THRESHOLD SELECTION FOR EXTREMAL INDEX ESTIMATION

Natalia M. Markovich and Igor V. Rodionov∗

July 25, 2021

Abstract

We propose a new threshold selection method for the nonparametric estimation of the extremal index of stochastic processes. The so-called discrepancy method was proposed as a data-driven smoothing tool for estimation of a probability density function. Now it is modified to select a threshold parameter of an extremal index estimator. To this end, a specific normalization of the discrepancy statistic based on the Cramér-von Mises-Smirnov statistic $\omega^2$ is calculated by the $k$ largest order statistics instead of an entire sample. Its asymptotic distribution as $k \to \infty$ is proved to be the same as the $\omega^2$-distribution. The quantiles of the latter distribution are used as discrepancy values. The rate of convergence of an extremal index estimate coupled with the discrepancy method is derived. The discrepancy method is used as an automatic threshold selection for the intervals and $K$-gaps estimators and it may be applied to other estimators of the extremal index.

Keywords: Threshold selection; Discrepancy method; Cramér-von Mises-Smirnov statistic; Nonparametric estimation; Extremal index

1 Introduction

Let $X^n = \{X_i\}_{i=1}^n$ be a sample of random variables (r.v.s) with cumulative distribution function (cdf) $F(x)$. By Leadbetter et al. 1983 the stationary sequence $\{X_n\}_{n \geq 1}$ is said to have extremal index $\theta \in (0, 1]$ if for each $0 < \tau < \infty$ there is a sequence of real numbers $u_n = u_n(\tau)$ such that

∗V.A.Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Profsoyuznaya 65, 117997 Moscow, Russia; e-mail: markovic@ipu.rssi.ru, vecsell@gmail.com
it holds
\[
\lim_{n \to \infty} n(1 - F(u_n)) = \tau, \quad \lim_{n \to \infty} \mathbb{P}\{M_n \leq u_n\} = e^{-\tau\theta},
\]
where \(M_n = \max\{X_1, \ldots, X_n\}\). The extremal index reflects a cluster structure of an underlying sequence or its local dependence. \(\theta = 1\) holds if \(X_1, \ldots, X_n\) are independent. For stationary sequences \(\theta = 1\) when mixing conditions \(D(u_n)\) and \(D'(u_n)\) hold (Leadbetter et al. 1983).

Nonparametric estimators of \(\theta\) require usually the choice of a threshold and/or a declustering parameter. The well-known blocks and runs estimators of the extremal index require an appropriate threshold \(u\) and the block size \(b\) or the number of consecutive observations \(r\) running below \(u\) to separate two clusters (Beirlant et al. 2004). A bias-corrected modification of the blocks estimator (Drees 2011) informs how to avoid the threshold selection by providing a rather stable plot of the extremal index estimates against \(u\) with some remaining uncertainty.

In Sun and Samorodnitsky (2018) the multilevel blocks estimator is proposed where a sequence of increasing levels and a weight function have to be defined. The sliding blocks estimator has asymptotic variance smaller than the disjoint blocks estimator (Robert et al. 2009b) and all of them require the selection of a pair \((u, b)\). The cycles estimator proposed by Ferreira (2018) needs both \(u\) and the cycle size \(s\) as parameters. The intervals estimator of \(\theta\) by Ferro and Segers (2003) and the estimators introduced by Robert (2009b) require the choice of \(u\). The \(K\)-gaps estimator is another threshold-based one (S"uveges and Davison 2010).

One of the high quantiles of the sample \(X^n\) is taken usually as \(u\) or \(u\) is selected visually corresponding to a stability interval of the plot of some estimate \(\hat{\theta}(u)\) against \(u\). Following S"uveges and Davison (2010), a list of pairs \((u, K)\) is selected according to the Information Matrix Test (IMT) in Fukutome et al. (2015). Then \(u\) is selected from such a pair that corresponds to the largest number of clusters of exceedances separated by more than \(K\) non-exceedances. The semiparametric maxima estimators (Berghaus and B"ucher 2018; Northrop 2015) depend on the block size only. The choice of the latter remains an open problem.

The objective of this paper is to propose a new nonparametric tool to find the threshold \(u\).

