Threshold Selection for Multivariate Heavy-Tailed Data

Phyllis Wan and Richard A. Davis

Department of Statistics
Columbia University
1255 Amsterdam Avenue, MC 4690
New York, NY 10027
e-mail: phyllis@stat.columbia.edu; rdavis@stat.columbia.edu

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 Θ as R goes to infinity. The conditional limiting distribution of Θ 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 Θ 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 and phrases: distance covariance, heavy-tailed data, multivariate regular variation, threshold selection.

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 $(R, \mathbf{\Theta}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|)$, where $\|\cdot\|$ is some norm, satisfies 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) $\mathbb{P}(\mathbf{\Theta} \in \cdot | R > r)$ converges weakly to a measure $S(\cdot)$ as $r \to \infty$.

The α 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 $\{(R_i, \mathbf{\Theta}_i)\}_{i=1}^n$, 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, $\mathbf{\Theta}_i$ 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 Θ 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 Stărică plots (Stărică, 1999) looked at the joint tail empirical measure, but was, in some way, equivalent to only examining the extremal behavior of R. It also required handpicking the threshold from several graphs. 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 $\mathbf{\Theta}$ given $R > r$ become independent as $r \to \infty$. Given a sequence of candidate

threshold levels, we test the degree of dependence between R and $\mathbf{\Theta}$ 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 Secion 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 $\mathbf{X} = (X_1, \ldots, X_d)$ be a d-dimensional random variable defined on the cone $\bar{\mathbb{R}}^d_+ = [\mathbf{0}, \infty] \backslash \{\mathbf{0}\}$. Define the polar coordinate transformation

$$T(\mathbf{X}) = (\|\mathbf{X}\|, \mathbf{X}/\|\mathbf{X}\|) =: (R, \mathbf{\Theta}). \tag{2.1}$$

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 $b(t) \to \infty$, such that

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

where $\xrightarrow{v}$ denotes vague convergence, and ν_α 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.,

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

This implies that

$$\mathbb{P}\left[\left(\frac{R}{r}, \mathbf{\Theta}\right) \in \cdot \middle| R > r\right] \xrightarrow{w} \nu_\alpha \times S, \quad r \to \infty, \quad \text{on } [1, \infty) \times \mathbb{S}^{d-1}, \tag{2.2}$$

where $\xrightarrow{w}$ denotes weak convergence. In other words, given that $R > r$ for r large, the conditional distribution of R/r and $\mathbf{\Theta}$ are asymptotically independent. The angular measure S characterizes the tail dependence structure of $\mathbf{X}$. If S is concentrated on $\{e_i, i = 1, \ldots d\}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, 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 $\{(e_1, \ldots, e_d) \in \mathbb{S}^{d-1} | e_i > 0, e_j > 0, i \neq j\}$, then an extreme observation in the X_i direction implicates that 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 (2.2):

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

This suggests estimating S using the angular data $(\mathbf{\Theta}_i)$ 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 $\mathbf{\Theta}$ are virtually independent. Given a candidate threshold sequence $\{r_k\}$, we formally test the independence between $(R_i, \mathbf{\Theta}_i)$ 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 $\mathbf{\Theta}$ 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

$$T(X,Y;\mu) = \int_{\mathbb{R}^{p+q}} \left|\varphi_{X,Y}(s,t) - \varphi_X(s)\,\varphi_Y(t)\right|^2 \mu(ds,dt)\,, \quad (s,t) \in \mathbb{R}^{p+q} \tag{3.1}$$

where $\varphi_{X,Y}(s,t), \varphi_X(s), \varphi_Y(t)$ denote the joint and marginal characteristic functions of (X,Y) and μ is a measure on $\mathbb{R}^{p+q}$. One advantage of distance covariance over, say, Pearson's covariance, is that, if μ 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 $(X_1,Y_1),\ldots,(X_n,Y_n)$, define the empirical distance covariance

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

where $\hat{\varphi}_{X,Y}(s,t) = \frac{1}{n}\sum_{j=1}^n \mathrm{e}^{\,i\,\langle s,X_j\rangle + i\,\langle t,Y_j\rangle}$ 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 $T_n(X,Y;\mu)$ 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 $\tilde{\mu}(x) = \int (1 - \cos\langle s,x\rangle)\,\mu(ds)$ (Davis et al., 2017).

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

$$\mu(ds,dt) = c_{p,q}|s|^{-\kappa-p}|t|^{-\kappa-q}ds\,dt\,. \tag{3.2}$$

where $c_{p,q}$ is as defined in Lemma 1 of Székely et al. (2007). This choice of μ gives $\tilde{\mu}(x)\tilde{\mu}(y) = |x|^\kappa|y|^\kappa$. Moreover, this is the only choice of μ for which the distance covariance is invariant relative to scale and orthogonal transformations. Note that in order for the integral (3.1) to exist, it is required that

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

We will utilize the described weight measure (3.2) with $\kappa = 1$ in our simulations and data analyses 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

$$T_n(X,Y;\mu) \xrightarrow{\text{a.s.}} T(X,Y;\mu). \tag{3.3}$$

Further, if X and Y are independent, then under an α-mixing condition,

$$n\,T_n(X,Y;\mu) \xrightarrow{d} \int_{\mathbb{R}^{p+q}} |G(s,t)|^2\mu(s,t) \tag{3.4}$$

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'_μ, implying that $n\,T_n(X,Y;\mu)$ diverges as $n \to \infty$. 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 χ if $n\,T_n(X,Y;\mu) > c$, where c is the upper χ-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 (2.1) and (2.2), and $\{(R_i, \mathbf{\Theta}_i)\}_{i=1}^n$ be their polar coordinate transformations. Given a threshold r_n, we measure the dependence between R and $\mathbf{\Theta}$ conditional on $R > r_n$ by the empirical distance covariance of $(R, \mathbf{\Theta})$ on the set $\{R > r_n\}$:

$$T_n := \int_{\mathbb{R}^{d+1}} |C_n(s,t)|^2 \mu(ds, dt), \tag{4.1}$$

with

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

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

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

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

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

In this section, we establish the limiting results (3.3) and (3.4) adapted to the conditional distance covariance. For ease of notation, let

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

be the theoretical and empirical probability of exceedance, and let

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

and

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

be the theoretical conditional joint and marginal characteristic functions.

Theorem 4.1. *1. Assume that $\mathbf{X}$ is multivariate regularly varying with index $\alpha > 1$. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be iid observations generated from $\mathbf{X}$ and T_n be the conditional distance covariance between the angular and radial component defined in (4.1). Further assume that $np_n \to \infty$ and the weight measure μ satisfies*

$$\int_{\mathbb{R}^{d+1}} (1 \wedge |s|^\beta)(1 \wedge |t|^2) \mu(ds, dt) < \infty \tag{4.2}$$

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

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

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

$$np_n \int_{\mathbb{R}^{d+1}} |\varphi_{R,\Theta|r_n}(s,t) - \varphi_{R|r_n}(s)\varphi_{\Theta|r_n}(t)|^2 \mu(ds, dt) \to 0, \tag{4.3}$$

then

$$n\hat{p}_n T_n \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds, dt), \tag{4.4}$$

where Q is a centered Gaussian process with covariance function given in (A.9).

Remark 4.2. *In the case where* $\mathbf{X}$ *is regularly varying with index* $\alpha \leq 1$*, similar results hold if we replace* R/r_n *with* $\log(R/r_n)$*.*

The proof of the theorem is delayed to Appendix A. We now provide a sufficient condition for assumption (4.3).

