A nonparametric estimator of the extremal index

Juan-Juan Cai *

Delft University of Technology, The Netherlands

November 18, 2019

Abstract

Clustering of extremes has a large societal impact. The extremal index, a number in the unit interval, is a key parameter in modelling the clustering of extremes. We build a connection between the extremal index and the stable tail dependence function, which enables us to compute the value of extremal indices for some time series models. We also construct a nonparametric estimator of the extremal index and an estimation procedure to verify $D^{(d)}(u_n)$ condition, a local dependence condition often assumed for studying the extremal index. We prove that the estimator is asymptotically normal. The simulation study which compares our estimator to two existing methods shows that our method has better finite sample properties. We apply our method to estimate the expected durations of heatwaves in the Netherlands and in Greece.

Keywords: Serial dependence; Extreme clusters; Stable tail dependence function; $D^{(d)}(u_n)$ condition

1 Introduction

A cluster of extremes refers to the occurrence of multiple extreme observations (i.e. high level exceedances) within a short period of time. When extreme events happen sequentially, it often has a destructive impact on our society. For instance, hot temperature extremes in successive days increase the risk of mortality, drought, wildfire and others. The clustering of extremes is due to the serial extremal dependence of time series data. The extremal index $\theta \in (0,1]$ (cf. Definition 2.1) is a key parameter to measure the strength of such dependence. The smaller value of θ indicates stronger extremal dependence, while $\theta=1$ corresponds to the case where the extremes are independent, meaning that there is no clustering of extremes. The extremal index provides two other insights on the behavior of the clustering. First, it equals the reciprocal of the expected clustering size, that is the number of extreme observations in a cluster (Leadbetter [1983]). Second, O'Brien [1987] shows that θ equals a conditional probability that measures to what extent extremes cluster together. The primarily goal of this paper is to develop an estimation for the extremal index.

The existing methods for estimating θ can be grouped into two categories. The first category of estimators requires two tuning parameters, namely a threshold and a block length.

^{*}Address for correspondence: Juan-Juan Cai, Delft Institute of Applied Mathematics, Delft University of Technology, van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands. $\textit{E-mail:}\ j.j.cai@tudelft.nl$

The threshold indicates a level above which observations are considered as extremes. And the block length defines the size of a cluster. The representatives of this group are the blocks and runs estimators (Smith and Weissman [1994], Weissman and Novak [1998]), which are based on the two aforementioned interpretations of θ , respectively. The choice of tuning parameters is notoriously difficult to make in extreme value theory in general. The second category of estimation methods needs only one tuning parameter. For instance, the estimator developed by Ferro and Segers [2003] requires only a choice of the threshold. The maximum likelihood estimators (MLE) of θ studied by Northrop [2015] and Berghaus and Bücher [2018] needs a choice of the block length. Note that this MLE is a nonparametric method and is based on the following intuition: for (X_i) a stationary sequence and F the marginal distribution function, $F(\max_{i=1,\dots,r_n} X_i)$ is approximately exponentially distributed with parameter θ , where r_n denotes the block length. There are many other references on the estimation of extremal index such as Hsing [1993], Laurini and Tawn [2003], Robert [2009], and Ancona-Navarrete and Tawn [2000].

In this paper, we develop an estimator of θ assuming a local dependence condition, namely $D^{(d)}(u_n)$ condition (cf. Section 2.1). Our estimator requires two inputs: a threshold and the value of d. It turns out that the choice of d is generally not unique (cf. Theorem 2.2). We construct a consistent estimator for the smallest d such that $D^{(d)}(u_n)$ condition is valid. Our proposed estimation procedure for θ requires only a choice of the threshold. There are three major contributions of this paper. First, we prove the asymptotic normality of the proposed estimator of θ . Among the existing literature on estimating θ , only a few have addressed the asymptotic normality, cf. Hsing [1993], Weissman and Novak [1998] and Berghaus and Bücher [2018]. Yet, our result is proved under a rather general setting compared to the existing ones. Second, we build a representation of θ using the stable tail dependence function of the random vector (X_1, X_2, \dots, X_d) . The result links the extremal behavior of a stationary sequence to multivariate extreme value theory and it provides a convenient way to compute the value of θ for some time series models. Third, our estimation procedure for d verifies the $D^{(d)}(u_n)$ condition, which is often assumed for studying the extremal index. For instance Leadbetter et al. [1983] has assumed $D^{(1)}(u_n)$ and Süveges [2007] has studied a likelihood estimator of θ under $D^{(2)}(u_n)$ condition. To the best of our knowledge, this is the first attempt to propose an inference procedure for verifying $D^{(d)}(u_n)$ condition based on asymptotic results.

The rest of this paper is organized as follows. The estimations of θ and d and their asymptotic properties are discussed in Section 2. In Section 3, we demonstrate the computation of θ via the stable tail dependence function for some examples, and compare the finite sample performance of our estimator of θ to two existing methods. In Section 4, we apply our estimator to compare the durations of heatwaves in the Netherlands and in Greece, using the temperature data from two weather stations of the two countries, respectively. Sections 5 and 6 contain the proofs for the main theorems.

2 Estimations and asymptotic properties

2.1 Estimation of θ assuming that d is known

Let $\{X_i, i = 1, ..., n\}$ be a stretch of length n from a strictly stationary sequence of random variables. The extremal index of the sequence is defined as below.

Definition 2.1 Suppose that the distribution function F of X_1 is in the domain of attraction of some extreme value distribution, that is, there exists sequences $a_n > 0$ and $b_n \in \mathbf{R}$ such that for all x,

$$F^n(a_n x + b_n) = G(x),$$

where G is an extreme value distribution. If there exists some $0 < \theta \le 1$ such that

$$\mathbb{P}\left(\frac{\max_{1\leq i\leq n} X_i - b_n}{a_n} \leq x\right) = G^{\theta}(x)$$

for all x, then θ is the extremal index of the sequence $\{X_i, i \geq 1\}$.

Our goal is to estimate the extremal index θ . The starting point is Corollary 1.3 from Chernick et al. [1991]. For stating the existence of θ , the corollary assumes two "mixing" conditions in the dependence of the sequence, namely $D(u_n)$ and $D^{(d)}(u_n)$. Let $u_n(\tau), \tau > 0$ be such that

$$\lim_{n \to \infty} n \mathbb{P}(X_1 > u_n(\tau)) = \tau.$$

When it does not cause any misunderstanding, we write u_n instead of $u_n(\tau)$.

Condition $D(u_n)$ For any integers $1 \leq i_1 < \cdots < i_q < j_1 < \cdots < j_{q'} \leq n$ for which $j_1 - i_q \geq l$, we have

$$\left| \mathbb{P} \left(\max_{1 \le t \le q} X_{i_t} \le u_n, \max_{1 \le t \le q'} X_{j_t} \le u_n \right) - \mathbb{P} \left(\max_{1 \le t \le q} X_{i_t} \le u_n \right) \mathbb{P} \left(\max_{1 \le t \le q'} X_{j_t} \le u_n \right) \right| \le \alpha_{n,l},$$

and $\lim_{n\to\infty} \alpha_{n,l_n} = 0$ for some sequence $l_n = o(n)$ and $l_n \to \infty$.

Condition $D^{(d)}(u_n)$ There exist a positive integer d, sequences of integers r_n and l_n such that $r_n \to \infty$, $n\alpha_{n,l_n}/r_n \to 0$, $l_n/r_n \to 0$, and

$$\lim_{n \to \infty} n \mathbb{P}(X_1 > u_n \ge M_{2,d}, M_{d+1,r_n} > u_n) = 0, \tag{2.1}$$

where $M_{i,j} := -\infty$ for i > j and $M_{i,j} := \max_{1 \le t \le j} X_t$ for $i \le j$.

Condition $D(u_n)$ is a standard condition on long range dependence when studying the extremal behaviour of a stationary sequence (see e.g. Leadbetter et al. [1983]). Condition $D^{(d)}(u_n)$ is a local mixing condition, which will be further studied in Section 2.2.

Proposition 2.1 (Chernick et al. [1991], Corollary 1.3) Let $\{X_n\}$ be a stationary sequence of random variables such that for some $d \ge 1$ the conditions $D(u_n)$ and $D^{(d)}(u_n)$ hold for $u_n = u_n(\tau)$ for all $\tau > 0$. Then the extremal index of $\{X_n\}$, θ exists if and only if

$$\lim_{n \to \infty} \mathbb{P}(M_{2,d} \le u_n | X_1 > u_n) = \theta, \tag{2.2}$$

for all $\tau > 0$.

We develop our estimator of θ based on this limit result assuming that d is known. First, we choose the intermediate quantile $F^{-1}(1-k/n)$ as the threshold u_n , where k=k(n) is an intermediate sequence such that $k\to\infty$ and $n/k\to\infty$ as $n\to\infty$. We shall estimate this threshold with $X_{n-k,n}$, where $\{X_{1,n} \leq X_{2,n} \cdots \leq X_{n,n}\}$ are the order statistics of the sample. Then making use of the empirical probability measure, we have

$$\theta \approx \frac{n}{k} \mathbb{P}(M_{2,d} \le F^{-1}(1 - k/n) < X_1)$$

$$\approx \frac{n}{k} \cdot \frac{1}{n} \sum_{i=1}^{n-d+1} \mathbb{1} \left\{ M_{i+1,i+d-1} \le X_{n-k,n} < X_i \right\} =: \hat{\theta}_n(d),$$

where $\mathbbm{1}$ denotes an indicator function. Throughout we assume that F is a continuous function. Let $U_i = 1 - F(X_i)$, i = 1, ..., n, which become uniform random variables and let $\{U_{i,n}, i = 1, ..., n\}$ be the order statistics. Observe that $\mathbbm{1}$ $\{M_{i+1,i+d-1} \leq X_{n-k,n} < X_i\} = \mathbbm{1}$ $\{mU_{i+1,i+d-1} \geq U_{k,n} > U_i\}$, where $mU_{i+1,i+d-1} = \min_{i+1 \leq j \leq i+d-1} U_j$. Therefore $\hat{\theta}_n(d)$ can be written as

$$\hat{\theta}_n(d) = \frac{1}{k} \sum_{i=1}^{n-d+1} \mathbb{1} \left\{ mU_{i+1,i+d-1} \ge U_{k,n} > U_i \right\}, \tag{2.3}$$

It is more convenient to formulate the conditions using U_i than using X_i .

To obtain the asymptotic normality, we need a ϕ -mixing conditions on the sequence $\{\mathbb{1}\left\{U_i \leq \frac{k}{n}\right\}, i = 1, \ldots, n\}$. Let $\mathcal{H}_l^s = \sigma(\mathbb{1}\left\{U_i \leq \frac{k}{n}\right\}, l \leq i \leq s)$. Define

$$\phi(l) = \max_{s \ge 1} \sup_{A \in \mathcal{H}_i^s, B \in \mathcal{H}_{s+l}^n, \mathbb{P}(A) > 0} |\mathbb{P}(B|A) - P(B)|. \tag{2.4}$$

(A1) $\lim_{l\to\infty} \phi(l) = 0$

Note that condition (A1) implies $D(u_n)$ condition and also the so called absolute regularity of the sequence, cf. Bradley [2005]. We also need a strengthening condition of $D^d(u_n)$.

(A2) There exist r_n and l_n such that $r_n \to \infty$, $\frac{n}{r_n}\phi(l_n) \to 0$, $\frac{l_n}{r_n} \to 0$ and for any $x,y \in [1/2,3/2]$,

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=d+1}^{r_n} \mathbb{P}(U_1 < \frac{kx}{n} < mU_{2,d}, U_i < \frac{ky}{n}) = 0.$$
 (2.5)

When d = 1, Condition (A2) becomes the so called $D'(u_n)$ condition which implies $\theta = 1$ and for d = 2, it is equivalent to $D''(u_n)$ condition, cf. Leadbetter and Nandagopalan [1989].