The so-called discrepancy method was proposed in Markovich (1989) and Vapnik et al. (1992) as a data-driven smoothing tool for a probability density function (pdf) estimation by i.i.d. data. We aim to extend this method for an extremal index estimation. The idea was to find an unknown parameter \(h\) of the pdf as a solution of the discrepancy equation
\[
\rho(\hat{F}_h, F_n) = \delta.
\]
Here, \(\hat{F}_h(x) = \int_{-\infty}^{x} \hat{f}_h(t)dt\) holds, \(\hat{f}_h(t)\) is some pdf estimate, \(\delta\) is a discrepancy value of the estimation of \(F(x)\) by the empirical distribution function \(F_n(x)\), i.e. \(\delta = \rho(F, F_n)\). \(\rho(\cdot, \cdot)\) is a metric in the space of cdf’s. Since \(\delta\) is usually unknown, quantiles of the limit distribution of
the Cramér-von Mises-Smirnov (C-M-S) statistic
\[ \omega_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x), \]
were proposed as \( \delta \). The latter limit distribution which is rather complicated can be found in Bolshev and Smirnov (1965) or Markovich (2007). One can choose other nonparametric statistics like the Kolmogorov-Smirnov or Anderson-Darling ones instead of \( \omega_n^2 \). Limit distributions of these statistics are invariant regarding \( F(x) \) (Bolshev and Smirnov 1965). We will focus on \( \omega_n^2 \). Regarding practical applications the bandwidth \( h \) was proposed in Markovich (1989) as a solution of the equation
\[ \hat{\omega}_n^2(h) = 0.05. \] (1.1)
Here,
\[ \hat{\omega}_n^2(h) = \sum_{i=1}^{n} \left( \hat{F}_h(X_{i,n}) - \frac{i - 0.5}{n} \right)^2 + \frac{1}{12n} \]
was calculated by the order statistics \( X_{1,n} \leq \ldots \leq X_{n,n} \) corresponding to the sample \( X^n \), and the value 0.05 corresponding to the mode of the pdf of the statistic \( \omega_n^2 \) and thus, the maximum likelihood value of the \( \omega_n^2 \) was found by tables of the statistic \( \omega_n^2 \) (Bolshev and Smirnov 1965) as the discrepancy value \( \delta \). A similar idea was explored in Markovich (2015) to estimate the extremal index.
Following (Ferro and Segers 2003; Markovich 2014, 2016b, 2017) one can determine a cluster as the number of consecutive observations exceeding the threshold \( u \) between two consecutive non-exceedances. Ferro and Segers (2003) state that the times between exceedances of a threshold \( u \) by the process \( \{X_i\} \), is a random variable \( T(u) \) equal in distribution to \( T_1(u) \). It holds
\[ T_1(u) = \min\{j \geq 1 : M_{1,j} \leq u, X_{j+1} > u | X_1 > u\}, \]
where \( M_{i,j} = \max\{X_{i+1}, \ldots, X_j\}, M_{1,1} = -\infty \). \( T_1(u_n) \) normalized by the tail function \( \{Y = F(u_n)T_1(u_n)\} \), is derived to be asymptotically exponentially distributed with a weight \( \theta \) and with an atom at zero with a weight \( 1 - \theta \) (Ferro and Segers 2003).
Taking the exceedance times \( 1 \leq S_1 < \ldots < S_{N_u} \leq n \), the observed interexceedance times are \( T_i = S_{i+1} - S_i \) for \( i = 1, \ldots, N_u - 1 \), where \( N_u = \sum_{i=1}^{n} X_i > u \) is the number of observations which exceed a predetermined high threshold \( u \) (Ferro and Segers 2003). Denote further \( L \equiv L(u) = N_u - 1 \). In case of the statistic \( \omega_n^2 \), the discrepancy equation may be calculated by

\footnote{Theoretically, events \( \{T_i = 1\} \) are allowed. In practice, such cases that mean single inter-arrival times between consecutive exceedances are meaningless.}
the $k$, $1 \leq k \leq L$, largest order statistics of a sample $\{Y_i = (N_u/n)T_i\}$ as follows

$$\hat{\omega}^2_L(u) = \sum_{i=L-k+1}^L \left( \hat{G}(Y_{i,L}) - \frac{i - 0.5}{L} \right)^2 + \frac{1}{12L} = \delta.$$  \hspace{1cm} (1.2)

Here, the distribution model of the normalized inter-exceedance times $\hat{G}(Y_{i,L})$ is determined by $G(t) = 1 - \theta \exp(-\theta t)$ with a substitution of $\theta$ by some estimate $\hat{\theta}(u)$ and of $t$ by the order statistic $Y_{i,L}$ (Markovich 2015). A value of $u$ can be found as a solution of (1.2) with regard to any consistent nonparametric estimator of $\theta$. The calculation (1.2) by the entire sample may lead to the lack of a solution of the discrepancy equation regarding $u$ the same way as for the heavy-tailed pdf estimation in (Markovich 2007; Markovich 2016a) or to too large $u$’s which may be not appropriate for the estimation of $\theta$.

The selection of $k$ and $\delta$ remains a problem. To overcome this problem we find a specific normalization of the discrepancy statistic $\hat{\omega}^2_L(u)$ in (1.2) such that its limit distribution is the same as for the C-M-S statistic. Then its quantiles may be used as $\delta$. Although the limit distribution of the C-M-S does not depend on $k$ as $k \to \infty$, for moderate samples the selection of $k$ is necessary. From Theorems 1 and 2 it is natural to select $k$ such that $k \leq \lfloor \theta L(u) \rfloor$.

Specifically, for the discrepancy statistic the choice of $k$ has to be modified to satisfy Theorem 3. The discrepancy method can be easily applied not only to the threshold-based but generally to any extremal index estimator. In contrast to approaches based on the minimum of the mean squared error with regard to the threshold or another tuning parameter, e.g. Robert et al. (2009b), the discrepancy method does not attract any knowledge about the asymptotic rates of the variance and the bias of estimates. Similarly to Hall (1990), the minimization of the bootstrap mean squared error may require two additional parameters to select the size of a resample and to relate thresholds based on the resamples and the entire sample.

The paper is organized as follows. In Section 2 related work is recalled. In Section 3 a normalization of $\hat{\omega}^2_L(\theta)$ denoted as $\tilde{\omega}^2_L(\hat{\theta})$ is found which has the limit distribution of $\omega^2_n$ (Theorem 2). The convergence of $\tilde{\omega}^2_L(\hat{\theta})$ to the $\omega^2$-distribution is derived when the difference $\sqrt{m_n}(\hat{\theta} - \theta)$, where $m_n$ is some sequence relating to $k$ and $L$, has a nondegenerate distribution (Theorem 3). In Theorem 4 the consistency and the inconsistency conditions for the normalized statistic $\tilde{\omega}^2_L(\hat{\theta})$ are given. The rate of convergence of the extremal index estimates with the threshold selected by the discrepancy method is derived in Corollary 1. The choice of the $k$ largest order statistics for $\tilde{\omega}^2_L$ by samples of moderate sizes is discussed. Finally, an algorithm and a simulation study of the discrepancy method based on the normalized $\tilde{\omega}^2_L(\hat{\theta})$ statistic is given in Section 4 and an illustration with real data is stated in Section 5. Proofs can be found in Section 6.
2 Important mathematical results

Our results are based on Lemmas 2.2.3, 3.4.1 by de Haan and Ferreira (2006) concerning the limit distributions of the order statistics and Theorem 1 by Ferro and Segers (2003).