Remark 4.3. *If we assume* $\mu = \mu_1 \times \mu_2$*, where* μ_1, μ_2 *are measures on* $\mathbb{R}$ *and* $\mathbb{R}^d$*, respectively, and symmetric about the origin, then from Section 2.2 of Davis et al. (2017), condition* (4.3) *is equivalent to*

$$\begin{aligned} np_n \, (\mathbb{E}[\tilde{\mu}_1(R-R')\,\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}')] + \mathbb{E}[\tilde{\mu}_1(R-R')]\,\mathbb{E}[\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}')] \\ -2\,\mathbb{E}[\tilde{\mu}_1(R-R')\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}'')]) \quad \to \quad 0 \end{aligned}$$

where

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

or equivalently,

$$\begin{aligned} & np_n \int_{\mathbb{R}^{d+1}} \tilde{\mu}_1(R-R')\,\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}')\,\big(P_{R,\Theta|r_n}(dR,d\boldsymbol{\Theta})P_{R,\Theta|r_n}(dR',d\boldsymbol{\Theta}') \\ & \qquad + P_{R|r_n}(dR)P_{\Theta|r_n}(d\boldsymbol{\Theta})P_{R|r_n}(dR')P_{\Theta|r_n}(d\boldsymbol{\Theta}') \\ & \qquad - 2P_{R,\Theta|r_n}(dR,d\boldsymbol{\Theta})P_{R|r_n}(dR')P_{\Theta|r_n}(d\boldsymbol{\Theta}')\big) \\ = \quad & \int_{\mathbb{R}^{d+1}} \tilde{\mu}_1(R-R')\,\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}')\,\sqrt{np_n}\,\big(P_{R,\Theta|r_n}(dR,d\boldsymbol{\Theta}) - P_{R|r_n}(dR)P_{\Theta|r_n}(d\boldsymbol{\Theta})\big) \\ & \quad \sqrt{np_n}\,\big(P_{R,\Theta|r_n}(dR',d\boldsymbol{\Theta}') - P_{R|r_n}(dR')P_{\Theta|r_n}(d\boldsymbol{\Theta}')\big) \qquad (4.5) \\ \to \quad & 0. \end{aligned}$$

One way to approach this requirement is to assume a second-order like condition on the distribution of $(R, \boldsymbol{\Theta})$*. For example, we may assume that*

$$\frac{P_{R,\Theta|r_n} - \nu_\alpha \times S}{A(r_n)} \xrightarrow{w} \chi, \qquad (4.6)$$

where χ *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* $A(t) \to 0$ *as* $t \to \infty$*. If we choose the sequence* r_n *such that* $\sqrt{np_n} \to \infty$ *and* $\sqrt{np_n}A(r_n) \to 0$*, then*

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

on $[1,\infty]\times\mathbb{S}^{d-1}$*. And* (4.5) *is satisfied since*

$$\begin{aligned} \int_{\mathbb{R}^{d+1}} \Bigg[\int_{\mathbb{R}^{d+1}} & \tilde{\mu}_1(R-R')\,\tilde{\mu}_2(\boldsymbol{\Theta}-\boldsymbol{\Theta}')\; np_nA^2(r_n) \\ & \frac{P_{R,\Theta|r_n}(dR,d\boldsymbol{\Theta}) - P_{R|r_n}(dR)P_{\Theta|r_n}(d\boldsymbol{\Theta})}{A(r_n)}\Bigg] \\ & \frac{P_{R,\Theta|r_n}(dR',d\boldsymbol{\Theta}') - P_{R|r_n}(dR')P_{\Theta|r_n}(d\boldsymbol{\Theta}')}{A(r_n)} \quad \to \quad 0. \end{aligned}$$

In the special case that $|A| \in RV_\rho$ for $\rho < 0$, r_n can be chosen such that

$$O(n^{\frac{1}{\alpha+2\rho}+\epsilon}) \leq r_n \leq o(n^{\frac{1}{\alpha}})$$

for some $\epsilon > 0$.

If the components of $\mathbf{X}$ are asymptotically independent, this is equivalent to the second order condition for multivariate regular variation (Resnick, 2002). In the following, we give an example of (4.6) where $\mathbf{X}$ is asymptotically dependent.

Example 4.4. Let $\mathbf{X}$ follow a bivariate logistic distribution, then $\mathbf{X}$ has cdf

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

and is asymptotically independent if and only if $\gamma = 1$. It can be shown that $\mathbf{X}$ is regularly varying with index $\alpha = 1$. Taking the pseudo-polar coordinate transform $(R, \Theta) = (X_1 + X_2, X_1/(X_1 + X_2))$, the pdf of (R, Θ) is

$$\begin{aligned}
& f_{R,\Theta}(r,\theta) \\
= \; & r^{-2}\left(\theta(1-\theta)\right)^{-\frac{\gamma+1}{\gamma}}\left(\theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma-2} e^{-r^{-1}\left(\theta^{-\frac{1}{\gamma}}+(1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma}} \\
& \left(r^{-1}\left(\theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma} - \frac{\gamma-1}{\gamma}\right) \\
\sim \; & 2r^{-2}\left(-\frac{1}{2}\frac{\gamma-1}{\gamma}(\theta(1-\theta))^{-\frac{\gamma+1}{\gamma}}\left(\theta^{-\frac{1}{\gamma}} + (1-\theta)^{-\frac{1}{\gamma}}\right)^{\gamma-2}\right) \\
=: \; & f_R(r) f_\Theta(\theta). \qquad (4.7)
\end{aligned}$$

Hence

$$f_{R,\Theta}(r,\theta) - f_R(r) f_\Theta(\theta) \sim O(r^{-3}),$$

and for any B Borel in $\mathbb{S}^{d-1}$ and $r \geq r_n$,

$$\begin{aligned}
P_{R,\Theta|r_n}((r,\infty] \times B) - \nu_1(r/r_n, \infty]S(B) &= \frac{\int_{rr_n}^\infty \int_B (f_{R,\Theta}(r,\theta) - f_R(r) f_\Theta(\theta)) dr d\theta}{r_n^{-1}} \\
&= O(r_n^{-1}).
\end{aligned}$$

Therefore (4.6) holds for $A(r_n) = r_n^{-1}$ and the conditions $np_n \to \infty$ and $\sqrt{np_n}A(r_n) \to 0$ can be satisfied with the choice of

$$O(n^{\frac{1}{3}+\epsilon}) \leq r_n \leq o(n),$$

for some $\epsilon > 0$.

The result in Theorem 4.1 can be generalized from iid to α-mixing case, which we present in the next theorem.

Theorem 4.5. *1. Let $\{\mathbf{X}_t\}$ be a stationary α-mixing sequence with mixing coefficients $\{\alpha_h\}_{h\geq 0}$ and marginal distribution $\mathbf{X}$ that is multivariate regularly varying with index $\alpha > 1$. Let T_n be the conditional distance covariance between the angular and radial component defined in (4.1) such that the weight measure μ satisfies (4.2). Further, assume that $np_n^{1+\delta} \to \infty$ and $\sum_{h=0}^\infty h^2 \alpha_h^\delta < \infty$, for some $\delta \in (0, 1)$. Then*

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

2. In addition, if $\{r_n\}$ satisfies (4.3), 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.

The proof of Theorem 4.5 is given in Appendix B.

Note that the limiting distributions Q in Theorem 4.1 and Q' in Theorem 4.5 are both intractable and in practice, quantiles of the distributions are formed using resampling methods. While in the iid case this can be done straightforwardly, in the weakly dependent case this is more complicated and convergence may not be guaranteed. In our framework for the threshold selection, presented in the following section, we incorporate a subsampling scheme that overcomes this difficulty.