In order to obtain the asymptotic variance of $\hat{\theta}_n(d)$, we need the following two assumptions on the tail dependence structure of (X_1, \ldots, X_{r_n}) .

(A3) For $x, y \in [1/2, 3/2]$,

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P}\left(U_1 < \frac{kx}{n}, U_{i+1} < \frac{ky}{n}\right) = \Lambda_1(x, y) \in [0, \infty).$$
 (2.6)

(A4)

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P}\left(U_1 < \frac{k}{n}, U_{i+1} < \frac{k}{n} < mU_{2+i, i+d}\right) = \lambda_1 \in [0, \infty). \tag{2.7}$$

The next condition makes sure that θ exists, cf. (2.12) and further it is used to control the bias of the limit distribution of the estimator.

(A5) There exists a $\rho > 0$ such that for j = d and j = d - 1, as $t \to 0$,

$$\sup_{1/2 \le x \le 3/2} \left| \frac{1}{t} \mathbb{P} \left(mU_{1,j} < tx \right) - \ell_j(x, \dots, x) \right| = O(t^{\rho}),$$
 (2.8)

where ℓ_j is defined in (2.10). Here we also assume that $\ell_d(x_1, \ldots, x_d)$ exists for $(x_1, \ldots, x_d) \in \mathbb{R}^d_+$.

Theorem 2.1 Assume that Condition $D(u_n)$ and Conditions A(1)-A(5) hold, $\sum_{i=r_n}^n \left(1 - \frac{i}{n}\right) \phi(i) = o(1)$, $\frac{r_n k}{n} = o(1)$ and that $k = o\left(n^{2\rho/(2\rho-1)}\right)$. Then,

$$\sqrt{k}(\hat{\theta}_n(d) - \theta) \stackrel{d}{\to} N(0, \sigma^2),$$
 (2.9)

where $\sigma^2 = \theta(1 - 2\lambda_1) + \theta^2(2\Lambda_1(1, 1) - 1)$.

The proof of this theorem is provided in Section 5.

Remark 2.1 Hsing [1993] has already studied this estimator defined in (2.3) and obtained its asymptotic normality under m-dependence. Weissman and Novak [1998] has established the asymptotic normality for runs and blocks estimators for a deterministic threshold, that is using $F^{-1}(1-k/n)$ (instead of $X_{n-k,n}$) as the threshold. Using a data-depending threshold and assuming a general mixing condition impose more technical challenges for proving the asymptotic result.

2.2 Estimation of d^* : statistical inference on $D^{(d)}(u_n)$ condition

In practice, d is unknown and hence has to be estimated. In this subsection, we provide a procedure to estimate d. We aim to identify a d such that both (2.2) and $D^{(d)}(u_n)$ condition hold. It turns out that the choice of d to meet such a requirement is not unique.

Observe that (2.2) and $D^{(d)}(u_n)$ condition are about the tail dependence structure of the random vectors (X_1, \ldots, X_d) and (X_1, \ldots, X_{r_n}) , respectively. To specify the notion of tail dependence, we consider the stable tail dependence function ℓ_d of (X_1, \ldots, X_d) , which, if exists, is defined as

$$\ell_d(x_1, \dots, x_d) = \lim_{n \to \infty} n \mathbb{P} \left(X_1 > u_n(x_1) \text{ or } \dots \text{ or } X_d > u_n(x_d) \right),$$

$$= \lim_{n \to \infty} n \mathbb{P} \left(U_1 < \frac{x_1}{n} \text{ or } \dots \text{ or } U_d < \frac{x_d}{n} \right), \tag{2.10}$$

for $(x_1, \ldots, x_d) \in \mathbb{R}^d_+$ and $d \geq 1$. It is easy to see that the stable tail dependence function has homogeneity of order 1: for a positive constant a,

$$\ell_d(ax_1, \dots, ax_d) = a\ell_d(x_1, \dots, x_d). \tag{2.11}$$

See Chapter 6 in de Haan and Ferreira [2006] for more properties of ℓ_d . The tail process introduced in Basrak and Segers [2009] is a commonly used tool for modeling the extremal dependence of a stationary sequence. Suppose that the tail process of (X_t) exists and we denote the tail process by (Y_t) . Then ℓ_d also exists for any finite d and moreover for $x_1 \ge 1$,

$$\ell_d(x_1, \dots, x_d) = \mathbb{P}\left(Y_1 > x_1^{\gamma} \text{ or } \dots \text{ or } Y_d > x_d^{\gamma}\right),$$

where $\gamma > 0$ is the extreme value index of X_1 . The existence of the tail process requires that any finite-dimensional distribution of (X_t) are multivariate regularly varying whereas ℓ_d exists for a more general setting and it does not impose any constraint on the distribution of X_1 (apart from continuity).

We denote by $\ell_d(\mathbf{1}_d)$ the value of function ℓ_d evaluated at the d-dimensional unity vector $\mathbf{1}_d := (1, \ldots, 1), \ d \geq 1$, and $\ell_0(1) := 0$. The link between θ and ℓ_d is established in the following theorem.

Theorem 2.2 Assume that conditions $D(u_n)$ and $D^{(d)}(u_n)$ hold for some $d \ge 1$ and ℓ_d exists. Then,

$$\theta = \ell_d(\mathbf{1}_d) - \ell_{d-1}(\mathbf{1}_{d-1}) \in [0, 1]. \tag{2.12}$$

Moreover, if ℓ_s exists for any finite s, then there exists a positive integer $d^* \leq d$ such that for any $s \geq d^*$, $D^{(s)}(u_n)$ condition holds and

$$\ell_s(\mathbf{1}_s) - \ell_{s-1}(\mathbf{1}_{s-1}) = \theta;$$

and, for any $1 \le s < d^*$,

$$\ell_s(\mathbf{1}_s) - \ell_{s-1}(\mathbf{1}_{s-1}) > \theta.$$

Proof. By Proposition 2.1, in order to prove (2.12) it is sufficient to show that for any $u_n(\tau)$ such that $\lim_{n\to\infty} n\mathbb{P}(X_1 > u_n(\tau)) = \tau$, we have $\lim_{n\to\infty} \mathbb{P}(M_{2,d} \leq u_n|X_1 > u_n) = \ell_d(\mathbf{1}_d) - \ell_{d-1}(\mathbf{1}_{d-1})$. In fact, this follows from,

$$\lim_{n \to \infty} \mathbb{P}(M_{2,d} \le u_n(\tau) | X_1 > u_n(\tau)) = \lim_{n \to \infty} \frac{n}{\tau} \mathbb{P}(M_{2,d} \le u_n(\tau), X_1 > u_n(\tau))$$

$$= \lim_{n \to \infty} \frac{n}{\tau} \left(\mathbb{P}(M_{2,d} \le u_n(\tau)) - \mathbb{P}(M_{1,d} \le u_n(\tau)) \right)$$

$$= \lim_{n \to \infty} \frac{n}{\tau} \left(\mathbb{P}(M_{1,d} > u_n(\tau)) - \mathbb{P}(M_{2,d} > u_n(\tau)) \right)$$

$$= \ell_d(\mathbf{1}_d) - \ell_{d-1}(\mathbf{1}_{d-1}).$$

Under the assumption that ℓ_s exists,

$$\Delta(s) := \lim_{n \to \infty} \mathbb{P}(M_{2,s} \le u_n | X_1 > u_n) = \ell_s(\mathbf{1}_s) - \ell_{s-1}(\mathbf{1}_{s-1})$$

is well defined for any finite s. Now since $\Delta(s)$ is a non-increasing function in s, we have, for any $s_0 > d$,

$$\Delta(d) \ge \Delta(s_0) \ge \lim_{n \to \infty} \mathbb{P}(M_{2,r_n} \le u_n | X_1 > u_n).$$

On the other hand, $D^{(d)}(u_n)$ condition is equivalent to

$$\Delta(d) = \lim_{n \to \infty} \mathbb{P}(M_{2,r_n} \le u_n | X_1 > u_n). \tag{2.13}$$

Therefore,

$$\{d, d+1, \ldots, s_0\} \subseteq \{s : \Delta(s) = \theta\}.$$

Therefore, d^* exists and

$$d^* = \min\{s : \Delta(s) = \theta\}. \tag{2.14}$$

Remark 2.2 If $d^* = 1$, it implies that $D^{(1)}(u_n)$ condition holds and thus $\theta = 1$ by (2.12) with d = 1. And if $d^* \geq 2$, from the monotonicity of $\Delta(s)$ it follows that $\theta < 1$. Moreover, (2.13) and (2.12) motivate to check $D^{(d)}(u_n)$ condition and to obtain the value of θ by evaluating the function ℓ_s . In Section 3, we demonstrate the calculation for four models.

In view of Theorem 2.2, there are multiple choices of d (namely any $d^* \leq d \leq r_n$) that work for the estimation given in (2.3). In fact, $\hat{\theta}_n(r_n)$ coincides with the runs estimator, cf. Weissman and Novak [1998]. We shall derive a procedure to estimate d^* . Let $\delta(s) = \Delta(s) - \Delta(s+1)$. Then $\delta(s) = 0$, for $s \geq d^*$; and $\delta(d^*-1) > 0$. We estimate $\delta(s)$ by $\hat{\theta}_n(s) - \hat{\theta}_n(s+1)$. The following theorem states the asymptotic property of this estimator, based on which we shall derive an estimator for d^* .

Theorem 2.3 Suppose that all the conditions of Theorem 2.1 hold for d = s + 1. Assume that Condition (A5) holds for d = s and that the following limits exist and are finite. For any $x, y \in [1/2, 3/2]$,

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=s+1}^{r_n} \mathbb{P}\left(U_1 < \frac{kx}{n} < mU_{2,s}, U_i < \frac{ky}{n} < mU_{i+1,i+s-1}\right) = \Lambda_2(x,y) \in [0,\infty),$$

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P}\left(U_1 < \frac{k}{n}, U_{i+1} < \frac{k}{n} < mU_{i+2,i+s}\right) = \tilde{\lambda}_1 \in [0,\infty),$$

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=s+1}^{r_n} \mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,s}, U_i < \frac{k}{n} < mU_{i+1,i+s}\right) = \lambda_2 \in [0,\infty),$$

and

$$\lim_{n \to \infty} \frac{n}{k} \sum_{i=s+1}^{r_n} \mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,s}, U_i < \frac{k}{n}\right) = \lambda_3 \in [0, \infty).$$

Then,

$$\sqrt{k} \left(\hat{\theta}_n(s) - \hat{\theta}_n(s+1) - \delta(s) \right) \overset{d}{\to} N(0, \sigma_1^2),$$

where
$$\sigma_1^2 = \delta^2(s) \left(2\Lambda_1(1,1) - 1 \right) - 2\delta(s) \left(\tilde{\lambda}_1 - \lambda_1 + \lambda_3 - \frac{1}{2} \right) + 2\Lambda_2(1,1) - 2\lambda_2$$
.

The proof of Theorem 2.3 is provided in Section 5.

Corollary 2.4 Assume that the conditions of Theorem 2.3 hold and Condition (A2) holds for $d = d^*$. If $s \ge d^*$, then

$$\sqrt{k}\left(\hat{\theta}_n(s) - \hat{\theta}_n(s+1)\right) \stackrel{d}{\to} 0.$$
 (2.15)

And,

$$\sqrt{k} \left(\hat{\theta}_n(d^* - 1) - \hat{\theta}_n(d^*) - \delta(d^* - 1) \right) \stackrel{d}{\to} N(0, \sigma_*^2),$$

where $\sigma_*^2 > 0$.