Lemma 1. (de Haan and Ferreira 2006; Smirnov 1952) Let \( U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n} \) be the \( n \)th order statistics from a standard uniform distribution. Then, as \( n \to \infty, k \to \infty, n - k \to \infty \),

\[
\frac{U_{k,n} - b_n}{a_n}
\]

is asymptotically standard normal with

\[
b_n = \frac{k - 1}{n - 1}, \quad a_n = \sqrt{\frac{b_n(1 - b_n)}{n - 1}}.
\]

Lemma 2. (de Haan and Ferreira 2006) Let \( X, X_1, X_2, \ldots, X_n \) be i.i.d. r.v.s with common cdf \( F \), and let \( X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n} \) be the \( n \)th order statistics. The joint distribution of \( \{X_{i,n}\}_{i=n-k+1}^n \) given \( X_{n-k,n} = t \), for some \( k \in \{1, \ldots, n-1\} \), equals the joint distribution of the set of order statistics \( \{X_{i,k}^*\}_{k=1}^n \) of i.i.d. r.v.s \( \{X_i^*\}_{i=1}^n \) with cdf

\[
F_i(x) = P\{X \leq x | X > t\} = \frac{F(x) - F(t)}{1 - F(t)}, \quad x > t.
\]

Definition 1. (Ferro and Segers 2003) For real \( u \) and integers \( 1 \leq k \leq l \), let \( F_{k,l}(u) \) be the \( \sigma \)-field generated by the events \( \{X_i > u\}, k \leq i \leq l \). Define the mixing coefficients \( \alpha_{n,q}(u) \),

\[
\alpha_{n,q}(u) = \max_{1 \leq k \leq n-q} \sup |P(B|A) - P(B)|,
\]

where the supremum is taken over all \( A \in F_{1,k}(u) \) with \( P(A) > 0 \) and \( B \in F_{k+q,n}(u) \) and \( k, q \) are positive integers.

The next theorem states that

\[
F(u_n)T_1(u_n) \to^d T_\theta = \begin{cases} \eta, & \text{with probability } \theta, \\ 0, & \text{with probability } 1 - \theta, \end{cases}
\]

where \( \eta \) is exponentially distributed with mean \( \theta^{-1} \). The zero asymptotic inter-exceedance times (the intracluster times) imply the times between the consecutive exceedances of the same cluster. The positive asymptotic inter-exceedance times are the inter-cluster times. \( \to^d \) denotes convergence in distribution.

Theorem 1. (Ferro and Segers 2003) Let \( \{X_n\}_{n \geq 1} \) be a stationary process of r.v.s with tail function \( \overline{F}(x) = 1 - F(x) \). Let the positive integers \( \{r_n\} \) and the thresholds \( \{u_n\}, n \geq 1 \), be such that \( r_n \to \infty \), \( r_n \overline{F}(u_n) \to \tau \) and \( P\{M_{r_n} \leq u_n\} \to \exp(-\theta \tau) \) hold as \( n \to \infty \) for some
If \( \tau \in (0, \infty) \) and \( \theta \in (0, 1] \). If there are positive integers \( q_n = o(r_n) \) such that \( \alpha_{cr_n q_n}(u_n) = o(1) \) for any \( c > 0 \), then we get for \( t > 0 \)

\[
P\{F(u_n)T_1(u_n) > t\} \to \theta \exp(-\theta t), \quad n \to \infty.
\]

The intuition for declustering of a sample is given in Ferro and Segers (2003). One can assume that the largest \( C - 1 = [\theta L] \) inter-exceedance times are approximately independent inter-cluster times. The larger \( u \) corresponds to the larger inter-exceedance times whose number \( L \equiv L(u) \) may be small. It leads to a larger variance of the estimates based on \( \{T_i(u)\} \). The intervals estimator follows from Theorem 1. It is defined as (Beirlant et al. 2004, p. 391),

\[
\hat{\theta}_n(u) = \begin{cases} 
\min(1, \hat{\theta}_1^n(u)), & \text{if } \max\{T_i : 1 \leq i \leq L\} \leq 2, \\
\min(1, \hat{\theta}_2^n(u)), & \text{if } \max\{T_i : 1 \leq i \leq L\} > 2,
\end{cases}
\] (2.2)

where

\[
\hat{\theta}_1^n(u) = \frac{2(\sum_{i=1}^{L} T_i)^2}{L \sum_{i=1}^{L} T_i^2}, \quad \hat{\theta}_2^n(u) = \frac{2(\sum_{i=1}^{L} (T_i - 1))^2}{L \sum_{i=1}^{L} (T_i - 1)(T_i - 2)}.
\]

The \( K \)-gaps estimator was proposed in S"uveges and Davison (2010) as alternative to the intervals estimator, where the \( K \)-gaps

\[
S(u_n)^{(K)} = (\max (T_1(u_n) - K, 0)), \quad K = 0, 1, 2, ..., 
\]

have the same limiting mixture law (2.1). The \( K \)-gaps estimator is obtained by the maximum likelihood method using the model (2.1) and assuming that the \( K \)-gaps observations are independent. It has the following form

\[
\hat{\theta}^K = 0.5 \left( \frac{a + b}{c + 1} + \sqrt{((a + b)/c + 1)^2 - 4b/c} \right),
\] (2.3)

with \( a = L - N_C \), \( b = 2N_C \), \( c = \sum_{i=1}^{L} F(u_n)S(u_n)^{(K)}_i \). \( N_C \) is the number of non-zero \( K \)-gaps.

The \( K \)-gaps estimator is consistent and asymptotically normal as \( L \to \infty \). Since the \( K \)-gaps may have a distribution different from (2.1) for moderate samples, \( K \) is selected by a model misspecification test. The iterative weighted least squares estimator of S"uveges (2007) explores the inter-exceedance times with \( K = 1 \). The automatic selection of an optimal pair \((u, K)\) is proposed in Fukutome et al. (2015) by a choice of pairs for which values of the statistic of the information matrix test (the IMT) are less than 0.05. The test works satisfactorily when the number of exceedances is not less than 80.