5. Threshold selection

In this section, we propose an procedure to select the threshold for estimating the spectral measure S from observations $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Let us first consider the case where a specific threshold r_n is given. Then (4.1) specifies the empirical distance covariance between R/r_n and $\mathbf{\Theta}$ conditional on $R > r_n$. Under the assumption (4.3), we have from Theorem 4.1,

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

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

$$\begin{aligned} H_0 &: \quad R/r_n \text{ and } \mathbf{\Theta} \text{ are independent given } R > r_n \\ H_1 &: \quad R/r_n \text{ and } \mathbf{\Theta} \text{ are not independent given } R > r_n. \end{aligned}$$

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

$$pv = \mathbb{P}\left(\int_{\mathbb{R}^{d+1}} |Q|^2 \mu(ds, dt) > u \right) \Big|_{u = n\hat{p}_n T_n}. \tag{5.1}$$

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 $\{r_k\}$. For convenience, we choose $\{r_k\}$ to be the empirical quantiles of $\{R_i\}$ based on a sequence of pre-specified levels $\{q_k\}$. From (5.1), 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 $R > r^*$, Θ 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 $\hat{F}_{(\Theta,R)|R>r_k}$ of (R_i, Θ_i) with $R_i > r_k$, $i = 1, \ldots, n$. 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 n_k from

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

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

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

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 $\mathbf{\Theta}$ are independent given $R > r_k$, then the $pv_{k1}, \ldots, pv_{km}$ will be iid $U(0,1)$ random variables, so that $\overline{pv}_k$ should center around 0.5.

- If R and $\mathbf{\Theta}$ are dependent given $R > r_k$, then the pv_{kj}'s will be well below 0.5 (close 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 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. For all the examples, we choose the weight function μ for distance covariance to be (3.2) with $\kappa = 1$. To satisfy the moment conditions, a log-transform is applied to the radial part R for all.

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 Θ become independent.

Let R be the absolute value of a t-distribution with 2 degrees of freedom and set

$$\Theta \overset{iid}{\sim} \begin{cases} U(0,1), & \text{if } R > r_{0.2}, \\ Beta(3,3), & \text{if } R \le r_{0.2}, \end{cases}$$

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

A sequence of candidate thresholds $\{r_k\}$ is selected to be the empirical upper quantiles of R corresponding to the levels $\{q_k\}$. 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 1(d) shows the mean p-value path. For the WBS algorithm, we set the threshold to be the smallest r_k when the fitted spline of the p-value remains above 0.45. 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 Θ_i's corresponding to the chosen thresholds is shown in Figure 1(e). We can see that the true tail angular cdf (i.e., $U(0,1)$) is accurately recovered.

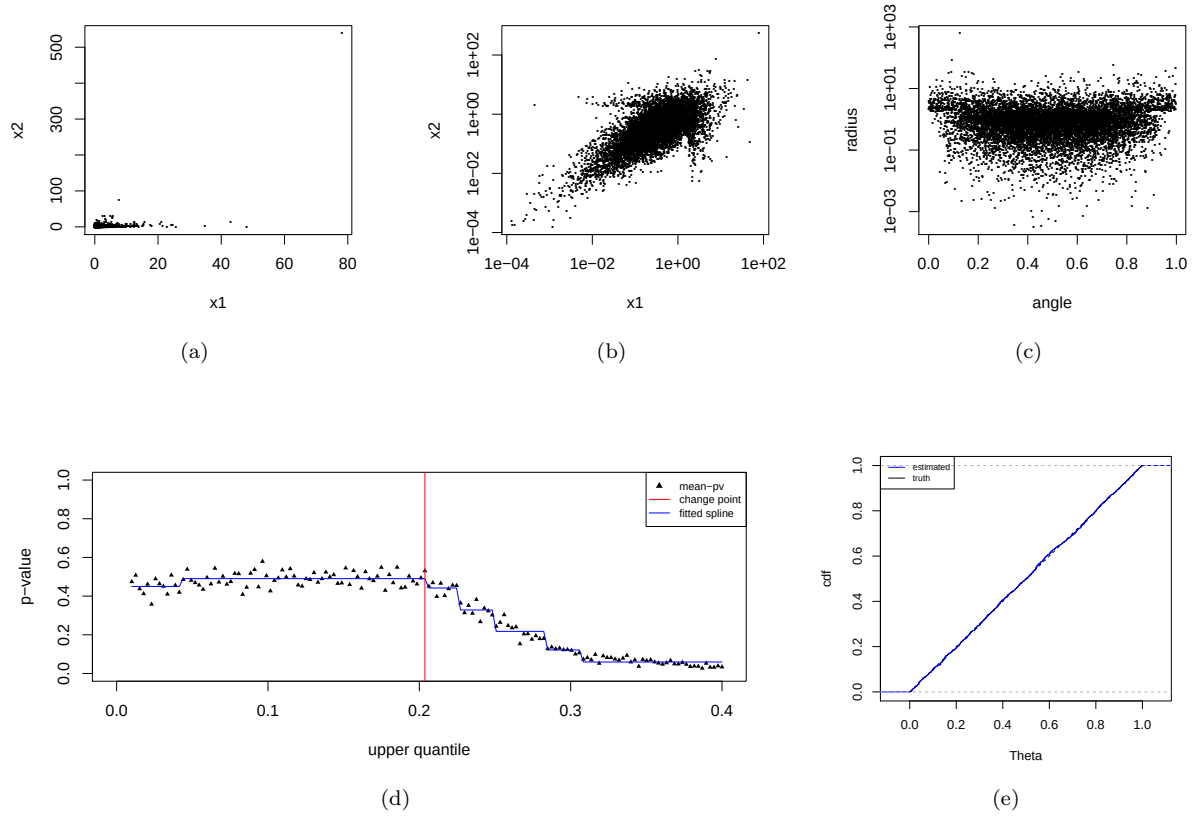


FIG 1. *Example 6.1. (a) scatterplot of* (X_{i1}, X_{i2})*; (b) scatterplot of* (X_{i1}, X_{i2}) *in log-log scale; (c) scatterplot of* (R_i, Θ_i)*; (d) mean p-value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); (e) estimated cdf of* Θ *using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

6.2. Simulated logistic data

We simulate data from a bivariate logistic distribution, which is bivariate regularly varying. Recall from Example 4.4 that (X_1, X_2) follows a bivariate logistic distribution if it has cdf

$$\mathbb{P}(X_1 < x_1, X_2 < x_2) = \exp(-(x_1^{-1/\gamma} + x_2^{-1/\gamma})^{\gamma}), \qquad \gamma \in (0, 1]. \tag{6.1}$$

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, the mean p-value $\overline{pv}_k$ is calculated using $m = 60$ random subsamples of size $n_k = 500 \cdot q_k$ from the observations with $R_i > r_k$.

Figures 2(a) , 2(b) and 2(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 (4.7), in Figure 2(e). So even though R and Θ are not independent for any threshold r_k, our procedures produce good estimates of the limiting distribution of Θ.