Proof. For $s \ge d^*$, $\delta(s) = \theta - \theta = 0$. Moreover, Condition (A2) with $d = d^*$ implies that $\Lambda_2(1,1) = 0$ and $\lambda_2 = 0$. Thus, $\sigma^2 = 0$ for $s \ge d^* + 1$ and (2.15) follows.

Remark 2.3 Observe that the result in Corollary 2.4 implies that

$$\hat{\theta}_n(s) - \hat{\theta}_n(s+1) = \begin{cases} o_p\left(\frac{1}{\sqrt{k}}\right) & \text{if } s \ge d^*; \\ \delta(d^*-1) + O_p\left(\frac{1}{\sqrt{k}}\right) & \text{if } s = d^* - 1. \end{cases}$$

Based on this, one can develop inference procedures to estimate d^* or to test the hypothesis that $D^{(d)}(u_n)$ condition holds for some given d. In Süveges [2007], the author has proposed to check the $D^{(2)}(u_n)$ condition by looking at the empirical value of $\mathbb{P}(X_2 \leq u_n < M_{3,r_n}|X_1 > u_n)$. Ferreira and Ferreira [2018] has used the same idea for checking $D^{(d)}(u_n)$ for a $d \geq 2$. However, no formal consistency result has been established for such a procedure in both papers.

We propose the following estimator for d^* :

$$\hat{d}^*(k) = \min \left\{ h : \max_{h \le i \le d^u} \left(\hat{\theta}_n(i) - \hat{\theta}_n(i+1) \right) < \frac{1}{\sqrt{k}} \right\}, \tag{2.16}$$

where d^u is a pre-specified upper bound of the searching range. Provided that $d^u \ge d^*$ and the assumptions of Corollary 2.4 hold, we have that (cf. Proposition 6.1) as $n \to \infty$,

$$\mathbb{P}\left(\hat{d}^*(k) = d^*\right) \stackrel{p}{\to} 1.$$

Plugging $\hat{d}^* = \hat{d}^*(k)$ into (2.3) yields our final estimator of θ :

$$\hat{\theta}_n(\hat{d}^*) = \frac{1}{k} \sum_{i=1}^{n - \hat{d}^* + 1} \mathbb{1} \left\{ m U_{i+1, i + \hat{d}^* - 1} \ge U_{k, n} > U_i \right\}.$$
(2.17)

Due to the consistency of \hat{d}^* , we conclude that the estimator $\hat{\theta}_n(\hat{d}^*)$ retains the asymptotic normality as shown in Theorem 2.1 with $d = d^*$.

3 Examples and simulation study

To investigate the finite sample performance of our estimators, we generate data from six distributions, of which four satisfy $D^{(d)}(u_n)$ condition for some d. In the sequel, we drop the subscript of the unit vector $\mathbf{1}$ when it is clear from the context: $\ell_s(\mathbf{1}) = \ell_s(\mathbf{1}_s)$. In this section, u_n is chosen such that $n\mathbb{P}(X > u_n) = 1$.

• Moving Maxima Let $X_i = \max_{0 \le j \le m} \epsilon_{j+i}$, for $i \ge 1$, where $m \ge 2$ is a fixed constant and ϵ_i an i.i.d sequence with $\mathbb{P}(\epsilon_i \le x) = \exp\left(-\frac{1}{mx}\right)$. Then (X_i) is a stationary sequence with marginal distribution $F(x) = \exp\left(-\frac{1}{x}\right)$ and for $s \ge 2$,

$$\ell_s(\mathbf{1}) = \lim_{n \to \infty} n \mathbb{P}\left(\max_{1 \le i \le s} X_i > u_n\right) = \lim_{n \to \infty} n \left(1 - F^{\frac{s+m-1}{m}}(u_n)\right) = \frac{s+m-1}{m}.$$

Thus (2.13) and therefore also $D^{(d)}(u_n)$ are satisfied for any $d \geq 2$. For this model, $d^* = 2$ and $\theta = \ell_2(1) - 1 = \frac{1}{m}$

• AR-C This is an AR(1) model with Cauchy margin. For $z \in (-1,1)$, define

$$X_i = zX_{i-1} + \epsilon_i, i \ge 1, \tag{3.1}$$

where X_0 has a standard Cauchy density $\frac{1}{\pi(1+x^2)}$ and ϵ_i i.i.d. with density $\frac{1-|z|}{\pi(x^2+(1-|z|)^2)}$. By Proposition 6.2, for $z \geq 0$, $d^* = 2$ and $\theta = 1 - z$; and for z < 0, $d^* = 3$ and $\theta = 1 - |z|^2$. • AR-N This is a classical AR(1) model with Gaussian margin. For $z \in (-1,1)$, define

$$X_i = zX_{i-1} + \epsilon_i, \quad i \ge 1,$$

where $X_0 \sim N(0, 1/(1-z^2))$ and ϵ_i i.i.d from N(0, 1). This is a model with $\theta = 1$ because $\ell_s(\mathbf{1}) = s$, for $s \geq 2$ and $d^* = 1$.

• Max Auto-Regressive models For $z \in [0,1)$, define $X_i = \max\{zX_{i-1}, \epsilon_i\}, i \geq 2$, where $X_1 = \frac{\epsilon_1}{1-z}$ and ϵ_i i.i.d with $\mathbb{P}(\epsilon_i \leq x) = \exp\left(-\frac{1}{x}\right)$. Then for $s \geq 2$,

$$\ell_s(\mathbf{1}) = \lim_{n \to \infty} n \left(1 - \mathbb{P} \left(\max_{1 \le i \le s} X_i \le u_n \right) \right)$$
$$= \lim_{n \to \infty} n \left(1 - \mathbb{P} \left(\epsilon_1 \le (1 - z)u_n, \max_{2 \le i \le s} \epsilon_i \le u_n \right) \right) = s - sz + z.$$

Thus, $d^* = 2$ and $\theta = 1 - z$.

Table 1 presents the parameters for each distribution that we have chosen for the simulation.

Table 1: Parameters of four distributions

	MM	AR-C	AR-N	MAX-AR
$m \text{ or } z$ d^* θ	$\begin{array}{c} 3 \\ 2 \\ \frac{1}{3} \end{array}$	$-\frac{1}{2}$ 3 $\frac{3}{4}$	$\begin{array}{c} \frac{1}{2} \\ 1 \\ 1 \end{array}$	$\begin{array}{c} \frac{1}{2} \\ 2 \\ \frac{1}{2} \end{array}$

Finally, we also consider two distributions, which do not satisfy $D^{(d)}(u_n)$ condition for any finite d. The proof for this is given in Proposition 6.3.

• sARCH This is a squared ARCH model given by

$$X_i = (2 \times 10^{-5} + \frac{1}{2} X_{i-1}) \epsilon_i^2, \quad i \ge 1,$$

where (ϵ_i) are i.i.d. from N(0,1). For this model, $\theta = 0.727$ (Table 3.1 in de Haan et al. [1989]).

• ARCH This is an ARCH model given by

$$X_i = (2 \times 10^{-5} + 0.7X_{i-1}^2)^{1/2} \epsilon_i, \quad i \ge 1,$$
 (3.2)

where (ϵ_i) are i.i.d. from N(0,1). For this model, $\theta = 0.721$ (Table 3.2 in de Haan et al. [1989]).

From each distribution, we generate m = 1000 samples with sample size n = 5000. We first look at the percentage of identifying the true d^* :

$$c(k) = \frac{\#\{1 \le i \le m, \ \hat{d}_i^*(k) = d^*\}}{m}.$$

Table 2: Percentage of correctly identified d^* for AR-N model

\underline{k}	30	40	50	60	70	80
c(k)	0.915	0.826	0.718	0.421	0.295	0.125

For the MM and ARMAX models, c(k) = 1 for $50 \le k \le 100$. For AR-C model, the estimate of d^* is either 3 or 1. Mostly, the correct d^* is identified: c(k) increases from 96.5% to 1 as k increases from 50 to 75 and c(k) = 1 for $75 \le k \le 100$. For AR-N model, smaller k leads to higher accuracy. Table 2 present c(k) for difference choices of k; note that the most miss-specified d^* is 2; $\hat{d}^* = 3$ only occurs a few times.

For the estimation of θ , we compare our estimator defined in (2.17) with $d^u = 10$ to two other methods.

- The pseudo maximum likelihood estimators based on the sliding blocks, $\hat{\theta}_n^{B,sl}$ defined on page 7 in Berghaus and Bücher [2018]. In this simulation, we take n/k as the block length r_n for this estimator.
- The interval estimator from Ferro and Segers [2003], which is given by $\hat{\theta}_n(u)$ on page 548 of that paper. We denote this estimator by $\hat{\theta}^{int}$.

The simulated MSE defined by $MSE(k) = \frac{1}{m} \sum_{i=1}^{m} (\hat{\theta}_i(k) - \theta)^2$, where $\hat{\theta}_i$ is the estimate based on the *i*-th generated sample by one of the three methods, is plotted in Figure 1. For the four distributions that satisfy $D^{(d)}(u_n)$ condition, our estimator outperforms the other two methods because it has the smallest MSE among the three for a wide range of k and it has the smallest minimum MSE taken over $k \in [30, 300]$. Even for the ARCH model, the minimum MSE of our estimator is also smaller than that of the other two methods. For the squared ARCH model, our estimator is better than the interval method but slightly worse than the sliding blocks method in terms of MSE.

For ARCH and squared ARCH models, $\theta = \lim_{n\to\infty} \mathbb{P}(M_{2,r_n} \leq u_n|X_1 > u_n)$. So θ can be well estimated by the runs estimator $\hat{\theta}(r_n)$, with $\hat{\theta}(\cdot)$ defined in (2.3). Our procedure leads to an estimator $\hat{\theta}(\hat{d}^*)$ such that the difference between $\hat{\theta}(\hat{d}^*)$ and $\hat{\theta}(d^u)$ is very small (cf. (2.16)). Therefore, when d^u is large enough our estimation of d^* can be viewed as a selection procedure for block lengths for the runs estimator.

4 An application on heatwaves

We investigate the clustering of high temperature in summer by estimating the extremal index for data measured at two weather stations: de Bilt (N 52.11°, E 5.18°) in the Netherlands and Larissa (N 39.64°, E 22.41°) in Greece. We consider the daily maximum temperatures in June, July and August from 1955 to 2018 (Chamberlain [2019]). The sample size is 5888 for de Bilt and 5796 for Larissa. The difference in sample sizes is due to the missing data of year 2005 for the station Larrisa. In terms of the measurement time, there is a natural gap between the data of two consecutive years, which are considered independent. To account for this feature, our estimator is adjusted as following:

$$\hat{\theta}_n(d) = \frac{1}{k} \sum_{i=0}^{N-1} \sum_{j=1}^{L-d+1} \mathbb{1} \left\{ U_{iL+j} < U_{k,n} \le m U_{iL+j+1,iL+j+d-1} \right\}, \tag{4.1}$$

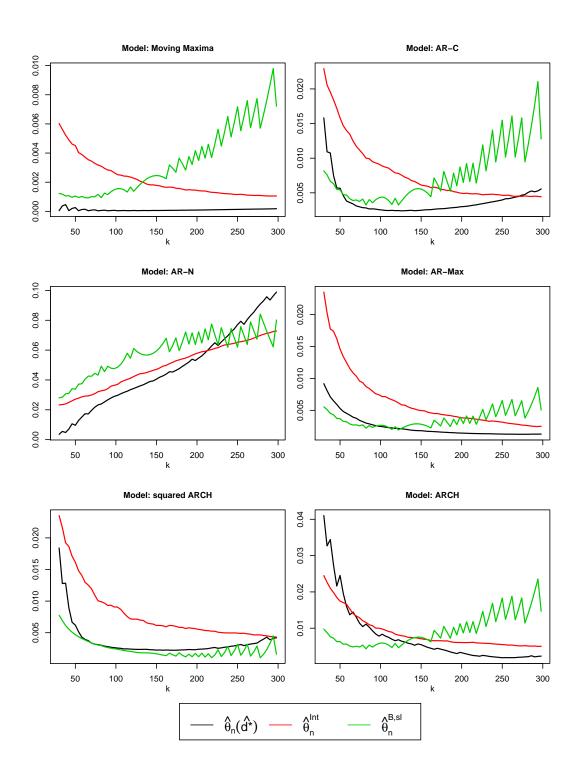


Figure 1: The simulated MSE of three estimators for six distributions, where $\hat{\theta}(\hat{d}^*)$ is defined in (2.17), $\hat{\theta}_n^{B,sl}$ defined in Berghaus and Bücher [2018] and $\hat{\theta}^{\text{int}}$ defined in Ferro and Segers [2003].