The intervals estimator is derived to be consistent for \( m \)-dependent processes (Ferro and Segers 2003). Asymptotic normality \( \sqrt{m_n}(\hat{\theta}_n(u) - \theta) \to^d N(0, V) \) as \( n \to \infty \) is derived for several extremal index estimators and different values of variance \( V \). Here, \( m_n = np_n \) holds for the blocks and runs estimators, where \( p_n = P\{X_1 > u_n\} \) holds in Weissman and Novak (1978);
$m_n = L(u_n)$ is the number of inter-exceedance times $\{T_i(u_n)\}$ for the intervals estimator in Robert (2009a), $\lfloor x \rfloor$ denotes the integer part of $x$; $m_n = n/p_n$ is taken for the multilevel blocks estimator in Sun and Samorodnitsky (2018), where $p_n \varphi(u_n^s) \rightarrow \tau_s$ and $s \in \{1, ..., m\}$ is the number of levels $\{u_n^s\}$; $m_n = \lfloor n/r_n \rfloor$ is used for the disjoint and sliding blocks estimators, where $r_n = o(n)$ are positive integers related to a mixing condition (Robert 2009a); $m_n = k_n$ is taken for the disjoint and sliding blocks estimators by Berghaus and Bücher (2018) and Northrop (2015), where $k_n$ is a number of blocks of length $b_n$ such that $k_n = o(b_n^2)$ holds as $n \rightarrow \infty$.

3 Main results: $\omega^2$-distribution of the normalized Cramér-von Mises-Smirnov statistic

3.1 Normalized Cramér-von Mises-Smirnov statistic for known $\theta$

Let us rewrite the left-hand side (1.2) in the following form

$$\hat{\omega}_L^2(u) = \sum_{i=L-k+1}^{L} \left( 1 - \theta \exp(-Y_{i,L}\theta) - \frac{i-0.5}{L} \right)^2 + \frac{1}{12L}$$

and derive its limit distribution. Note that $L = L(u_{r_n})$ is a sequence of r.v.s converging in probability to infinity due to (Theorem 2.1, Robert 2009a). In sequel, the limit distribution of the concerned statistics does not depend on $L$, so we can neglect its randomness. The threshold sequence $u_{r_n}$ introduced in Robert (2009a) corresponds to the point process of time normalized exceedances defined on $(0, \infty)$, in contrast to a traditional point process defined on $(0, 1)$ (Beirlant et al. 2004).

According to (Martynov 1978; Smirnov 1952) the limit distribution of the C-M-S statistic (1.1) or of $\omega_n^2 = n \int_0^1 (F_n(t) - t)^2 dt$ coincides with the distribution of

$$\Omega = \int_0^1 B^2(t) dt,$$

where $B(t)$ is a Brownian bridge on $[0, 1]$, i.e. the Gaussian random process with zero mean and the covariance function $R(s, t) = \min(s, t) - st$, $s, t \in [0, 1]$. Then the statistic (3.1) built by the $k$ largest order statistics tends to 0 for $k = o(L)$ as $L \rightarrow \infty$, since the interval over which we integrate $B^2(t)$ tends to an empty set. Thus, (3.1) has to be normalized. Let us consider
the normalization of (3.1)
\[ \tilde{\omega}^2_L(\theta) = \frac{1}{(1 - t_k)^2}. \]
\[ \cdot \sum_{i=L-k+1}^{L} \left( 1 - \theta \exp(-Y_{i,L}\theta) - t_k - \frac{i - (L - k) - 0.5}{k} (1 - t_k) \right)^2 + \frac{1}{12k}, \]
where \( t_k = 1 - \theta \exp(-Y_{L-k,L}\theta) \). Let us explain in more detail why we need such normalization. It follows from (Robert 2009a, Theorem 2.1), that there are a probability, \( \theta \), of asymptotic positive inter-exceedance times (the inter-cluster times) and a probability, \( 1 - \theta \), of zero asymptotic inter-exceedance times (the intra-cluster times). Moreover, the inter-cluster times are asymptotically independent exponential with mean \( 1/\theta \). Let us consider the following statistic
\[ \omega^2_k(\theta) = \frac{1}{Z^2_{L-k,L}} \cdot \sum_{i=L-k+1}^{L} \left( Z_{L-k,L} - Z_{i,L} - \frac{i - (L - k) - 0.5}{k} Z_{L-k,L} \right)^2 + \frac{1}{12k}, \]
where \( Z_{i,L} = \theta \exp(-T_{i,L}^*\theta) \), \( T_{1,L}^* \leq \ldots \leq T_{L,L}^* \) are order statistics of a sample \( \{T_{i}^*\} \), \( \{T_{i}^*\} \) are independent copies of \( T_\theta \).

It follows from (3.3), Theorem 1 and Lemma 2, that the conditional distribution of the set of order statistics \( \{1 - Z_{i,L}\} \) given \( 1 - Z_{L-k,L} = t_k \) asymptotically agrees for \( \lim \sup_{n \to \infty} k/L < \theta \) with the distribution of the set of order statistics \( \{U_{j,k}^*\} \), \( j = i - (L - k) \), of an i.i.d. sample \( \{U_j^*\} \) from the uniform distribution on \( [t_k, 1] \). The asymptotical distribution of \( \omega^2_k(\theta) \) is given in the next theorem.

**Theorem 2.** It holds
\[ \omega^2_k(\theta) \overset{d}{\to} \xi \]
as \( k \to \infty, L \to \infty, \lim \sup_{n \to \infty} k/L < \theta \), where \( \xi \) has the distribution function \( A_1 \), which is the limit distribution function of the C-M-S statistic \( \omega_n^2 \).