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

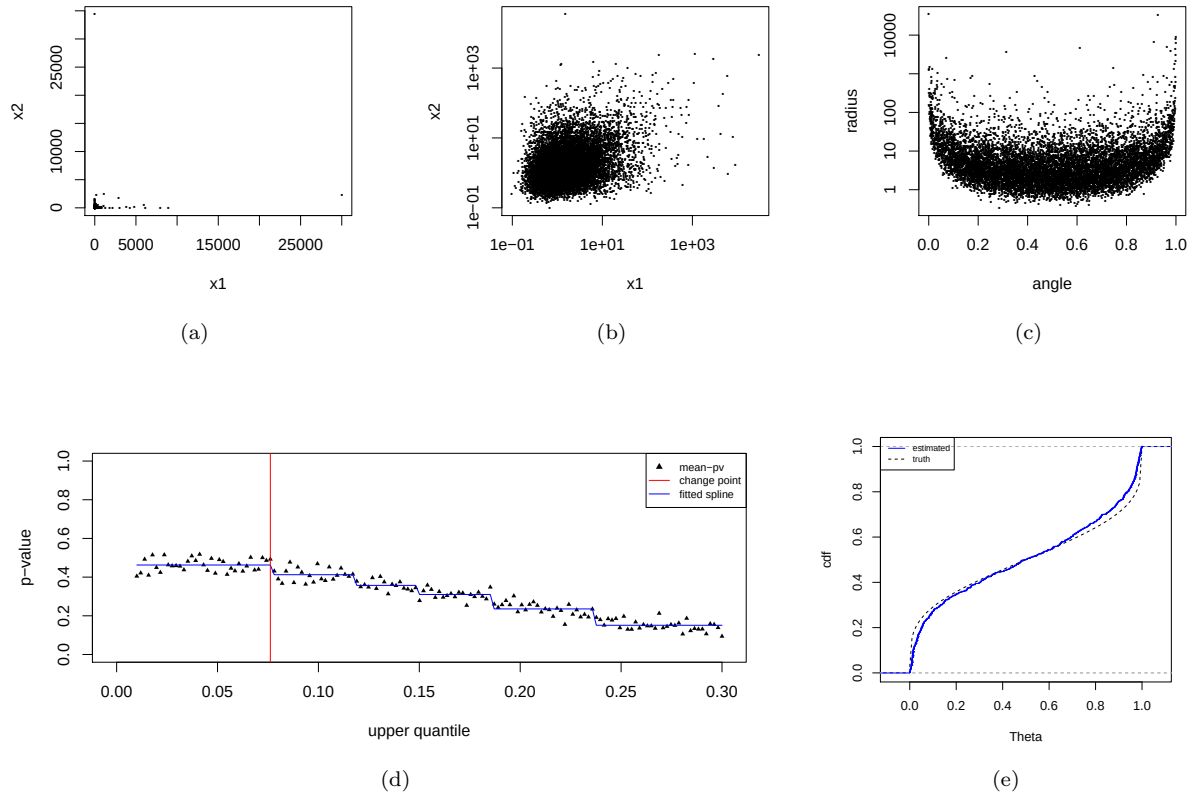


FIG 2. *Example 6.2. (a) scatterplot of* (X_{i1}, X_{i2})*; (b) scatterplot of* (X_{i1}, X_{i2}) *in log-log scale; (c) scatterplot of* (R_i, Θ_i)*; (d) mean p-value path (black triangles), fitted WBS spline (blue line), and the chosen threshold quantile (red vertical line); (e) estimated cdf of* Θ *using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

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. Figures 3(a)–3(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):

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

Again the mean p-value $\overline{pv}_k$ 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 Θ.

The mean p-value paths are shown in Figures 4(a)–4(c). The threshold levels selected for the three pairs are 9.6%, 7.4%, 16%, respectively. Figures 3(d)–3(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 DEM and CAD are almost independent.

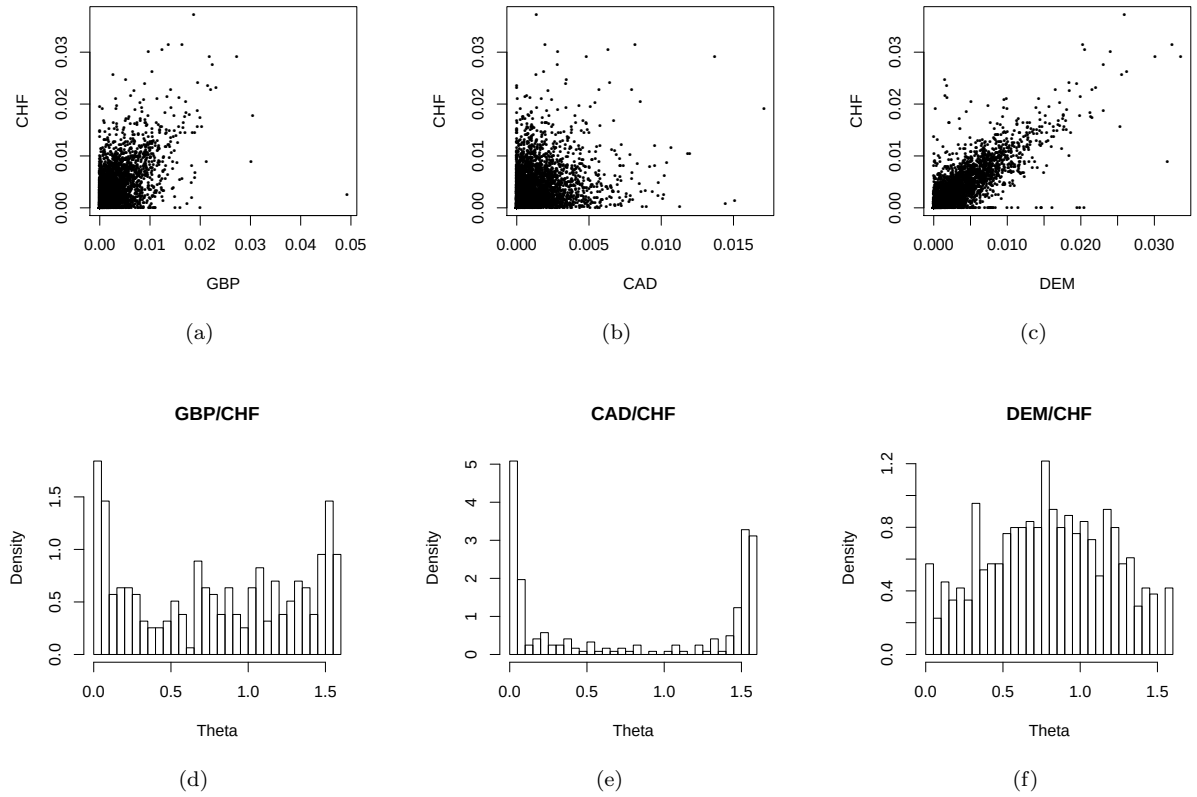
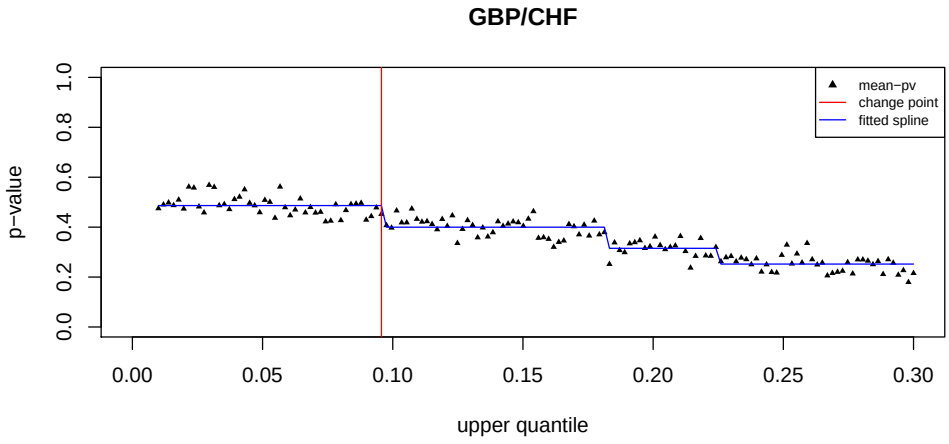
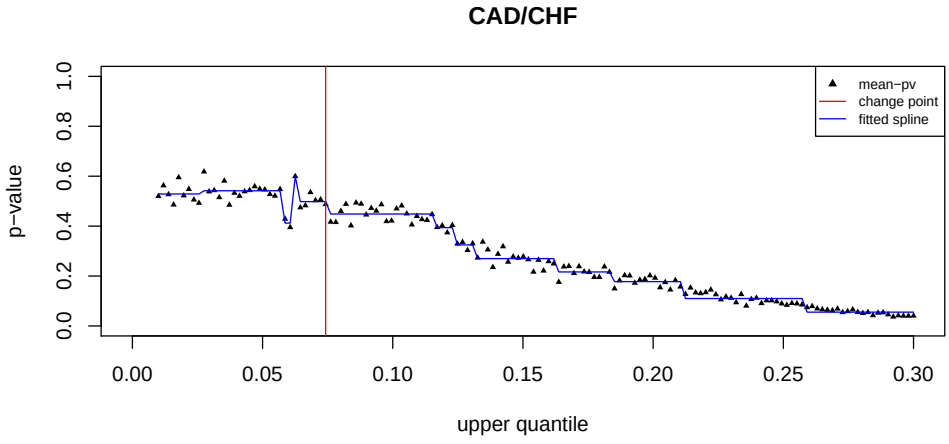