Estimation of d*: de Bilt Estimation of d*: Larissa h=6 0.4 0.3 $\frac{\hat{\theta}(h) - \hat{\theta}(h+1)}{0.2}$ $\hat{\theta}(h) - \hat{\theta}(h+1)$ 0.1 0.1 0.0 0.0 50 100 150 200 50 100 150 200

Figure 2: The curves represent $\hat{\theta}_n(h) - \hat{\theta}_n(h+1)$ (solid lines) and the threshold $1/\sqrt{k}$ (the dash lines), based on the summer temperature data observed in de Bilt from 1955 to 2018 and in Larissa from 1955 to 2018 except 2005.

where N denotes the number of years, and L denotes the number of observations for each year, 92 in this case.

First, we check the $D^{(d)}(u_n)$ condition by obtaining the estimates of $\delta(h)$ and compare those to the threshold $1/\sqrt{k}$. Based on the results in Figure 2, we conclude that $\hat{d}^* = 2$ for the data from de Bilt and $\hat{d}^* = 5$ for the data from Larrisa. The estimates of θ are plotted in Figure 3 by plugging d = 2 into (4.1) for de Bilt and d = 5 for Larrisa. Both results are quite stable for $k \in [50, 200]$. Note that the threshold $X_{n-k,n}$ in Larrisa is much higher than that in de Bilt. We present the statistics and estimates in Table 3, where the fourth and seventh columns are the estimates of the expected number of days during a heatwave. For instance, the results for k = 200 are interpreted as when a heatwave occurs, the expected number of very hot days (above 38.6 °C) in the area of Larrisa in Greece is 7.7 while the expected number of warm days in the area of de Bilt is only 1.7.

Table 3: Estimates of extremal indices of daily maximum temperatures in De Bilt and in Larrisa

	De Bilt			Larrisa		
\underline{k}	$X_{n-k,n}$	$\hat{\theta}(2)$ expected duration		$X_{n-k,n}$	$\hat{\theta}(5)$	expected duration
50	32.2 °C	0.76	$1.3 \mathrm{\ days}$	41.4°C	0.14	$7.1 \mathrm{days}$
100	31.3 °C	0.65	$1.5 \mathrm{days}$	40.0 °C	0.12	$8.3 \mathrm{days}$
200	30.0 °C	0.60	$1.7 \mathrm{\ days}$	$38.6^{\circ}\mathrm{C}$	0.13	$7.7 \mathrm{days}$

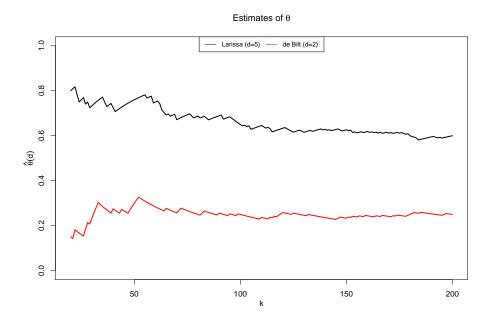


Figure 3: Estimates of θ based on the summer temperature data observed in de Bilt from 1955 to 2018 and in Larissa from 1955 to 2018 except 2005.

5 Proof

To prove Theorems 2.1 and 2.3, we first show two propositions. Define

$$\tilde{\theta}_n(x,d) = \frac{1}{k} \sum_{i=1}^{n-d+1} \mathbb{1} \left\{ U_i < \frac{kx}{n} < mU_{i+1,i+d-1} \right\}, \tag{5.1}$$

where $U_i = 1 - F(X_i)$ and $mU_{i+1,i+d-1} = \min_{i+1 \le j \le i+d-1} U_j$. Note that $\tilde{\theta}_n$ is a pseudo estimator because U_i 's are not observable since F is unknown. By the stationarity of U_i 's,

$$\mathbb{E}\left(\tilde{\theta}_n(x,d)\right) = \frac{n-d+1}{k} \mathbb{P}\left(U_1 < \frac{kx}{n} < mU_{2,d}\right) \to \ell_d(x\mathbf{1}) - \ell_{d-1}(x\mathbf{1}) = x\theta,$$

by Condition (A5), the homogeneity of ℓ and (2.12). And we also have,

$$\hat{\theta}_n(d) = \frac{1}{k} \sum_{i=1}^{n-d+1} \mathbb{1} \left\{ U_i < U_{k,n} \le m U_{i+1,i+d-1} \right\} = \tilde{\theta}_n \left(\frac{n}{k} U_{k,n}, d \right).$$

Since $\frac{n}{k}U_{k,n} \stackrel{p}{\to} 1$, we will first obtain the asymptotic properties for $\tilde{\theta}_n(x,d)$ for $x \in [1/2,3/2]$. Precisely, we shall prove the asymptotic normality of

$$\sqrt{k}\left(\tilde{\theta}_n(x,d)-\theta_n(x,d)\right)=:\nu_n(x,d), \quad x\in[1/2,3/2],$$

where $\theta_n(x,d) = \frac{n}{k} \mathbb{P}\left(U_1 < \frac{kx}{n} < mU_{2,d}\right)$.

Proposition 5.1 Under the conditions of Theorem 2.1,

$$\nu_n(x,d) \stackrel{d}{\to} W_d(x),$$

in D([1/2,3/2]), where W_d is a mean-zero Gaussian process with covariance structure

$$\mathbb{E}(W_d(x)W_d(y)) = \lim_{n \to \infty} \sum_{i=0}^{d-1} \frac{n}{k} \mathbb{P}\left(U_1 < \frac{kx}{n} < mU_{2,d}, U_{1+i} < \frac{ky}{n} < mU_{2+i,d+i}\right),$$

for $x \leq y$. In particular, $\mathbb{E}(W_d^2(x)) = x\theta$.

Proof of Proposition 5.1

When there is no confusion, we drop the subscript d from the notation of $W_d(x)$ and denote $\nu_n(x,d) =: \nu_n(x)$. We prove convergence of finite-dimensional distributions plus tightness. For the convergence of finite-dimensional distribution, it is sufficient to prove for each $m \in \mathbb{N}$ and for any $x_i \in [1/2, 3/2]$ and $a_i \in \mathbf{R}$, $i = 1, \ldots, m$,

$$\sum_{i=1}^{m} a_i \nu_n(x_i) \stackrel{d}{\to} \sum_{i=1}^{m} a_i W(x_i).$$

We show the proof for the case of m=2. For other cases, the proof is more tedious but can be done in the same way. Let $1/2 \le x \le y \le 3/2$, $I_i=\mathbbm{1}\left\{U_i < \frac{kx}{n} < mU_{i+1,i+d-1}\right\}$ and $J_i=\mathbbm{1}\left\{U_i < \frac{ky}{n} < mU_{i+1,i+d-1}\right\}$, $i=1,\ldots,n-d+1$. Then

$$a_1 \nu_n(x) + a_2 \nu_n(y) = \frac{1}{\sqrt{k}} \sum_{i=1}^{n-d+1} \left(a_1 I_i + a_2 J_i - \mathbb{E}(a_1 I_i + a_2 J_i) \right) + O\left(\frac{1}{\sqrt{k}}\right)$$
$$=: \sum_{i=1}^{n-d+1} \xi_{i,n} + O\left(\frac{1}{\sqrt{k}}\right).$$

We apply the main theorem in Utev [1990] to prove that $\sum_{i=1}^{n-d+1} \xi_{i,n} \stackrel{d}{\to} a_1 W(x) + a_2 W(y)$. We begin with computing the variance.

$$\mathbf{Var}\left(\sum_{i=1}^{n-d+1} \xi_{i,n}\right) = \frac{a_1^2}{k} \mathbf{Var}\left(\sum_{i=1}^{n-d+1} I_i\right) + \frac{a_2^2}{k} \mathbf{Var}\left(\sum_{i=1}^{n-d+1} J_i\right) + \frac{2a_1a_2}{k} \mathbf{Cov}\left(\sum_{i=1}^{n-d+1} I_i, \sum_{i=1}^{n-d+1} J_i\right) + \frac{a_2^2}{k} \mathbf{Var}\left(\sum_{i=1}^{n-d+1} I_i, \sum_{i=1}^{n-d+1} I_i\right) + \frac{a_2^2}{k} \mathbf{Var}\left(\sum_{i=1}^{n-d+1} I_i, \sum_{i=1}^{n-d+1} I_i\right) + \frac{a_2^2}{k} \mathbf{Var}\left(\sum_{i=1}^{n-d+1} I_i\right) + \frac{a_2^2$$

The first two terms can be dealt with in the same way. Note that

$$\mathbb{E}(I_1) = \mathbb{P}\left(U_i < \frac{kx}{n} < mU_{i+1,i+d-1}\right) = O\left(\frac{k}{n}\right),\,$$

and $\mathbf{Var}(I_1) = \mathbb{E}(I_1) \left(1 - O\left(\frac{k}{n}\right)\right)$. The same results hold for J_i 's. Thus, by stationarity,

$$\mathbf{Var}\left(\sum_{i=1}^{n-d+1} I_i\right) = (n-d+1)\mathbf{Var}(I_1) + 2\sum_{i < j} \mathbf{Cov}(I_i, I_j)$$
$$= (n-d+1)\mathbb{E}(I_1)\left(1 - O\left(\frac{k}{n}\right)\right) + 2\sum_{i=1}^{n-d} (n-d+1-i)\mathbf{Cov}(I_1, I_{1+i})$$

$$= (n - d + 1)\mathbb{E}(I_1)\left(1 - O\left(\frac{k}{n}\right)\right) + o(k),$$

by Lemmas 6.1 and 6.2. Thus,

$$\frac{a_1^2}{k} \mathbf{Var} \left(\sum_{i=1}^{n-d+1} I_i \right) = a_1^2 \frac{n}{k} \mathbb{E}(I_1) + o(1).$$

Similarly, one obtains that

$$\frac{a_2^2}{k} \mathbf{Var} \left(\sum_{i=1}^{n-d+1} J_i \right) = a_2^2 \frac{n}{k} \mathbb{E}(J_1) + o(1).$$

As for the covariance term, we have that, again by stationarity and Lemmas 6.1 and 6.2,

$$\mathbf{Cov}\left(\sum_{i=1}^{n-d+1} I_{i}, \sum_{i=1}^{n-d+1} J_{i}\right) \\
= \sum_{i=0}^{n-d+1} (n-d+1-i)\mathbf{Cov}\left(I_{1}, J_{1+i}\right) + \sum_{i=1}^{n-d+1} (n-d+1-i)\mathbf{Cov}\left(I_{1+i}, J_{1}\right) \\
= \sum_{i=0}^{d-1} (n-d+1-i)\mathbf{Cov}\left(I_{1}, J_{1+i}\right) + \sum_{i=1}^{d-1} (n-d+1-i)\mathbf{Cov}\left(I_{1+i}, J_{1}\right) + o(k) \\
= \sum_{i=0}^{d-1} (n-d+1-i)\mathbb{E}\left(I_{1}J_{1+i}\right) + \sum_{i=1}^{d-1} (n-d+1-i)E\left(I_{1+i}J_{1}\right) + O\left(\frac{k^{2}}{n}\right) + o(k),$$