**Remark 1.** Based on the proof of Theorem 2, one can propose the goodness-of-fit test of von Mises’ type to check the hypothesis \( H_0 : F(x) = F_0(x) \) for sufficiently large \( x \) using the largest order statistics of a sample \( \{X_i\}_{i=1}^n \). The test statistic is the following
\[ \omega^2_k = \sum_{i=0}^{k-1} \left( \frac{F_0(X_{n-i,n}) - F_0(X_{n-k,n})}{1 - F_0(X_{n-k,n})} - \frac{k - i - 0.5}{k} \right)^2 + \frac{1}{12k}. \]

Theorem 2 implies, that the limit distribution of the statistic \( \omega_k^2 \) under the hypothesis \( H_0 \) does not depend on \( k \) and \( n \). It is equal to the limit distribution of the C-M-S statistic, if the hypothesis
is true. The consistency of the proposed test follows from the equality in distribution of the test statistic given $F_0(X_{n-k,n}) = t$ and the C-M-S statistic. The test is based on the largest order statistics of a sample. It is reasonable both if only the upper tail of the distribution is of interest and/or if the largest order statistics of a sample are only available.

3.2 Normalized Cramér-von Mises-Smirnov statistic for unknown $\theta$

Here, we check whether one can substitute $\theta$ by its estimate $\hat{\theta}$ in (3.2) and find conditions imposed on $\hat{\theta}$ under which the limit distribution of $\tilde{\omega}_2^2(\theta)$ will be the same as the limit distribution of $\omega_2^2(\theta)$. Recall again that the number of inter-exceedance times $L = L(u_{r_n})$ is a sequence of r.v.s tending to $+\infty$ (Robert 2009a). It follows from Theorem 2 that the limit distribution of $\omega_2^2(L, \theta)$ does not depend on $L(u_{r_n})$. We can assume that $L(u_{r_n})$ is a numerical sequence tending to infinity as $n \to \infty$. We will write $L$ instead of $L(u_{r_n})$ in the sequel.

In spirit of Theorem 3.2 (Robert 2009a), the limit distribution of the following statistic

$$\sqrt{L} \left( \sum_{i=1}^{L-1} f(Y_i) - Ef(Y_1) \right)$$

for some continuous $f$ may not depend on a substitution of the set of r.v.s $\{T_i^*\}_{i=1}^{L-1}$ appearing in (3.3) instead of $\{Y_i\}_{i=1}^{L-1}$. Moreover, $T_i^* \stackrel{d}{=} T_\theta$, $i \in \{1, \ldots, L-1\}$ and there is a probability $\theta$ of the nonzero elements of this set that are independent exponentially distributed with parameter $\theta$. For convenience, we accept further $L$ instead of $L(u_{r_n})$. For these r.v.s, Theorem 2.2.1 (de Haan and Ferreira 2006) implies that if $k/L \to 0$ and $k \to \infty$ as $L \to \infty$, then

$$\sqrt{k}(T_{L-k,L}^* - \ln(L\theta/k)/\theta) = O_P(1).$$

In light of these remarks let us assume that there exists a sample of independent exponentially distributed r.v.s $\{E_i^{(L)}\}_{i=1}^{l}$ with mean $\theta^{-1}$ for all large enough $L$ such that

$$Y_{L-k,L} - E_{l-k,L} = o_P \left( \frac{1}{\sqrt{k}} \right)$$

(3.4)

if $k/L \to 0$ and $k \to \infty$ as $L \to \infty$, where we denote $l = \lfloor \theta L \rfloor$ and assume $k < l$. Here and further, we denote for brevity a sequence of positive integers $\{k_n\}$ as $k$. Theorem 2 remains valid when $T_{i,L}^*, i \in \{L-k, \ldots, L\}$ in $\omega_2^2(\theta)$ are substituted by $E_{i,L}^{(L)}$, $i \in \{l-k, \ldots, l\}$.

**Theorem 3.** Let the conditions of Theorem 7 and the condition (3.4) be fulfilled and the estimator of the extremal index $\hat{\theta} = \hat{\theta}_n$ be such that

$$\sqrt{m_n(\hat{\theta}_n - \theta)} \xrightarrow{d} \zeta, \quad n \to \infty,$$

(3.5)
where the r.v. $\zeta$ has a nondegenerate distribution function $H$. Let us assume that the sequence $m_n$ is such that
\[ \frac{k}{m_n} = o(1) \quad \text{and} \quad \frac{(\ln L)^2}{m_n} = o(1) \] (3.6)
as $n \to \infty$. Then
\[ \tilde{\omega}_L^2(\hat{\theta}_n) \overset{d}{\to} \xi \sim A_1 \]
holds, where $A_1$ is the limit distribution function of the C-M-S statistic.

**Remark 2.** Normal distributions give examples of $H$ regarding the intervals, blocks and sliding blocks estimators of the extremal index (Northrop 2015; Robert 2009a; Robert et al. 2009; Sun and Samorodnitsky 2018).

**Remark 3.** The replacement of $o(1)$ by $O(1)$ in (3.4) violates Theorem 4. The assumption $k = O(m_n)$ may lead to a limit distribution of $\tilde{\omega}_L^2(\hat{\theta}_n)$ different from $A_1$ that is out of scope of the paper.

**Remark 4.** The limit process of the point process of exceedance times is a compound Poisson process (Hsing et al. 1988). The condition (3.4) shows in fact the rate of this convergence required for the limit distribution of the normalized C-M-S statistic that is built by the largest $k$ order statistic to preserve the limit distribution of $\omega^2$.

**Theorem 4.** Let the conditions of Theorem 1 and (3.4) be fulfilled. Assume that the sequence of estimates $\{\hat{\theta}_n\}$ is such that for some $\alpha \in [0, 1/2]$
\[ k_n^{\alpha} |\hat{\theta}_{ns} - \theta| \overset{P}{\to} +\infty, \quad \text{if} \quad 0 < \alpha \leq 1/2, \]
\[ |\hat{\theta}_{ns} - \theta| > \varepsilon \quad \text{for some} \quad \varepsilon > 0, \quad \text{if} \quad \alpha = 0 \]
hold as $n \to \infty$ for some subsequence $\{k_{ns}\}$, $s \geq 1$, of the sequence $\{k_n\}$. Then for corresponding subsequence $\{L_s\}$ of the sequence $\{L\}$
\[ \tilde{\omega}_{L_s}^2(\hat{\theta}_{ns})/k_n^{1-2\alpha} \overset{P}{\to} +\infty \]
holds as $n \to \infty$.