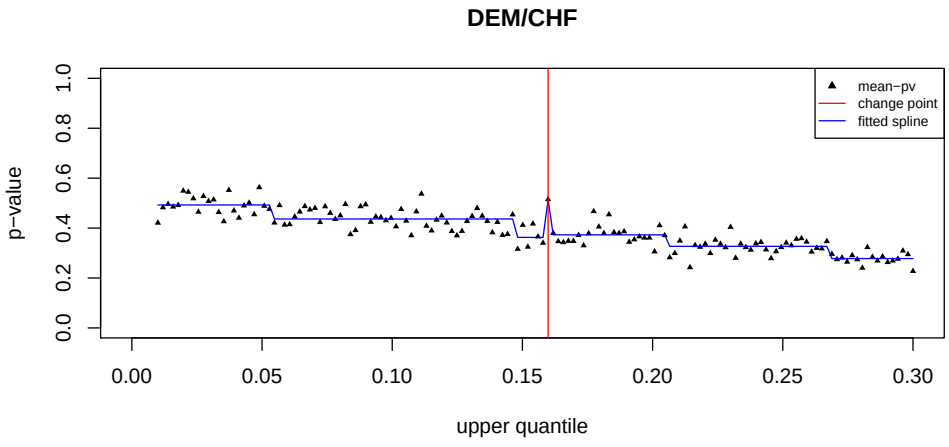
FIG 3. *Example 6.3. Analysis of the paired exchange rate returns: CHF/DEM, CHF/GBP, CHF/CAD with respect to USD between 1990-01-01 to 1998-12-31. (a)–(c): Scatter plots of the standardized paired exchange rate returns; (d)–(f): Estimated angular densities using the estimated thresholds chosen.*



(a)



(b)



(c)

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).*

7. Discussion

In this paper, we propose a threshold selection procedure for multivariate regular variation, for which R and Θ 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.

We note that the choice of norm in the polar coordinate transformation (2.1) 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 $\{(X_{i1}, X_{i2})\}_{i=1,\ldots,n}$ from the bivariate logistic distribution, where the cdf is given in (6.1), with $\gamma = 0.95$ and $n = 10000$. We apply the polar coordinate transformation with respect to the L_p-norm for $p = 0.2, 1, 5$ and compare the threshold selection results in Figure 5. Note that in the cases of the 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 Θ at all levels were shown to be significant. Indeed, this can be seen in Figure 6, 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. This is an interesting topic and is the subject of ongoing research.

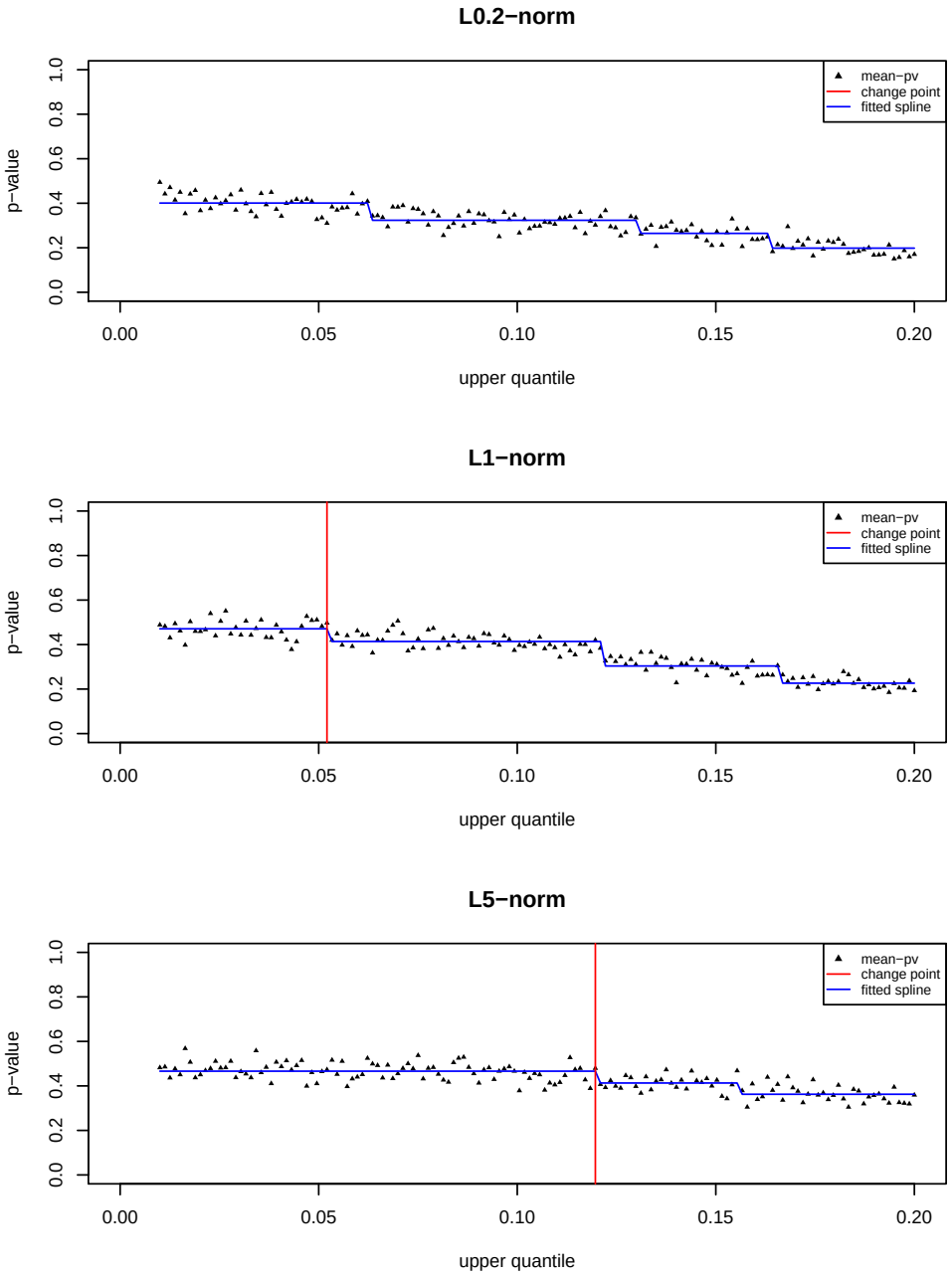


FIG 5. *Simulated logistic data of sample size* $n = 10000$ *with* $\gamma = 0.95$. *Threshold selection algorithm applied under the* $L_{0.2}$*-,* L_1*- and* L_5*-norms: mean p-value paths (black triangles), fitted WBS splines (blue lines) and the chosen threshold quantiles (red vertical line).*

