{cov } S_n^{(i)}, S_n^{(j)} \Big| = \sigma^2(f_i, f_j), \quad i, j = 1, 2.$$

This completes the proof of the lemma.  $\square$

**Remark 3** Lemma 1 itself is a more general result of independent interest. The result can be generalized for functions  $f_i$  defined on  $\overline{\mathbb{R}^d} \setminus \{\mathbf{0}\}$  with compact support. In this case, condition (22) should be modified to

$$\lim_{h \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{p_n} \sum_{j=h}^{l_n} \mathbb{P}(\|\mathbf{X}_0\| > \epsilon r_n, \|\mathbf{X}_j\| > \epsilon r_n) = 0$$

for some  $\epsilon > 0$ , where  $\text{support}(f) \subseteq \overline{\mathbb{R}^d} \setminus B(\mathbf{0})$ . Also, as seen during the proof of the lemma, the conditions on  $p_n$ ,  $l_n$ , and  $\alpha_t$  can be further relaxed.

## References

- Bickel, P.J., Wichura, M.J.: Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.* **42**, 1656–1670 (1971)
- Billingsley, P. *Probability and Measure*, 3rd edition. Wiley, New York (1995)
- Davis, R.A., Mikosch, T.: The extremogram: a correlogram for extreme events. *Bernoulli* **15**(4), 977–1009 (2009)
- Davis, R.A., Mikosch, T., Cribben, I.: Towards estimating extremal serial dependence via the bootstrapped extremogram. *J. Econometrics* **170**(1), 142–152 (2012)
- Davis, R.A., Matsui, M., Mikosch, T., Wan, P.: Applications of distance covariance to time series. [arXiv:1606.05481](https://arxiv.org/abs/1606.05481) (2017)
- de Haan, L., de Ronde, J.: Sea and wind: multivariate extremes at work. *Extremes* **1**, 7–46 (1998)
- Doukhan, P.: *Mixing: Properties and Examples*. Springer, New York (1994)
- Feuerverger, A.: A consistent test for bivariate dependence. *Internat. Statis. Rev.* **61**(3), 419–433 (1993)
- Fryzlewicz, P.: Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42**(6), 2243–2281 (2014)

- Jeon, S., Smith, R.L.: Dependence structure of spatial extremes using threshold approach. arXiv:[1209.6344v1](https://arxiv.org/abs/1209.6344v1) (2014)
- Joe, H., Smith, R.L., Weissman, I.: Bivariate threshold methods for extremes. *JRSS B.* **54**(1), 171–183 (1992)
- Mallik, A., Sen, B., Banerjee, M., Michailidis, G.: Threshold estimation based on a  $p$ -value framework in dose-response and regression settings. *Biometrika* **98**(4), 887–900 (2011)
- Page, E.S.: Continuous inspection schemes. *Biometrika* **41**(1), 100–115 (1954)
- Resnick, S.I.: Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5**, 303–336 (2002)
- Resnick, S.I.: *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York (2007)
- Stáříčá, C.: Multivariate extremes for models with constant conditional correlations. *J. Empir. Finan.* **6**, 515–553 (1999)
- Székely, G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–2794 (2007)