**Remark 5.** Theorem 4 implies that the non-consistency of the estimator $\hat{\theta}_n$ or the consistency with a sufficiently slow rate leads to the non-consistency of $\tilde{\omega}_{L_s}^2(\hat{\theta}_{ns})$ in a sense that its limit distribution does not exist or the latter statistic tends to $+\infty$. In case that $\alpha \neq 0$ holds, the estimator $\hat{\theta}_n$ may be consistent but with the rate of convergence slower than $k_n^{-\alpha}$. Hence, $\tilde{\omega}_L^2(\hat{\theta}_n)$ may be considered as a quality functional of $\hat{\theta}_n$. 

10
The consistency of the corresponding extremal index estimates follows from Theorem 4. The next corollary states, if the solutions of the discrepancy equation exist for each \( n \), then the consistency is fulfilled.

**Corollary 1.** Let \( \hat{\theta}_n(u_n) \) be an estimator of \( \theta \) and \( \{\tilde{u}_{k,L}\} \) be some sequence of solutions of the discrepancy equation. Then \( \hat{\theta}_n(\tilde{u}_{k,L}) \xrightarrow{P} \theta \) and for arbitrary \( \varepsilon > 0 \)

\[ k^{1/2-\varepsilon} |\hat{\theta}_n(\tilde{u}_{k,L}) - \theta| \xrightarrow{P} 0 \]

hold as \( k \to \infty, L/k \to \infty, L = o(n), n \to \infty \).

The proof of the corollary is based on a negation of the assertion of Theorem 4.

### 3.3 The choice of \( k \)

According to Theorem 3 the asymptotic distribution of \( \tilde{\omega}_L^2(\hat{\theta}_n) \) does not depend on \( k \). The \( k \)-selection gives another viewpoint that using only the largest inter-exceedance times screens out the smallest inter-exceedance times. It is helpful for the reasons discussed in Ferro and Segers (2003) and is the motivation for the introduction of the tuning parameter \( K \) in the \( K \)-gaps estimator of \( \theta \) proposed in S"uveges and Davison (2010).

In practice, for each predetermined \( \delta, u \) and \( L(u) \) one may decrease the \( k \)-value such that \( k \leq \min\{\hat{\theta}_0L(u), L(u)^\beta\}, 0 < \beta < 1, (\hat{\theta}_0 \text{ is some pilot estimate of } \theta) \) until the discrepancy equations have solutions and select the largest one among such \( k \)’s. This choice satisfies Theorem 3 but it is not unique. For instance, one can select \( k = \lfloor (\ln L)^2 \rfloor \). The following simulation shows an ideal case when the accuracy is the best. Namely, the \( K \)-gaps estimator with \( K = 0 \) and \( N_C = k \) coupled with the discrepancy method demonstrates the best choice when \( k = \lfloor \theta L \rfloor \) is chosen. Since \( \theta \) is in reality unknown, one has to take \( k = \lfloor \hat{\theta}_0L \rfloor \). This choice requires an accurate consistent pilot estimate \( \hat{\theta}_0 \).

### 4 Simulation study

In our simulation study we focus on the threshold-based intervals and \( K \)-gaps estimators. We propose also a modification of the \( K \)-gaps estimator with \( K = 0 \) and \( N_C = k \) in (2.3) notated as \( \hat{\theta}^{K_0} \). The latter coupled with the discrepancy method demonstrates the best choice when \( k = \lfloor \theta L \rfloor \) is close to \( \theta \). The natural drawback of the intervals estimator is that it needs a large sample size \( n \) to obtain a moderate size \( L(u) \) for a large \( u \). The same concerns the \( K \)-gaps estimator.
Algorithm 1. 1. Using $X^n = \{X_i\}_{i=1}^n$ and taking thresholds $u$ corresponding to quantile levels $q \in \{0.90, 0.905, ..., 0.995\}$, generate samples of the inter-exceedance times $\{T_i(u)\}$ and the normalized r.v.s

$$\{Y_i\} = \{\overline{F}(u)T_i(u)\} = \{(N_u/n)T_i(u)\}, \ i \in \{1, 2, ..., L\}, \ L = L(u),$$

where $N_u$ is the number of exceedances over threshold $u$.

2. For each $u$ select $k = \lceil \theta_0 L \rceil, \ k = \min\{\lceil \theta_0 L \rceil, \sqrt{L}\}$ (in case $\theta_0 = 1$, accept $k = L - 1$) or $k = \lceil (\ln L)^2 \rceil$, where the intervals estimator \[2.4\] may be selected as a pilot estimator $\hat{\theta}_0 = \hat{\theta}_0(u)$ with the same $u$ as in Item 1.

3. Use a sorted sample $Y_{L-k+1} \leq ... \leq Y_{L,L}$ and find among considered quantiles all solutions $u_1, ..., u_l$ (here, $l$ is a random number) of the following discrepancy equation

$$\tilde{\omega}_L^2(\hat{\theta}) = \frac{1}{(1 - \hat{\theta}^2)^2} \sum_{i=L-k+1}^L \left( 1 - \hat{\theta} \exp(-Y_{i,L,\hat{\theta}}) - \hat{\theta}_k - \frac{i - (L - k) - 0.5}{k} (1 - \hat{\theta}_k) \right)^2 + \frac{1}{12k} = \delta_1,$$

where $\hat{\theta}_k = 1 - \hat{\theta} \exp(-Y_{L-k,L,\hat{\theta}})$, $\hat{\theta} = \hat{\theta}(u)$ is calculated by \[2.2\], and $\delta_1 = 0.05$ is the mode of the C-M-S statistic. If $L < 40$ we should replace $\tilde{\omega}_L^2(\hat{\theta})$ by

$$(\tilde{\omega}_L^2(\hat{\theta}))' = \left( \tilde{\omega}_L^2(\hat{\theta}) - \frac{0.4}{L} + \frac{0.6}{L^2} \right) \left( 1 + \frac{1}{L} \right)$$

and use quantiles of the C-M-S statistic as the discrepancy $\delta$ (Kobzar 2006).