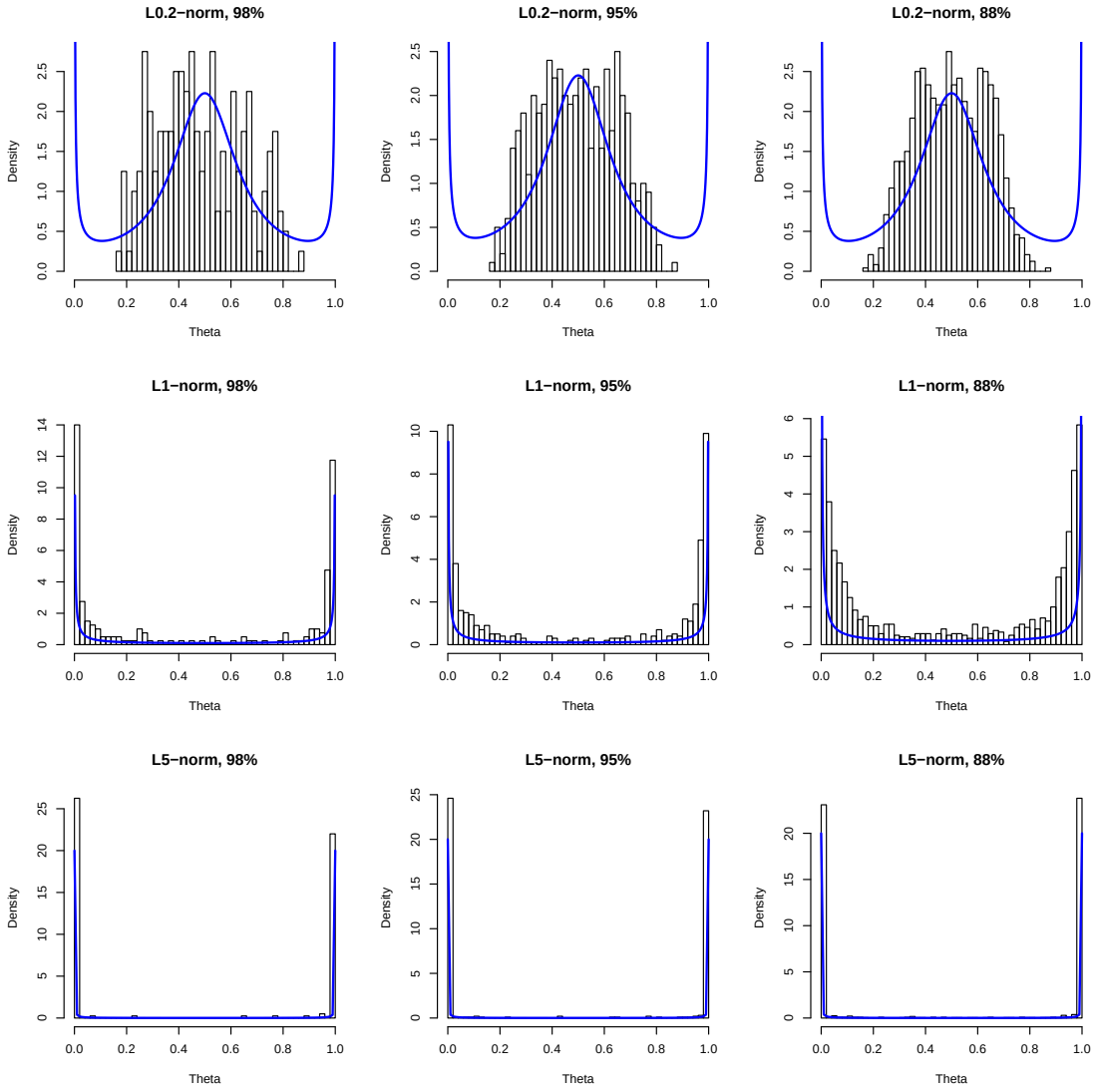


FIG 6. *Simulated logistic data of sample size* $n = 10000$ *with* $\gamma = 0.95$. *Histogram of* $X_1^p/(X_1^p + X_2^p)$ *for truncated levels* 2%, 5% *and* 12% *for* $p = 0.2, 1, 5$.

Acknowledgement

The foreign exchange rate data are obtained from OANDA from through R-package 'qrmtools'. We would like to thank Bodhisattva Sen for helpful discussions.

References

P. Bickel and M. Wichura. Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.*, 42:1656–1670, 1971.

P. Billingsley. *Probability and measure.* Wiley-Interscience, 3rd edition, 1995.

R. Davis, M. Matsui, T. Mikosch, and P. Wan. Applications of distance covariance to time series. *Bernoulli*, 2017.

L. de Haan and J. de Ronde. Sea and wind: Multivariate extremes at work. *Extremes*, 1:7–46, 1998.

P. Doukhan. *Mixing: Properties and Examples.* Springer-Verlag, New York, 1994.

A. Feuerverger. A consistent test for bivariate dependence. *Internat. Statis. Rev.*, 61(3):419–433, 1993.
P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42(6):2243–2281, 2014.
S. Jeon and R. Smith. Dependence structure of spatial extremes using threshold approach. arXiv:1209.6344v1, 2014.
H. Joe, R. Smith, and I. Weissman. Bivariate threshold methods for extremes. *JRSS. B.*, 54(1):171–183, 1992.
A. Mallik, B. Sen, M. Banerjee, and G. Michailidis. Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika*, 98(4):887–900, 2011.
E. Page. Continuous inspection schemes. *Biometrika*, 41(1):100–115, 1954.
D. Politis, J. Romano, and M. Wolf. Subsampling for heteroskedastic time series. *J. Econometrics*, 81: 281–317, 1997.
S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.
S. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer-Verlag, New York, 2007.
C. Stărică. Multivariate extremes for models with constant conditional correlations. *J. Empir. Finan.*, 6: 515–553, 1999.
G. Székely, M. Rizzo, and N. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35:2769–2794, 2007.

Appendix A: Proof of Theorem 4.1

We first note that

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

Writing $U_{jn} = \left(e^{isR_j/r_n}-\varphi_{R|r_n}(s)\right)\mathbf{1}_{\{R_j>r_n\}}$, $V_{jn} = \left(e^{it^T\mathbf{\Theta}_j}-\varphi_{\Theta|r_n}(t)\right)\mathbf{1}_{\{R_j>r_n\}}$, 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 $\mathbb{E}U_{jn}V_{jn}/p_n = \varphi_{R,\Theta|r_n}(s,t) - \varphi_{R|r_n}(s)\varphi_{\Theta|r_n}(t)$, it is convenient to mean correct the summands and obtain

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

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 4.1. The first part of Theorem 4.1 follows easily in a similar fashion. From here on, c will denote a generic a constant whose value may change from line to line.

Proof of Theorem 4.1(2). In order to show (4.4), it suffices to establish that

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

and

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

where (A.2) can be implied by

$$n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_2|^2 \mu(ds,dt) + n\hat{p}_n \int_{\mathbb{R}^{d+1}} \left(\frac{p_n}{\hat{p}_n}\right)^2 |\tilde{E}_3|^2 \mu(ds,dt) \xrightarrow{p} 0. \tag{A.3}$$

Notice that

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

Hence $\hat{p}_n/p_n \xrightarrow{p} 1$ and for (A.1) and (A.3), it is equivalent to prove that

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds,dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q(s,t)|^2 \mu(ds,dt) \tag{A.4}$$

and

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds,dt) + np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_3|^2 \mu(ds,dt) \xrightarrow{p} 0. \tag{A.5}$$

We will show the convergence (A.4) in Proposition A.1. By (4.3),

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

So that (A.5) holds provided

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds,dt) \xrightarrow{p} 0, \tag{A.6}$$

which follows in a similar fashion as Proposition A.1.