The last equality follows from that $\mathbb{E}(I_1)E(J_1)=O\left(\frac{n^2}{k^2}\right)$. Morever, because of the disjointness causing by the fact that $x \leq y$, $\mathbb{E}(I_{i+1}J_1)=0$, for $i=1,\ldots,d-1$. Therefore,

$$\frac{2a_1a_2}{k}\mathbf{Cov}\left(\sum_{i=1}^{n-d+1}I_i,\sum_{i=1}^{n-d+1}J_i\right) = 2a_1a_2\frac{n}{k}\sum_{i=0}^{d-1}\mathbb{E}\left(I_1J_{1+i}\right) + o(1).$$

We have shown that

$$\lim_{n \to \infty} \mathbf{Var} \left(\sum_{i=1}^{n-d+1} \xi_{i,n} \right) = \lim_{n \to \infty} \left(a_1^2 \frac{n}{k} \mathbb{E}(I_1) + a_2^2 \frac{n}{k} \mathbb{E}(J_1) + 2a_1 a_2 \frac{n}{k} \sum_{i=0}^{d-1} \mathbb{E}\left(I_1, J_{1+i}\right) \right)$$

$$= a_1^2 \mathbb{E}(W^2(x)) + a_2^2 \mathbb{E}(W^2(y)) + 2a_1 a_2 \mathbb{E}(W(x)W(y)).$$

Next, we check the Condition (2) in Utev [1990]. Denote $\sigma_n^2 = \mathbf{Var}(\sum_{i=1}^{n-d+1} \xi_{i,n})$. Choosing $j_1 = j_2 = \cdots = 1$, we have, for any $\epsilon > 0$,

$$\sigma_{n}^{-2} \sum_{i=1}^{n-a+1} \mathbb{E}(\xi_{i,n}^{2} \mathbb{1}(|\xi_{i,n}| \geq \epsilon \sigma_{n})) \leq \sigma_{n}^{-3} n \epsilon^{-1} \mathbb{E}(|\xi_{i}|^{3})$$

$$= \sigma_{n}^{-3} n k^{-3/2} \epsilon^{-1} \mathbb{E}\left(|a_{1}I_{1} + a_{2}J_{1} - \mathbb{E}(a_{1}I_{1} + a_{2}J_{1})|^{3}\right)$$

$$= \sigma_{n}^{-3} n k^{-3/2} \epsilon^{-1} \mathbb{E}\left(|a_{1}I_{1} + a_{2}J_{1} - O(\frac{k}{n})|^{3}\right)$$

$$=c\sigma_n^{-3}nk^{-3/2}\epsilon^{-1}O\left(\frac{k}{n}\right)\to 0.$$
 (5.2)

Thus, by the main theorem in Utev [1990], the central limit theorem holds for $\{\xi_{i,n}, i = 1, \ldots, n-d+1\}$.

Now we prove the tightness of ν_n . Note that for any $x \in [1/2, 3/2]$, $\mathbbm{1}\left\{U_i < \frac{kx}{n} < mU_{i+1,i+d-1}\right\} = \mathbbm{1}\left\{mU_{i,i+d-1} < \frac{kx}{n}\right\} - \mathbbm{1}\left\{mU_{i+1,i+d-1} < \frac{kx}{n}\right\}$. Thus,

$$\nu_n(x) = \nu_{1,n}(x) - \nu_{2,n}(x),$$

where $\nu_{1,n}(x) = \frac{1}{\sqrt{k}} \sum_{i=1}^{n-d+1} \left(\mathbbm{1}\left\{ mU_{i,i+d-1} < \frac{kx}{n} \right\} - \mathbb{P}\left(mU_{i,i+d-1} < \frac{kx}{n} \right) \right)$, and $\nu_{2,n}(x) = \frac{1}{\sqrt{k}} \sum_{i=1}^{n-d+1} \left(\mathbbm{1}\left\{ mU_{i+1,i+d-1} < \frac{kx}{n} \right\} - \mathbb{P}\left(mU_{i+1,i+d-1} < \frac{kx}{n} \right) \right)$. To prove the tightness of $\nu_n(x)$, it is sufficient to prove the tightness of $\nu_{1,n}(x)$ and $\nu_{2,n}(x)$. We demonstrate the proof for $\nu_{1,n}(x)$. Let $t_n = \lfloor \frac{n-d+1}{2r_n} \rfloor$. We split the sum into $2t_n$ blocks of length r_n and a remaining block of length less than $2r_n$. To simplify the notation, we denote $M_i = mU_{i,i+d-1}$. Precisely,

$$\begin{split} \nu_{1,n}(x) &= \frac{1}{\sqrt{k}} \sum_{i=1}^{n-d+1} \left(\mathbbm{1} \left\{ M_i < \frac{kx}{n} \right\} - \mathbb{P} \left(M_i < \frac{kx}{n} \right) \right) \\ &= \frac{1}{\sqrt{k}} \sum_{i=0}^{t_n - 1} \sum_{j=1}^{r_n} \left(\mathbbm{1} \left\{ M_{2ir_n + j} < \frac{kx}{n} \right\} - \mathbb{P} \left(M_{2ir_n + j} < \frac{kx}{n} \right) \right) \\ &+ \frac{1}{\sqrt{k}} \sum_{i=0}^{t_n - 1} \sum_{j=1}^{r_n} \left(\mathbbm{1} \left\{ M_{(2i+1)r_n + j} < \frac{kx}{n} \right\} - \mathbb{P} \left(M_{(2i+1)r_n + j} < \frac{kx}{n} \right) \right) \\ &+ \frac{1}{\sqrt{k}} \sum_{i=2t_n r_n + 1}^{n - d + 1} \left(\mathbbm{1} \left\{ M_i < \frac{kx}{n} \right\} - \mathbb{P} \left(M_i < \frac{kx}{n} \right) \right) \\ &= : \mu_{1,n}(x) + \mu_{2,n}(x) + \mu_{3,n}(x). \end{split}$$

Define $\tilde{\mu}_n(x) = \frac{1}{\sqrt{k}} \sum_{i=1}^{t_n r_n} \left(\mathbb{1} \left\{ \tilde{M}_i < \frac{kx}{n} \right\} - \mathbb{P} \left(\tilde{M}_i < \frac{kx}{n} \right) \right)$, where for any $i = 1, \dots, t_n$, $\left\{ \tilde{M}_{r_n(i-1)+i}, \dots, \tilde{M}_{ir_n} \right\} \stackrel{d}{=} \left\{ M_1, \dots, M_{r_n} \right\},$

and $\{\tilde{M}_{r_n(i-1)+j}, j=1,\ldots,r_n\}_{i=1}^{t_n}$ are t_n independent blocks. So \tilde{M}_i 's form a special r_n -dependent arrays for each n, which is not a strictly stationary sequence. We first apply a fluctuation inequality for m-dependent arrays given by Theorem 4.1 in Einmahl and Ruymgaart [2000] to prove the tightness of $\tilde{\mu}_n(x)$. Then, the tightness of $\mu_{1,n}$ and $\mu_{2,n}$ follows from the bounded variation distance between $\tilde{\mu}_n$ and $\mu_{1,n}$ and between $\tilde{\mu}_n$ and $\mu_{2,n}$, respectively.

For each n, let $q = \lceil r_n^{1+\epsilon} \rceil$, where ϵ is some positive number such that $q/\sqrt{k_n} \to 0$. Define $I_i = [\frac{1}{2} + \frac{i}{q}, \frac{1}{2} + \frac{i+1}{q}], \quad i = 0, \ldots, q-1$. Choose $\delta_n = \frac{1}{q}$. For any $x, y \in [1/2, 3/2]$ and $|x-y| < \delta_n$, there exists an $i \in \{1, \ldots, q-1\}$ such that

$$\left|x - \frac{1}{2} - \frac{i}{q}\right| < \frac{1}{q}$$
 and $\left|y - \frac{1}{2} - \frac{i}{q}\right| < \frac{1}{q}$.

Thus, for any $\lambda > 0$,

$$\begin{split} & \mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n\\1/2\leq x< y\leq 2}} |\tilde{\mu}_n(x) - \tilde{\mu}_n(y)| \geq \lambda\right) \\ \leq & \mathbb{P}\left(\max_{\substack{1\leq i\leq q-1\\|y-\frac{1}{2}-\frac{i}{q}|<\frac{1}{q}\\|y-\frac{1}{2}-\frac{i}{q}|<\frac{1}{q}}} \left(\left|\tilde{\mu}_n(x) - \tilde{\mu}_n(\frac{1}{2} + \frac{i}{q})\right| + \left|\tilde{\mu}_n(\frac{1}{2} + \frac{i}{q}) - \tilde{\mu}_n(y)\right|\right) \geq \lambda\right) \\ \leq & 2\mathbb{P}\left(\max_{1\leq i\leq q-1} \sup_{x\in I_i} \left|\tilde{\mu}_n(x) - \tilde{\mu}_n(\frac{1}{2} + \frac{i}{q})\right| \geq \frac{\lambda}{2}\right) + 2\mathbb{P}\left(\max_{1\leq i\leq q-1} \sup_{x\in I_i} \left|\tilde{\mu}_n(x) - \tilde{\mu}_n(\frac{1}{2} + \frac{i+1}{q})\right| \geq \frac{\lambda}{2}\right) \\ \leq & 2\sum_{i=1}^{q-1} P\left(\sup_{x\in I_i} \left|\tilde{\mu}_n(x) - \tilde{\mu}_n(\frac{1}{2} + \frac{i}{q})\right| \geq \frac{\lambda}{2}\right) + 2\sum_{i=1}^{m-1} P\left(\sup_{x\in I_i} \left|\tilde{\mu}_n(x) - \tilde{\mu}_n(\frac{1}{2} + \frac{i+1}{q})\right| \geq \frac{\lambda}{2}\right) \\ \leq & 4\sum_{i=1}^{m-1} P\left(\sup_{x,y\in I_i} |\tilde{\mu}_n(x) - \tilde{\mu}_n(y)| \geq \frac{\lambda}{2}\right) =: 4\sum_{i=1}^{q-1} P_i. \end{split}$$

Next, we apply (4.4) in Einmahl and Ruymgaart [2000] to bound P_i . To use the notation in that paper, we define

$$\Delta_n(x) = \frac{1}{t_n r_n} \sum_{i=1}^{t_n r_n} \mathbb{1} \left\{ \tilde{M}_i < x \right\} - \mathbb{P}(\tilde{M}_i < x).$$