4. For each $u_j, \ j \in \{1, ..., l\}$ calculate $\hat{\theta}(u_j)$ and find

$$\hat{\theta}_1 = \frac{1}{l} \sum_{i=1}^l \hat{\theta}(u_i), \ \hat{\theta}_2 = \hat{\theta}(u_{\min}), \ \hat{\theta}_3 = \hat{\theta}(u_{\max})$$

as resulting estimates, where $u_{\min} = \min\{u_1, ..., u_l\}, \ u_{\max} = \max\{u_1, ..., u_l\}$.

We take the intervals estimator as $\hat{\theta}_0$ since it requires only $u$ as parameter.

Remark 6. For the $K$-gaps estimator the algorithm is the same, but instead of $\{Y_i\}$ one has to use the normalized $K$-gaps $\{\overline{F}(u)S(u)^{(K)}_i\}$. For each value of $u$ one can examine different values of $K$, for instance, $K \in \{1, 2, ..., 20\}$ may be taken. For $K = 0$ $\{Y_i\}$ are still used.

Remark 7. As the solutions of \[4.2\] may not exist among considered quantiles for given $k$ and $K$, we propose to use the inequality

$$\tilde{\omega}_L^2(\hat{\theta}) \leq \delta_2$$

as an alternative, where $\delta_2 = 1.49$ is the 99.98% quantile of the C-M-S statistic.
Figure 1: The best RMSE and Bias for the K-gaps estimator $\hat{\theta}^{K_0}$ with the threshold $u$ selected by the discrepancy equation (4.2) and the corresponding inequality (4.4) and with $k = \lfloor sL \rfloor$, where 'K0dis' and 'K0disEst' correspond to $s = \theta$ and $s = \hat{\theta}_0$ in Tables 1 and 2 respectively, $\hat{\theta}_0$ is a pilot intervals estimate; the best RMSE and Bias among all estimates in Tables 3-8 notated as 'BestEst' against the number of processes related to the column labels in Tables 1-8, and enumerated from left to right as in the tables for sample size $n = 10^5$ (the upper row) and $n = 5000$ (the lower row).
Figure 2: The best RMSE and Bias for the intervals estimator ('Intdis') and K-gaps ('Kdis') estimators with threshold $u$ selected by the discrepancy equation (4.2) and the corresponding inequality (4.4), and for the K-gaps estimator with $u$ selected by the test IMT ('Kimt') and the intervals estimator with the "plateau-finding" algorithm A1 to select $u$ ('IntA1') against the number of processes related to the column labels in Tables 3 and 4, and enumerated from left to right as in the tables for sample size $n = 10^5$.

Remark 8. The discrepancy method is somewhat similar to the multilevel approach by Sun and Samorodnitsky (2018), where a fixed number of levels $u_1 < \ldots < u_m$ is selected such that $F(u_n^s)/F(u_m^s) \to \tau_s/\tau_m$ for some $\tau_1 > \ldots > \tau_m > 0$. In our case, the number of thresholds, which are the solutions of the discrepancy equation, is random and thus, it cannot be considered as an additional parameter.

The discrepancy method is universal and any estimator depending on $u$ can substitute $\hat{\theta}$ in (4.2). In case of the free-threshold estimators one can express a cluster identification parameter such as the block size as depending on $u$ and find the latter by the discrepancy method. For example, the block size can be selected as $b(u) = \lfloor n/L(u) \rfloor$. The simulation study of this case is out of scope of our paper. Comparison of the threshold-based intervals and $K$-gaps estimators with other estimators based on other tuning parameters (like the block size or the length for runs of non-exceedances) is very complicated since the numbers $L(u)$ and $N_C = N_C(u)$ used for calculations are random. That is the reason our comparison concerns only intervals and $K$-gaps estimators coupled with different threshold choice methods.

4.1 Models

In our simulation study we consider the processes MM, ARMAX, AR(1), AR(2), MA(2) and GARCH(1,1) all with known values $\theta$. The simulation is repeated 1000 times with the sample
Figure 3: The best RMSE and Bias for the intervals estimator (left column) and K-gaps estimator (right column) obtained by the Algorithm with (4.2) notated as 'Intdiseq' and 'Kdiseq', and with the inequality (4.4) notated as 'Intdisineq' and 'Kdisineq' against the number of processes related to the column labels in Tables 3 and 4 for sample size $n = 10^5$. 
Figure 4: Ratios $R_1/R_3$ and $R_2/R_3$ of the best RMSE for the intervals estimator (left column) and K-gaps estimator (right column) obtained by the Algorithm with the equation (4.2) notated as 'Intdiseq' and 'Kdiseq', and with the inequality (4.4) notated as 'Intdisineq' and 'Kdisineq' against the number of processes related to the column labels in Tables 3, 5 and 7: The $R_1/R_3$ corresponds to the best results in Table 3 divided to those in Table 7, and the $R_2/R_3$ to those in Tables 5 and 7 respectively, for sample size $n = 10^5$. 
Figure 5: Ratios of the RMSEs $RMSE(\hat{\theta}_i) / \min_i(RMSE(\hat{\theta}_i))$ corresponding to estimates $\{\hat{\theta}_i\}$, $i \in \{1, 2, 3\}$ in (4.3) for the intervals estimator (left column) and the K-gaps estimator (right column) obtained by the Algorithm with (4.4) against the number of processes related to the column labels in Tables 3, 5 and 7. The upper figures correspond to $k = \lfloor \hat{\theta}_0 L \rfloor$ in Table 3, the middle figures to $k = \lfloor \min(\hat{\theta}_0 L, \sqrt{L}) \rfloor$ in Table 5 and the lower figures to $k = \lfloor (\ln L)^2 \rfloor$ in Table 7 for sample size $n = 10^5$. 