Recall from (2.2) that as $n \to \infty$, R/r_n and $\boldsymbol{\Theta}$ become asymptotically independent and converge to ν_α and S respectively. Denote the characteristic functions of the corresponding limit distributions by

$$\varphi_R(s) := \int_{\mathbb{R}} \exp(isr)\alpha r^{-\alpha-1} dr = \lim_{n\to\infty} \varphi_{R|r_n}(s) \tag{A.7}$$

$$\varphi_\Theta(t) := \int_{\mathbb{R}^d} \exp(it\theta) S(d\theta) = \lim_{n\to\infty} \varphi_{\Theta|r_n}(t). \tag{A.8}$$

□

Proposition A.1. *Assume that $np_n \to \infty$ as $n \to \infty$. 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

$$\operatorname{cov}(Q(s,t), Q(s',t')) = (\varphi_R(s-s') - \varphi_R(s)\varphi_R(-s'))(\varphi_\Theta(t-t') - \varphi_\Theta(t)\varphi_\Theta(-t')), \tag{A.9}$$

with $\varphi_R, \varphi_\Theta$ as defined in (A.7) and (A.8).

Proof of Proposition A.1. We first show that

$$\sqrt{np_n}\tilde{E}_1 \xrightarrow{d} Q(s,t), \quad \text{on } \mathcal{C}(\mathbb{R}^{d+1}) \tag{A.10}$$

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

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

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

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

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

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

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:

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

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 $\sqrt{np_n}\tilde{E}_1$ following the proof of Theorem 3.2 in Davis et al. (2017). Define $E_1((s,t),(s',t')]$ to be the process E_1 evaluated on the cube $((s,t),(s',t')] = (s,s'] \times \prod_{k=1}^{d}(t_i,t_i']$. Using the sufficient condition from Theorem 3 of Bickel and Wichura (1971), it suffices to show that

$$\mathbb{E}\left|\sqrt{np_n}\left(\tilde{E}_1((s,t),(s',t')]\right)\right|^2 \le c|s-s'|^\beta\prod_{k=1}^{d}|t_i-t_i'|^\beta$$

for some $\beta > 1$. It follows that

$$\begin{aligned}
&\mathbb{E}\left|\sqrt{np_n}\left(\tilde{E}_1((s,t),(s',t')]\right)\right|^2 \\
= \; & np_n\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\sum_{z_0=0,1}\sum_{z_1=0,1}\cdots\sum_{z_d=0,1}(-1)^{d+1-\sum_j p_j}\left(\left(e^{i(s+z_0(s'-s))R/r}\right)\prod_{k=1}^{d}e^{i(t_k+z_k(t_k'-t_k))\Theta_{jk}}\right.\right. \\
& \left.\left. -\mathbb{E}\left(e^{i(s+z_0(s'-s))R/r}\right)\prod_{k=1}^{d}e^{i(t_k+z_k(t_k'-t_k))\Theta_{jk}}\right)\right| \\
= \; & np_n\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\left((e^{isR_j/r_j}-e^{is'R_j/r_j})\prod_{k=1}^{d}(e^{it_k\Theta_{jk}}-e^{it_k'\Theta_{jk}})\right.\right. \\
& \left.\left. -\mathbb{E}\left[(e^{isR/r}-e^{is'R/r})\prod_{k=1}^{d}(e^{it_k\Theta_k}-e^{it_k'\Theta_k})\mathbf{1}_{\{R>r_n\}}\right]\right)\mathbf{1}_{\{R_j>r_n\}}\right|^2
\end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}\left[\left|(e^{isR/r}-e^{is'R/r})\prod_{k=1}^{d}(e^{it_k\Theta_k}-e^{it'_k\Theta_k})-\mathbb{E}\right|^2\middle|R>r_n\right]\\
&\leq c|s-s'|^\beta\prod_{k=1}^{d}|t_i-t'_i|^\beta\mathbb{E}\left[(R/r_n)^\beta\prod_{k=1}^{d}|\Theta_k|^\beta|R>r_n\right]\\
&< \infty,
\end{aligned}$$

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

Define the bounded set

$$K_\delta=\{(s,t)|\ \delta<|s|<1/\delta,\delta<|t|<1/\delta\}$$

for $\delta<.5$. Then, using (A.10), we have from the continuous mapping theorem,

$$np_n\int_{K_\delta}|\tilde{E}_1|^2\mu(ds,dt)\xrightarrow{d}\int_{K_\delta}|Q(s,t)|^2\mu(ds,dt). \tag{A.11}$$

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

$$\begin{aligned}
&\mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2\\
=\ & np_n\mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\frac{U_{jn}V_{jn}}{p_n}-\mathbb{E}\left[\frac{U_{jn}V_{jn}}{p_n}\right]\right)\right|^2\\
\leq\ & \frac{\mathbb{E}|U_{jn}V_{jn}-\mathbb{E}U_{jn}V_{jn}|^2}{p_n}\\
\leq\ & \frac{4\,\mathbb{E}|U_{jn}V_{jn}|}{p_n}\\
=\ & \frac{4\,\mathbb{E}\left[|e^{isR_j/r_n}-\varphi_{R|r_n}(s)||e^{it^T\boldsymbol{\Theta}_j}-\varphi_{\Theta|r_n}(t)|\mathbf{1}_{\{R_j>r_n\}}\right]}{p_n}\\
\leq\ & \frac{4\,\mathbb{E}\left[\left|\left(1\wedge|s|^\beta(|\frac{R_j}{r_n}|^\beta+\mathbb{E}[|\frac{R}{r_n}|^\beta|\frac{R}{r_n}>1])\right)\left(1\wedge|t|^2(|\boldsymbol{\Theta}_j|^2+\mathbb{E}[|\boldsymbol{\Theta}|^2|\frac{R}{r_n}>1])\right)\right|^2\mathbf{1}_{\{R_j>r_n\}}\right]}{p_n}\\
\leq\ & 4\,\mathbb{E}\left[\left|\left(1\wedge|s|^\beta(|R_j/r_n|^\beta+\mathbb{E}[|R/r_n|^\beta|R>r_n])\right)\left(1\wedge|t|^2\right)\right|^2|R>r_n\right]\\
\leq\ & c(1\wedge|s|^\beta)(1\wedge|t|^2).
\end{aligned}$$

Therefore for any $\epsilon>0$,

$$\begin{aligned}
\lim_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}\left[np_n\int_{K_\delta^c}|\tilde{E}_1|^2\mu(ds,dt)>\epsilon\right] &\leq \frac{1}{\epsilon}\lim_{\delta\to0}\limsup_{n\to\infty}\int_{K_\delta^c}\mathbb{E}|\sqrt{np_n}\tilde{E}_1|^2\mu(ds,dt)\\
&\leq \frac{1}{\epsilon}\lim_{\delta\to0}\limsup_{n\to\infty}\int_{K_\delta^c}c(1\wedge|s|^\beta)(1\wedge|t|^2)\mu(ds,dt)\\
&\to 0
\end{aligned}$$

by the dominated convergence theorem. This combined with (A.11) shows the convergence of $np_n\int|\tilde{E}_1|^2\mu(ds,dt)$ to $\int|Q(s,t)|^2\mu(ds,dt)$, and hence completes the proof of the proposition.