Then, by taking, in the notation of that paper, $\epsilon = \frac{1}{2}$ and $m = r_n$, we have

$$P_{i} = \mathbb{P}\left(\sup_{\frac{1}{2} + \frac{i}{q} \le x < y \le \frac{1}{2} + \frac{i+1}{q}} |\tilde{\mu}_{n}(x) - \tilde{\mu}_{n}(y)| \ge \lambda\right)$$

$$= \mathbb{P}\left(\frac{r_{n}t_{n}}{\sqrt{k}} \sup_{\frac{1}{2} + \frac{i}{q} \le x < y \le \frac{1}{2} + \frac{i+1}{q}} |\Delta_{n}\left(\frac{kx}{n}\right) - \Delta_{n}\left(\frac{ky}{n}\right)| \ge \lambda\right)$$

$$= \mathbb{P}\left(\sup_{\frac{k}{n}(\frac{1}{2} + \frac{i}{q}) \le a < b \le \frac{k}{n}(\frac{1}{2} + \frac{i+1}{q})} |\Delta_{n}(a) - \Delta_{n}(b)| \ge \frac{\sqrt{k}\lambda}{r_{n}t_{n}}\right)$$

$$\le c_{1} \exp\left(\frac{-r_{n}t_{n}\frac{k\lambda^{2}}{r_{n}^{2}t_{n}^{2}}}{4r_{n}p_{i}}\psi\left(\frac{\sqrt{r_{n}t_{n}}\frac{\sqrt{k}\lambda}{r_{n}t_{n}}}{\sqrt{r_{n}t_{n}}p_{i}}\right)\right) = c_{1} \exp\left(-\frac{k\lambda^{2}}{4r_{n}^{2}t_{n}p_{i}}\psi\left(\frac{\sqrt{k}\lambda}{r_{n}t_{n}p_{i}}\right)\right)$$

where $p_i = \mathbb{P}\left(\frac{k}{n}(\frac{1}{2} + \frac{i}{q}) < mU_{1,d} \leq \frac{k}{n}(\frac{1}{2} + \frac{i+1}{q})\right)$ and ψ is a continuous and decreasing function such that $\psi(0) = 1$. Observe that

$$\sup_{0 \le i \le q-1} \left| \frac{n}{k} p_i - \left(\ell_d \left(\mathbf{1} \left(\frac{1}{2} + \frac{i+1}{q} \right) \right) - \ell_d \left(\mathbf{1} \left(\frac{1}{2} + \frac{i}{q} \right) \right) \right) \right| = o\left(\left(\frac{n}{k} \right)^{\tau} \right),$$

and $\ell_d\left(\mathbf{1}\left(\frac{1}{2}+\frac{i+1}{q}\right)\right)-\ell_d\left(\mathbf{1}\left(\frac{1}{2}+\frac{i}{q}\right)\right)=\frac{1}{q}\ell_d(\mathbf{1})$ by the homogeneity of ℓ_d . Thus, for n large enough, $\frac{k}{2nq}\leq p_i\leq \frac{2dk}{nq}$, uniformly in i, due to the fact that $\ell_d(\mathbf{1})\in[1,d]$.

Then by the choice of q and that $n - r_n \leq 2r_n t_n \leq n$,

$$\frac{k}{4r_n^2t_np_i} \geq \frac{k}{2r_nnp_i} \geq \frac{m}{4dr_n} = \frac{1}{4d}r_n^\epsilon,$$

and

$$\psi\left(\frac{\sqrt{k}\lambda}{r_nt_np_i}\right) \ge \psi\left(\frac{3\sqrt{k}\lambda}{n\frac{k}{2nq}}\right) = \psi\left(\frac{6\lambda q}{\sqrt{k}}\right) \to \psi(0) = 1.$$

Thus,

$$\mathbb{P}\left(\sup_{\substack{|x-y|<\delta_n\\1/2\leq x< y\leq 2}} |\tilde{\mu}_n(x) - \tilde{\mu}_n(y)| \geq \lambda\right) \leq 4q \exp(-c_2 r_n^{\epsilon}) = 4r_n^{1+\epsilon} \exp(-c_2 r_n^{\epsilon}) \to 0.$$

So the tightness of $\tilde{\mu}_n$ follows from the tightness criterion by Theorem 1 in Aldous [1978]. By Lemma 2 in Eberlein [1984],

$$\| \Omega \left(\{ \tilde{M}_{r_n(i-1)+1}, \dots, \tilde{M}_{ir_n} \}_{i=1}^{t_n} \right) - \Omega \left(\{ M_{r_n(i-1)+1}, \dots, M_{ir_n} \}_{i=1}^{t_n} \right) \|$$

$$= \| \bigotimes_{i=1}^{t_n} \Omega \left(M_{r_n(i-1)+1}, \dots, M_{ir_n} \right) - \Omega \left(\{ M_{r_n(i-1)+1}, \dots, M_{ir_n} \}_{i=1}^{t_n} \right) \|$$

$$\leq \beta(r_n) t_n \leq \beta(r_n) \frac{n}{r_n} \to 0,$$

by the absolutely regular assumption on the sequence, and the condition that $\beta(r_n)\frac{n}{r_n}\to 0$, where $\Omega(X)$ denotes the distribution of X. Thus, the tightness of $\mu_{1,n}$ and $\mu_{2,n}$ follow from the tightness of $\tilde{\mu}_n$. To prove the tightness of $\nu_{1,n}$, it is remaining to show that $\sup_{1/2 \le x \le 3/2} |\mu_{3,n}(x)| \stackrel{P}{\to} 0$. Note that by the definition of t_n , the number of summands in $\mu_{3,n}$ is bounded by $2r_n$.

$$\mathbb{E}\left(\sup_{1/2 \le x \le 3/2} |\mu_{3,n}(x)\right) \le \mathbb{E}\left(\frac{1}{\sqrt{k}} \sum_{i=2t_n r_n+1}^{n-d+1} \left(\mathbb{1}\left\{M_i < \frac{3k}{2n}\right\} + \mathbb{P}\left(M_i < \frac{3k}{2n}\right)\right)\right)$$

$$\le \frac{4r_n}{\sqrt{k}} \cdot \frac{3k}{2n} \to 0,$$

by the assumption that $\frac{r_n\sqrt{k}}{n} \to 0$.

Proposition 5.2 Under the conditions of Theorem 2.1,

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{n}\mathbb{1}\left(U_{i}<\frac{kx}{n}\right)-x\right)\overset{d}{\to}\tilde{W}(x),$$

in D([1/2,3/2]), where \tilde{W} is a mean-zero Gaussian process with covariance structure $\mathbb{E}(\tilde{W}(x),\tilde{W}(y)) = \min(x,y) + \Lambda_1(x,y) + \Lambda_1(y,x)$.

Proposition 5.2 can be proved in a similar but simpler way as Proposition 5.1.

This result implies that, by Theorem A.0.1 and Lemma A.0.2 in de Haan and Ferreira [2006],

$$\sqrt{k} \left(\frac{n}{k} U_{k,n} - 1 \right) \stackrel{d}{\to} -\tilde{W}(1). \tag{5.3}$$

In particular, one has $\frac{n}{k}U_{k,n} \stackrel{p}{\to} 1$.

Proof for Theorem 2.1

For convenient presentation, all the processes involved in the proof are defined on the same probability space, via the Skorohod construction. We use the same notation, though they are only equal in distribution to the original ones. We start with the following decomposition: by the definition of $\tilde{\theta}_n$,

$$\sqrt{k} \left(\hat{\theta}(d) - \theta \right) = \sqrt{k} \left(\tilde{\theta}_n \left(\frac{n}{k} U_{k,n}, d \right) - \theta \right)
= \sqrt{k} \left(\tilde{\theta}_n \left(\frac{n}{k} U_{k,n}, d \right) - \theta_n \left(\frac{n}{k} U_{k,n}, d \right) \right) + \sqrt{k} \left(\theta_n \left(\frac{n}{k} U_{k,n}, d \right) - \theta \right)
=: I_1 + I_2.$$

Now $I_1 \stackrel{p}{\to} W(1)$ follows from the fact that

$$|I_{1} - W_{d}(1)| \le \left| \sqrt{k} \left(\tilde{\theta}_{n} \left(\frac{n}{k} U_{k,n}, d \right) - \theta_{n} \left(\frac{n}{k} U_{k,n}, d \right) \right) - W_{d} \left(\frac{n}{k} U_{k,n} \right) \right| + \left| W_{d} \left(\frac{n}{k} U_{k,n} \right) - W_{d}(1) \right| \xrightarrow{p} 0,$$

$$(5.4)$$

by Proposition 5.1 and that $\frac{n}{k}U_{k,n} \stackrel{p}{\to} 1$ by Proposition 5.2. Next, we deal with I_2 . Define $\ell_{d,n}(x\mathbf{1}_d) = \frac{n}{k}\mathbb{P}\left(mU_{1,d} \leq \frac{kx}{n}\right)$. Then $\lim_{n\to}\ell_{d,n}(x\mathbf{1}_d) = \ell_d(x\mathbf{1}_d)$ and $\theta_n(x,d) = \ell_{d,n}(x\mathbf{1}_d) - \ell_{d-1,n}(x\mathbf{1}_{d-1})$. Therefore,

$$I_{2} = \sqrt{k} \left(\ell_{d,n} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d} \right) - \ell_{d-1,n} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d-1} \right) - \theta \right)$$

$$= \sqrt{k} \left(\ell_{d,n} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d} \right) - \ell_{d} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d} \right) - \left(\ell_{d-1,n} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d-1} \right) - \ell_{d-1} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d-1} \right) \right) \right)$$

$$+ \sqrt{k} \left(\ell_{d} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d} \right) - \left(\ell_{d-1} \left(\frac{n}{k} U_{k,n} \mathbf{1}_{d-1} \right) - \theta \right) \right)$$

$$= L_{11} + L_{12}$$

By Assumption (A5), $k = o\left(n^{2\rho/(2\rho-1)}\right)$ and $\frac{n}{k}U_{k,n} \stackrel{p}{\to} 1$, $I_{21} \stackrel{p}{\to} 0$. By the homogeneity of ℓ_d and ℓ_{d-1} , and (5.3)

$$I_{22} = \sqrt{k} \left(\frac{n}{k} U_{k,n} - 1 \right) \left(\ell_d(\mathbf{1}_d) - \ell_{d-1}(\mathbf{1}_{d-1}) \right) = \theta \sqrt{k} \left(\frac{n}{k} U_{k,n} - 1 \right) \xrightarrow{p} -\theta \tilde{W}(1).$$
 (5.5)

Thus,

$$\sqrt{k}\left(\hat{\theta}(d) - \theta\right) \stackrel{p}{\to} W_d(1) - \theta \tilde{W}(1).$$
 (5.6)

It remains to prove that $W_d(1) - \theta \tilde{W}(1) = N(0, \sigma^2)$, where σ^2 is defined in Theorem 2.1. In view of

 $\sqrt{k}(\tilde{\theta}_n(1) - \theta_n(1,d)) \stackrel{p}{\to} W_d(1)$ and $\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^n\mathbb{1}\left(U_i < \frac{k}{n}\right) - 1\right) \stackrel{p}{\to} \tilde{W}(1)$, it suffices to show that

$$\sqrt{k} \left(\tilde{\theta}_n(1) - \frac{\theta}{k} \sum_{i=1}^n \mathbb{1} \left(U_i < \frac{k}{n} \right) - (\theta_n(1, d) - \theta) \right) \stackrel{p}{\to} N(0, \sigma^2). \tag{5.7}$$

Define $T_i = \mathbb{1}\left\{U_i < \frac{k}{n} < mU_{i+1,i+d-1}\right\} - \theta\mathbb{1}\left(U_i < \frac{k}{n}\right)$, then the left hand side of (5.7) becomes $\frac{1}{\sqrt{k}}\sum_{i=1}^{n-d+1}\left(T_i - \mathbb{E}(T_1)\right) + O_p(d/\sqrt{k})$. We shall apply the main Theorem in Utev [1990] to prove the central limit theorem for $\sum_{i=1}^{n-d+1}\xi_{i,n}$, where $\xi_{i,n} = \frac{1}{\sqrt{k}}\left(T_i - \mathbb{E}(T_1)\right)$. We begin with the variance:

$$\mathbf{Var}\left(\sum_{i=1}^{n-d+1} \xi_{i,n}\right) = \frac{n-d+1}{k} \mathbf{Var}(T_i) + \frac{2}{k} \sum_{i=1}^{n-d+1} (n-i) \mathbf{Cov}(T_1, T_{1+i})$$

First, $\frac{n}{k}\mathbb{E}(T_1) \to \theta - \theta = 0$. Thus,

$$\frac{n}{k} \mathbf{Var}(T_i) = \frac{n}{k} \mathbb{E}(T_1^2) + o(1)$$

$$= \frac{n}{k} \left(\mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,d} \right) + \theta^2 \mathbb{P}\left(U_1 < \frac{k}{n} \right) - 2\theta \mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,d} \right) \right) + o(1)$$

$$\rightarrow \theta - \theta^2.$$

Second, by the mixing condition, $\frac{2}{k}\sum_{i=r_n+1}^{n-d+1}(n-i)\mathbf{Cov}(T_1,T_{1+i})=o(1)$. For $i\leq d-1$,

$$T_1 T_{i+1} = \theta^2 \mathbb{1} \left(U_1 < \frac{k}{n}, U_{1+i} < \frac{k}{n} \right) - \theta \mathbb{1} \left(U_1 < \frac{k}{n}, U_{1+i} < \frac{k}{n} < m U_{i+2, i+d} \right).$$

Thus, by Conditions (A3) and (A4),

$$\frac{2}{k} \sum_{i=1}^{r_n} (n-i) \mathbf{Cov}(T_1, T_{1+i}) = \frac{2n}{k} \sum_{i=1}^{r_n} \left(1 - \frac{i}{n} \right) \mathbb{E}(T_1 T_{i+1}) + o(r_n k/n)$$

$$= \frac{2n}{k} \sum_{i=1}^{r_n} \theta^2 \mathbb{P}\left(U_1 < \frac{k}{n}, U_{1+i} < \frac{k}{n} \right) - \frac{2n}{k} \sum_{i=1}^{r_n} \theta \mathbb{P}\left(U_1 < \frac{k}{n}, U_{1+i} < \frac{k}{n} < mU_{i+2,i+d} \right)$$

$$- \frac{2n}{k} \sum_{i=d}^{r_n} \theta \mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,d}, U_{i+1} < \frac{k}{n} \right)$$

$$+ \frac{2n}{k} \sum_{i=d}^{r_n} \mathbb{P}\left(U_1 < \frac{k}{n} < mU_{2,d}, U_{i+1} < \frac{k}{n} < mU_{i+2,i+d} \right) \to 2\theta^2 \Lambda_1(1,1) - 2\theta \lambda_1.$$

And the Condition (2) in Utev [1990] follows from the same argument as that for (5.2). Therefore, (5.7) is proved.

Proof for Theorem 2.3

Because the conditions of Theorem 2.1 hold for d = s + 1, $\Delta(s + 1) = \theta$ and the result in Proposition 5.1 holds for d = s + 1.

We also need a similar convergence result for $\sqrt{k} \left(\tilde{\theta}_n(x,s) - \theta_n(x,s) \right)$. Clearly, the covariance of the limit is not necessary in the same form because $D^{(s)}$ is not guaranteed. However

the condition (2.3) makes sure that the covariance exists. The tightness of the process still holds for d = s. Follow the same line as in the proof for Proposition 5.1, we have

$$\sqrt{k}\left(\tilde{\theta}_n(x,s) - \theta_n(x,s)\right) \stackrel{d}{\to} V(x),$$

in D([1/2,3/2]), where V is a mean-zero Gaussian process with covariance structure $\mathbb{E}(V(x)V(y))=\lim_{n\to\infty}\frac{n}{k}\sum_{i=0}^{s-1}\mathbb{P}\left(U_1<\frac{kx}{n}< mU_{2,s}, U_{1+i}<\frac{ky}{n}< mU_{2+i,s+i}\right)+\Lambda_2(x,y)+\Lambda_2(y,x),$ for $x\leq y$. Note that if Condition (A2) holds for d=s, then $\Lambda_2(x,y)=0$ and $V=dW_s$. It is clear that

$$\mathbb{E}(V^2(1)) = \Delta(s) + 2\Lambda_2(1, 1). \tag{5.8}$$

Recall that $\delta(s) = \Delta(s) - \Delta(s+1) = \Delta(s) - \theta$. We have,

$$\sqrt{k} \left(\hat{\theta}(s) - \hat{\theta}(s+1) - \delta(s) \right)
= \sqrt{k} \left(\tilde{\theta}_n \left(\frac{n}{k} U_{k,n}, s \right) - \Delta(s) \right) - \sqrt{k} \left(\tilde{\theta}_n \left(\frac{n}{k} U_{k,n}, (s+1) \right) - \theta \right)
\stackrel{p}{\to} \left(V(1) - \Delta(s) \tilde{W}(1) \right) - \left(W_{s+1}(1) - \theta \tilde{W}(1) \right)
= V(1) - W_{s+1}(1) - \delta(s) \tilde{W}(1)$$

where the convergence from the same argument used in obtaining (5.6). Thus, to prove the result, it suffices to show that

$$\sqrt{k} \left(\tilde{\theta}_n(1,s) - \tilde{\theta}_n(1,s+1) - \frac{\delta(s)}{k} \sum_{i=1}^n \mathbb{1} \left(U_i < \frac{k}{n} \right) - \left(\theta_n(1,s) - \theta_n(1,s+1) - \delta(s) \right) \right)$$

$$\stackrel{d}{\to} N(0,\sigma_2^2).$$

$$(5.9)$$

Define $I_i = \mathbb{1}\left\{U_i < \frac{k}{n} < mU_{i+1,i+s-1}\right\}$, $J_i = \mathbb{1}\left\{U_i < \frac{k}{n} < mU_{i+1,i+s}\right\}$ and $K_i = \mathbb{1}\left\{U_i < \frac{k}{n}\right\}$. We need to show that

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{n-s} \left(I_i - J_i - \delta(s) K_i - \mathbb{E}(I_i - J_i - \delta(s) K_i) \right) \stackrel{d}{\to} N(0, \sigma_2^2).$$

We have

$$\mathbf{Var}\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n-s} (I_{i} - J_{i} - \delta(s)K_{i})\right) \\
= \mathbf{Var}\left(V(1)\right) + \mathbf{Var}\left(W_{s+1}(1)\right) + \delta^{2}(s)\mathbf{Var}\left(\tilde{W}(1)\right) + o(1) \\
- \frac{2}{k}\mathbf{Cov}\left(\sum_{i=1}^{n-s} I_{i}, \sum_{i=1}^{n-s} J_{i}\right) - \frac{2\delta(s)}{k}\mathbf{Cov}\left(\sum_{i=1}^{n-s} I_{i}, \sum_{i=1}^{n-s} K_{i}\right) + 2\delta(s)\mathbf{Cov}\left(W_{s+1}(1), \tilde{W}(1)\right) \\
= \Delta(s) + 2\Lambda_{2}(1, 1) + \theta + \delta^{2}(s)(1 + 2\Lambda_{1}(1, 1)) + o(1) \\
- \frac{2}{k}\mathbf{Cov}\left(\sum_{i=1}^{n-s} I_{i}, \sum_{i=1}^{n-s} J_{i}\right) - \frac{2\delta(s)}{k}\mathbf{Cov}\left(\sum_{i=1}^{n-s} I_{i}, \sum_{i=1}^{n-s} K_{i}\right) + 2\delta(s)(\theta + \lambda_{1}), \tag{5.10}$$

where we used that $\mathbf{Cov}\left(W_{s+1}(1), \tilde{W}(1)\right) = \theta + \lambda_1$, from the proof for Theorem 2.1. Next, we compute two covariance term. By stationarity and Lemma 6.2,

$$\frac{1}{k} \mathbf{Cov} \left(\sum_{i=1}^{n-s} I_i, \sum_{i=1}^{n-s} J_i \right) \\
= \frac{n}{k} \mathbf{Cov} (I_1, J_1) + \frac{1}{k} \sum_{i=1}^{r_n} (n - s - i) \mathbf{Cov} (I_1, J_{1+i}) + \frac{1}{k} \sum_{i=1}^{r_n} (n - s - i) \mathbf{Cov} (I_{1+i}, J_1) + o(1) \\
= \frac{n}{k} \mathbb{E} (I_1 J_1) + \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{E} (I_1 J_{1+i}) + \frac{n}{k} \sum_{i=1}^{r_n} E (I_{1+i} J_1) + O\left(\frac{kr_n}{n}\right) + o(1) \\
= \frac{n}{k} \mathbb{P} \left(U_1 < \frac{k}{n} < mU_{2,s+1} \right) + \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P} \left(U_1 < \frac{k}{n} < mU_{2,s}, U_{i+1} < \frac{k}{n} < mU_{i+2,i+s+1} \right) \\
+ \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P} \left(U_{i+1} < \frac{k}{n} < mU_{i+2,i+s}, U_1 < \frac{k}{n} < mU_{2,s+1} \right) + o(1) \\
= \theta + \lambda_2 + o(1), \tag{5.11}$$

by Condition (A2) for d = s + 1. Similarly, we have

$$\frac{1}{k} \mathbf{Cov} \left(\sum_{i=1}^{n-s} I_i, \sum_{i=1}^{n-s} k_i \right) \\
= \frac{n}{k} \mathbb{P} \left(U_1 < \frac{k}{n} < m U_{2,s} \right) + \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P} \left(U_1 < \frac{k}{n} < m U_{2,s}, U_{i+1} < \frac{k}{n} \right) \\
+ \frac{n}{k} \sum_{i=1}^{r_n} \mathbb{P} \left(U_{i+1} < \frac{k}{n} < m U_{i+2,i+s}, U_1 < \frac{k}{n} \right) + o(1) \\
= \Delta(s) + \tilde{\lambda}_1 + \lambda_3 + o(1). \tag{5.12}$$

Combining (5.10), (5.11) and (5.12), it yields that

$$\begin{aligned} & \mathbf{Var}\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{n-s}\left(I_i-J_i-\delta(s)K_i\right)\right) \\ \rightarrow & \delta^2(s)\left(2\Lambda_1(1,1)-1\right)-2\delta(s)\left(\tilde{\lambda}_1-\lambda_1+\lambda_3-\frac{1}{2}\right)+2\Lambda_2(1,1)-2\lambda_2=\sigma_2^2 \end{aligned}$$

And the Condition (2) in Utev [1990] follows from the same argument as that for (5.2). Therefore, (5.9) is proved.

6 Lemmas and three propositions

Lemma 6.1 Define $I_i(x) := \mathbb{1} \{ U_i < \frac{kx}{n} < mU_{i+1,i+d-1} \}$ for i = 1, ..., n-d+1 and $1/2 \le x \le 3/2$. Assume that $\frac{r_n k}{n} = o(1)$ and Condition A(2) holds. Then,

$$\sum_{i=1}^{r_n} (n-i)\mathbf{Cov}(I_1(x), I_{1+i}(x)) = o(k), \tag{6.1}$$

and

$$\sum_{i=d}^{r_n} (n-i)\mathbf{Cov}(I_1(x), I_{1+i}(y)) = o(k),$$
(6.2)

for any $y \in [1/2, 3/2]$.

Proof Observe that $\mathbb{E}(I_1(x)) \leq \mathbb{P}(U_1 < 2k/n) = O\left(\frac{k}{n}\right)$, for any $x \in [1/2, 2]$. By construction, $\mathbb{E}(I_1I_{1+i}) = 0$ for $1 \leq i \leq d-1$. On the other hand, for $d \leq i \leq r_n$, $\mathbb{E}(I_1(x)I_{1+i}(x)) \leq \mathbb{P}(U_1 < \frac{kx}{n} < mU_{2,d}, U_i < \frac{kx}{n})$. Thus, by the Condition A(2),

$$\sum_{i=1}^{r_n} (n-i) \mathbf{Cov}(I_1, I_{1+i}) = \sum_{i=1}^{r_n} (n-i) \left(\mathbb{E}(I_1 I_{1+i}) - (\mathbb{E}(I_1))^2 \right)$$

$$= \sum_{i=1}^{r_n} (n-i) \mathbb{E}(I_1 I_{1+i}) - (\mathbb{E}(I_1))^2 \sum_{i=1}^{r_n} (n-i)$$

$$= o(k) + O\left(\frac{k^2 r_n}{n}\right) = o(k).$$

Hence (6.1) is proved. And, (6.2) follows in the same way because for $d \leq i \leq r_n$, $\mathbb{E}(I_1(x)I_{1+i}(y)) \leq \mathbb{P}(U_1 < \frac{kx}{n} < mU_{2,d}, U_{1+i} < \frac{ky}{n})$.