□

Proof of Theorem 4.1(2) (cont.). Now it remains to show (A.6). Similar to the proof of Proposition A.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. □

Proof of Theorem 4.1(1). Similar to the proof of Theorem 4.1(2), it suffices to show that

$$\int |\tilde{E}_i|^2 \mu(ds, dt) \xrightarrow{p} 0, \quad i = 1, 2, 3. \tag{A.12}$$

The convergence (A.12) for $i = 1, 2$ follows trivially from the more general results (A.4) and (A.6) in the proof of Theorem 4.1(2). Hence it suffices to show

$$\int |\tilde{E}_3|^2 \mu(ds, dt) \to 0, \tag{A.13}$$

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

Let $P_{R,\Theta|r_n}(\cdot) = P\left[\left(\frac{R}{r_n}, \mathbf{\Theta}\right) \in \cdot | \frac{R}{r_n} > 1\right]$ and $P_{R|r_n}, P_{\Theta|r_n}$ be the corresponding marginal measures. Then from (2.2),

$$P_{R,\Theta|r_n} - P_{R|r_n}P_{\Theta|r_n} = (P_{R,\Theta|r_n} - c\nu_\alpha \times S) - (P_{R|r_n} - c\nu_\alpha) \times S - c\nu_\alpha \times (P_{\Theta|r_n} - S) \xrightarrow{v} 0,$$

and hence for fixed (s,t),

$$\tilde{E}_3(s,t) = \int e^{isR/r_n + it^T\mathbf{\Theta}} (P_{R,\Theta|r_n} - P_{R|r_n}P_{\Theta|r_n})(dR, d\mathbf{\Theta}) \to 0.$$

Since

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

the result (A.13) follows from (4.2) and dominated convergence.

This concludes the proof. □

Appendix B: Proof of Theorem 4.5

Following the same notation and steps as the proof of Theorem 4.1 in Appendix A, it suffices to prove the following convergences for the mixing case:

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

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_1|^2 \mu(ds, dt) \xrightarrow{d} \int_{\mathbb{R}^{d+1}} |Q'(s,t)|^2 \mu(ds, dt) \tag{B.2}$$

for some Gaussian process Q', and

$$np_n \int_{\mathbb{R}^{d+1}} |\tilde{E}_2|^2 \mu(ds, dt) \xrightarrow{p} 0. \tag{B.3}$$

We prove (B.1) and (B.2) in Propositions B.1 and B.2, respectively. The proof of (B.3) follows in a similar fashion.

Proposition B.1. *Assume that $np_n^{1+\delta} \to \infty$ for some $\delta \in (0,1)$, and that $\sum_{h=0}^{\infty} \alpha_h^\delta < \infty$, then (B.1) holds.*

Proof. We have

$$\mathbb{E}\left|\frac{\hat{p}_n}{p_n} - 1\right|^2 = \mathbb{E}\left|\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\mathbf{1}_{\{R_j > r_n\}}}{p_n} - 1\right)\right|^2$$

$$
\begin{aligned}
&\leq \frac{1}{n^2 p_n^2}\left(\mathbb{E}\left|\mathbf{1}_{\{R_1>r_n\}}-p_n\right|^2+2\sum_{h=2}^{n}\left(1-\frac{h}{n}\right)\mathbb{E}\left|\mathbf{1}_{\{R_1>r_n\}}-p_n\right|\left|\mathbf{1}_{\{R_{h+1}>r_n\}}-p_n\right|\right)\\
&\leq \frac{2}{np_n^2}\sum_{h=0}^{\infty}\mathbb{E}\left|\mathbf{1}_{\{R_1>r_n\}}-p_n\right|\left|\mathbf{1}_{\{R_{h+1}>r_n\}}-p_n\right|\\
&\leq \frac{c}{np_n^2}\sum_{h=0}^{\infty}\alpha_h^{\delta}(E\left|\mathbf{1}_{\{R_1>r_n\}}-p_n\right|^{2/(1-\delta)})^{1-\delta} \qquad \text{(B.4)}\\
&\leq \frac{c}{np_n^2}\sum_{h=0}^{\infty}\alpha_h^{\delta}(E\left|\mathbf{1}_{\{R_1>r_n\}}-p_n\right|^{2})^{1-\delta} \qquad \text{(B.5)}\\
&= \frac{c}{np_n^{1+\delta}}\sum_{h=0}^{\infty}\alpha_h^{\delta}\\
&\to 0,
\end{aligned}
$$

where (B.4) follows from the α-mixing condition (see Section 1.2.2, Theorem 3(a) of Doukhan (1994)), and (B.5) follows from the facts that $2/(1-\delta)>2$ and $|\mathbf{1}_{\{R_1>r_n\}}-p_n|<1$. □

Proposition B.2. *Assume that* $np_n\to\infty$, *and*

$$\sum_{h=0}^{\infty}h^2\alpha_h^{\delta}<\infty \qquad \text{(B.6)}$$

for some $\delta\in(0,1)$, *then* (B.2) *holds.*

Proof. For fixed (s,t), we can write

$$\sqrt{n\hat{p}_n}\tilde{E}_1=\frac{1}{\sqrt{n\hat{p}_n}}\sum_{k=1}^{\hat{p}_n}(U_{j_k^{(n)}n}V_{j_k^{(n)}n}-\mathbb{E}U_{1n}V_{1n}),$$

where $(j_k^{(n)})_{k=1,\ldots,\hat{p}_n}$ are the subsequence for which $R_{j_k}>r_n$. Then $U_{j_k^{(n)}n}V_{j_k^{(n)}n}-\mathbb{E}U_{1n}V_{1n}$ is a triangular array where the rows are stationary and α-mixing with coefficients $\alpha_h^{(n)}$ such that $\alpha_h^{(n)}<\alpha_h$. For any increasing b_n,

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{1}{\sqrt{b_n}}\sum_{k=1}^{b_n}(U_{j_k^{(n)}n}V_{j_k^{(n)}n}-\mathbb{E}U_{1n}V_{1n})\right)\\
\to\ &\mathbb{E}\left|U_{j_0^{(n)}n}V_{j_0^{(n)}n}-\mathbb{E}U_{1n}V_{1n}\right|^2+2\sum_{h=1}^{\infty}\mathrm{Re}\mathbb{E}\left((U_{j_0^{(n)}n}V_{j_0^{(n)}n}-\mathbb{E}U_{1n}V_{1n})\right)\left((U_{j_h^{(n)}n}V_{j_h^{(n)}n}-\mathbb{E}U_{1n}V_{1n})\right)\\
=:\ &\sigma^2(s,t)\\
\leq\ &2+c\sum_{h=1}^{\infty}(\alpha_h^{(n)})^{\delta'}\left(\mathbb{E}\left|U_{j_0^{(n)}n}V_{j_0^{(n)}n}-\mathbb{E}U_{1n}V_{1n}\right|^2\right)^{1-\delta'}\\
\leq\ &2+c\sum_{h=1}^{\infty}\alpha_h^{\delta'}<\infty.
\end{aligned}
$$

Since $|U_{j_k^{(n)}n}V_{j_k^{(n)}n}-\mathbb{E}U_{1n}V_{1n}|\leq 2$ and (B.6) holds, the central limit theory for α-mixing triangular arrays (see Theorem A.1 of Politis et al. (1997)) implies that for fixed (s,t), $\sqrt{n\hat{p}_n}\tilde{E}_1$ converges to a normal variable with mean zero and variance $\sigma^2(s,t)$.

The finite dimensional distributional convergence of $\sqrt{n\hat{p}_n}\tilde{E}_1$ to a Gaussian process $Q'(s,t)$ can be generalized using the Cramér-Wold device and we omit the calculation of the covariance structure. The tightness condition for the functional convergence and the condition (A.11) can be verified through a similar argument. □