Lemma 6.2 Let $A \in \sigma\left(\mathbb{1}\left\{U_j \leq \frac{k}{n}\right\}, l \leq j \leq d_1\right)$ and $B_i \in \sigma\left(\mathbb{1}\left\{U_j \leq \frac{k}{n}\right\}, i \leq j \leq i + d_2\right), i = 1, 2, \ldots, n - d_2$, where d_1 and d_2 are two positive integers. If $\frac{n}{k}\mathbb{P}(A) = O(1)$ and $\sum_{i=r_n}^n \left(1 - \frac{i}{n}\right)\phi(i) = o(1)$, then

$$\frac{1}{k} \sum_{i=r_n}^{n-d_2-1} (n-i) \mathbf{Cov}(A, B_{i+1}) = o(1).$$
(6.3)

Proof By the definition of $\phi(i)$ in (2.4), we have that $\mathbf{Cov}(A, B_{i+1}) = \mathbb{P}(A \cap B_{i+1}) - \mathbb{P}(A)\mathbb{P}(B_{i+1}) = \mathbb{P}(A)(\mathbb{P}(B_{i+1}|A) - \mathbb{P}(B_{i+1})) \leq \mathbb{P}(A)\phi(i+1-d_1)$. Thus,

$$\frac{1}{k} \sum_{i=r_n}^{n-d_2-1} (n-i) \mathbf{Cov}(A, B_{i+1}) \le \frac{n}{k} \mathbb{P}(A) \sum_{i=r_n}^{n-d_2-1} \left(1 - \frac{i}{n}\right) \phi(i+1-d_i) = o(1).$$

Proposition 6.1 Suppose that $d^u \geq d^*$ and the assumptions of Corollary 2.4 hold, then

$$\mathbb{P}\left(\hat{d}^*(k) = d^*\right) \stackrel{p}{\to} 1$$

where $\hat{d}^*(k)$ is defined in (2.16).

Proof Denote $\hat{\delta}_n(i) = \hat{\theta}_n(i) - \hat{\theta}_n(i+1)$. By the definition of $\hat{d}^*(k)$,

$$\{\hat{d}^*(k) \ge d^* + 1\} \subset \left\{ \max_{d^* \le i \le d^u} \hat{\delta}_n(i) \ge \frac{1}{\sqrt{k}} \right\},$$

the latter has a probability tending to zero by Corollary 2.4. On the other hand, for any $j \leq d^* - 1$,

$$\{\hat{d}^*(k) = j\} \subset \left\{ \max_{j \le i \le d^u} \hat{\delta}_n(i) < \frac{1}{\sqrt{k}} \right\} \subset \left\{ \hat{\delta}_n(d^* - 1) < \frac{1}{\sqrt{k}} \right\},\,$$

which also has a probability tending to zero. Therefore,

$$\mathbb{P}(\hat{d}^*(k) \neq d^*) \to 0.$$

Proposition 6.2 For the AR(1) model with Cauchy margin defined in (3.1), we have

(a) for
$$z \ge 0$$
, $\ell_s(1) = s - (s - 1)z$, for $s \ge 2$;

(b) for
$$z < 0$$
, $\ell_2(\mathbf{1}) = 2$ and $\ell_s(\mathbf{1}) = s - |z|^2$ for $s \ge 3$.

Proof This result is easily derived by using the independence of $(X_1, \epsilon_2, \dots, \epsilon_s)$. Let v_n be such that $\lim_{n\to\infty} n\mathbb{P}(\epsilon_i > v_n) = 1$. Then $v_n = (1-|z|)u_n$. For $s \geq 2$, we have

$$\begin{split} \ell_s(\mathbf{1}) &= \lim_{n \to \infty} n \mathbb{P} \left(X_1 > u_n \text{ or } \dots \text{ or } X_s > u_n \right) \\ &= \lim_{n \to \infty} n \mathbb{P} \left(X_1 > u_n \text{ or } \dots \text{ or } z^{s-1} X_1 + z^{s-2} \epsilon_2 + \dots + \epsilon_s > u_n \right) \\ &= \lim_{n \to \infty} n \mathbb{P} \left(\frac{X_1}{u_n} > 1 \text{ or } \dots \text{ or } z^{s-1} \frac{X_1}{u_n} + z^{s-2} \frac{\epsilon_2}{u_n} + \dots + \frac{\epsilon_s}{u_n} > 1 \right) \\ &= \lim_{n \to \infty} n \mathbb{P} \left(\frac{X_1}{u_n} > 1 \text{ or } \dots \text{ or } z^{s-1} \frac{X_1}{u_n} + z^{s-2} (1 - |z|) \frac{\epsilon_2}{v_n} + \dots + (1 - |z|) \frac{\epsilon_s}{v_n} > 1 \right) \\ &= \nu \{ (t_1, \dots, t_s) : t_1 > 1 \text{ or } \dots \text{ or } z^{s-1} t_1 + z^{s-2} (1 - |z|) t_2 + \dots + (1 - |z|) t_s > 1 \}, \end{split}$$

where ν denotes the exponent measure of $(X_1, \epsilon_2, \ldots, \epsilon_s)$ (For definition of exponent measure, see Section 6.1.3 in de Haan and Ferreira [2006]). The last convergence follows from Theorem 6.1.11 in de Haan and Ferreira [2006] and the fact that the distribution of $(X_1, \epsilon_2, \ldots, \epsilon_s)$ belongs to the max domain of attraction. Now because of the exact independence of X_1 and ϵ_i 's and therefore asymptotic independence, the exponent measure ν puts mass only in the axes: $\nu\{(t_1, \ldots, t_s) : t_i > a_1 \text{ and } t_j > a_2\} = 0$ for any $i \neq j$ and positive a_1 and a_2 . Then, the result readily follows from the property that $\nu\{(t_1, \ldots, t_s) : t_i > a_1\} = 1/a_1$.

Proposition 6.3 An ARCH model defined in (3.2) does not satisfy $D^{(d)}(u_n)$ condition for any finite d.

Proof We apply Proposition 6.2 of Ehlert et al. [2015] to show that for any finite d,

$$\lim_{n \to \infty} \mathbb{P}(M_{2,d} \le u_n < M_{d+1,r_n} | X_1 > u_n) > 0.$$

In this proof, all the cited equations are referred to the formulas in Ehlert et al. [2015]. Note that ARCH(1,1) model is a special case of the model considered in that paper, which corresponds $\delta_1 = \beta_1 = 0$ in the model given by relations (6.2) and (6.3) in that paper. Therefore $\phi(x) = \alpha_1^{-1/2}|x|$, for the ϕ appeared in the limit of (6.14) in that paper. Let W denote a random variable from Pareto distribution with parameter α and $(Z_i)_{i\geq 1}$ are i.i.d. standard normal random variables. Then by Proposition 6.2 of Ehlert et al. [2015],

$$\lim_{n \to \infty} \mathbb{P}(M_{2,d} \le u_n < M_{d+1,r_n} | X_1 > u_n)$$

$$\ge \lim_{n \to \infty} \mathbb{P}(M_{2,d} \le u_n < X_{d+1} | X_1 > u_n)$$

$$= \mathbb{P}\left(\max_{2 \le i \le d} \frac{W}{|Z_0|} Z_i \prod_{j=0}^{i-1} \phi(Z_j) \le 1 < \frac{W}{|Z_0|} Z_{d+1} \prod_{j=0}^{d} \phi(Z_j)\right)$$

$$= \mathbb{P}\left(W \max_{2 \le i \le d} \alpha_1^{i/2} Z_i \prod_{j=1}^{i-1} |Z_j| \le 1 < W \alpha_1^{(d+1)/2} Z_{d+1} \prod_{j=1}^{d} |Z_j|\right)$$

$$\geq \mathbb{P}\left(W > \alpha_1^{-(d+1)/2}, \max_{2 \leq i \leq d} Z_i \leq -1, Z_{d+1} > 1\right)$$
$$= \alpha_1^{\alpha(d+1)/2} (\Phi(-1))^d > 0,$$

where $\alpha_1 \in (0,1)$ (which equals 1/2 in our simulation example), $\alpha > 0$ and Φ is a standard normal distribution function.

7 Acknowledgements

The author would like to acknowledge Dr. Andrea Krajina for the collaboration and fruitful discussion in an early stage of this project.

References

- D. Aldous. Stopping times and tightness. The Annals of Probability, 6(2):335–340, 1978.
- M.A. Ancona-Navarrete and J.A. Tawn. A comparison of methods for estimating the extremal index. *Extremes*, 3(1):5–38, 2000.
- B. Basrak and J. Segers. Regularly varying multivariate time series. Stochastic Processes and their Applications, 119(4):1055 1080, 2009.
- B. Berghaus and A. Bücher. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *The Annals of Statistics*, 46(5):2307–2335, 2018.
- R.C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probability surveys*, 2:107–144, 2005.
- S. Chamberlain. rnoaa: 'NOAA' Weather Data from R, 2019. URL https://CRAN.R-project.org/package=rnoaa. R package version 0.8.4.
- M.R. Chernick, T. Hsing, and W.P. McCormick. Calculating the extremal index for a class of stationary sequences. *Advances in applied probability*, 23(4):835–850, 1991.
- L. de Haan and A. Ferreira. Extreme Value Theory: An Introduction. Springer Verlag, 2006.
- L. de Haan, S.I. Resnick, H. Rootzén, and C.G. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to arch processes. *Stochastic Processes and their Applications*, 32(2):213 224, 1989.
- E. Eberlein. Weak convergence of partial sums of absolutely regular sequences. Statistics & probability letters, 2(5):291-293, 1984.
- A. Ehlert, U.-R. Fiebig, A. Janßen, and M. Schlather. Joint extremal behavior of hidden and observable time series with applications to GARCH processes. *Extremes*, 18(1):109–140, 2015.
- J.H.J. Einmahl and F.H. Ruymgaart. Some results for empirical processes of locally dependent arrays. *Mathematical Methods of Statistics*, 9(4):399–414, 2000.

- H. Ferreira and M. Ferreira. Estimating the extremal index through local dependence. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 54(2):587–605, 2018.
- C.A.T. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):545–556, 2003.
- T. Hsing. Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics*, 21(4):2043–2071, 1993.
- F. Laurini and J. A. Tawn. New estimators for the extremal index and other cluster characteristics. *Extremes*, 6(3):189–211, 2003.
- M.R. Leadbetter. Extremes and local dependence in stationary sequences. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 65(2):291–306, 1983.
- M.R. Leadbetter and S. Nandagopalan. On exceedance point processes for stationary sequences under mild oscillation restrictions. *In: Hüsler J., Reiss RD. (eds) Extreme Value Theory. Lecture Notes in Statistics*, 51:69–80, 1989.
- M.R. Leadbetter, G. Lindgren, and H. Rootzén. Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York, 1983.
- P.J. Northrop. An efficient semiparametric maxima estimator of the extremal index. *Extremes*, 18(4):585–603, 2015.
- G.L. O'Brien. Extreme values for stationary and markov sequences. *The Annals of Probability*, 15(1):281–291, 1987.
- C.Y. Robert. Inference for the limiting cluster size distribution of extreme values. *The Annals of Statistics*, 37(1):271–310, 2009.
- R.L. Smith and I. Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):515–528, 1994.
- M. Süveges. Likelihood estimation of the extremal index. Extremes, 10:41–55, 2007.
- S.A. Utev. On the central limit theorem for φ -mixing arrays of random variables. Theory of Probability & Its Applications, 35(1):131–139, 1990.
- I. Weissman and S.Yu. Novak. On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference*, 66(2):281 288, 1998.