17
size $n = 10^5$ of initial measurements $\{X_1, \ldots, X_n\}$. Big sample sizes may lead, however, to moderate size samples $L(u)$ of normalized inter-exceedance times $\{Y_1, \ldots, Y_{L(u)}\}$. We recall the definitions of the processes. The $m$th order MM process is $X_t = \max_{i=0,\ldots,m} \{\alpha_i Z_{t-i}\}$, $t \in \mathbb{Z}$, where $\{\alpha_i\}$ are constants with $\alpha_i \geq 0$, $\sum_{i=0}^{m} \alpha_i = 1$, and $Z_t$ are i.i.d. standard Fréchet distributed r.v.s with the cdf $F(x) = \exp(-1/x)$, for $x > 0$. The extremal index of the process is equal to $\theta = \max_i \{\alpha_i\}$ (Ancona-Navarrete and Tawn 2000). The distribution of $\{X_t\}_{t \geq 1}$ is standard Fréchet. Values $m = 3$ and $\theta \in \{0.5, 0.8\}$ corresponding to $\alpha \in \{0.5, 0.3, 0.15, 0.05\}$ and $\alpha \in \{0.8, 0.1, 0.008, 0.02\}$, respectively, are taken for our study.

The ARMAX process is determined as $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \epsilon_t$, where $\{\epsilon_t\}$ are i.i.d standard Fréchet distributed r.v.s and $P\{\epsilon_t < 0\} = \exp(-1/x)$ holds assuming $X_0 = Z_0$. The extremal index of the process is given by $\theta = 1 - \alpha$, Beirlant et al. (2004). $P\{X_t < x\} = \exp(-1/x)$ holds assuming $X_0 = Z_0$. We consider $\theta \in \{0.25, 0.75\}$.

The positively correlated AR(1) process with uniform noise $(ARu^+)$ is defined by $X_j = (1/r)X_{j-1} + \epsilon_j$, $j \geq 1$ and $X_0 \sim U(0,1)$ with $X_0$ independent of $\epsilon_j$. $X_1 \sim U(0,1)$ holds. For a fixed integer $r = 2$ let $\epsilon_n$, $n \geq 1$ be i.i.d. r.v.s with $P\{\epsilon_1 = k/r\} = 1/r$, $k \in \{0,1,\ldots,r-1\}$. The extremal index of $ARu^+$ is $\theta = 1 - 1/r$ (Chernick et al. 1991). $\theta \in \{0.5, 0.8\}$ corresponding to $r \in \{2, 5\}$ are taken. The negatively correlated AR(1) process with uniform noise $(ARu^-)$ is defined by $X_j = -(1/r)X_{j-1} + \epsilon_j$ with the similar distributed $\epsilon_n$ but with $k \in \{1,\ldots,r\}$. Its extremal index is $\theta = 1 - 1/r^2$ (Chernick et al. 1991). The same $r$ were taken corresponding to $\theta \in \{0.75, 0.96\}$.

We simulate the MA(2) process (Sun and Samorodnitsky 2018) $X_i = pZ_{i-2} + qZ_{i-1} + Z_i$, $i \geq 1$, with $p > 0$, $q < 1$, and i.i.d. Pareto random variables $Z_{-1}, Z_0, Z_1, \ldots$ with $P\{Z_0 > x\} = 1$ if $x < 1$, and $P\{Z_0 > x\} = x^{-\alpha}$ if $x \geq 1$, for some $\alpha > 0$. The extremal index of the process is $\theta = (1 + p^\alpha + q^\alpha)^{-1}$. The cases $\alpha = 2$, $(p,q) = (1/\sqrt{2}, 1/\sqrt{2}), (1/\sqrt{3}, 1/\sqrt{6})$ with corresponding $\theta \in \{1/2, 2/3\}$ are considered. The distribution of the sum of weighted i.i.d. Pareto r.v.s behaves like a Pareto distribution in the tail and its exact form may be obtained by Ramsay (2008).

We consider also processes studied in (Ferreira 2018b; Northrop 2015; Süveges and Davison 2010). These comprise the AR(1) process $X_j = 0.7X_{j-1} + \epsilon_j$, where $\epsilon_j$ is standard Cauchy distributed and $\theta = 0.3$ (ARc); the AR(2) process $X_j = 0.95X_{j-1} - 0.89X_{j-2} + \epsilon_j$, where $\epsilon_j$ is Pareto distributed with tail index 2 and $\theta = 0.25$; $GARCH(1,1)$, $X_j = \sigma_j \epsilon_j$, with $\sigma_j^2 = \alpha + \lambda X_{j-1}^2 + \beta \sigma_{j-1}^2$, $\alpha = 10^{-6}$, $\beta = 0.7$, $\lambda = 0.25$, with $\{\epsilon_j\}_{j \geq 1}$ an i.i.d. sequence of standard Gaussian r.v.s and $\theta = 0.447$ (see Laurini and Tawn 2012).
4.2 Notations

Tables 1 and 2 contain the statistics (4.3) for the $K$-gaps estimates with $K = 0$ coupling with the discrepancy method (4.2) denoted as $\hat{\theta}^{K_{dis}}_i$, $i \in \{1, 2, 3\}$. $|\theta L|$ and $|\hat{\theta}_0 L|$ are considered as options for $k$, where $\hat{\theta}_0$ is a pilot intervals estimate. The sign '-' means that there are no solutions of the discrepancy equation.

The rest of the tables is a partition regarding $k$ for the intervals and $K$-gaps estimators coupled with the discrepancy method. We study $k = |\hat{\theta}_0 L|$ in Tables 3 and 4, $k = \lfloor \min(\hat{\theta}_0 L, \sqrt{L}) \rfloor$ in Tables 5 and 6, and $k = \lfloor (\ln L)^2 \rfloor$ in Tables 7 and 8. In Tables 3 and 4 the statistics (4.3) corresponding to the intervals estimates coupled with the discrepancy method (4.2) are denoted as $\{\hat{\theta}_i\}$, $i \in \{1, 2, 3\}$. The $K$-gaps estimates with pairs $(u, K)$ selected by (4.2) are denoted as $\hat{\theta}^{K_{dis}}_i$, $i \in \{1, 2, 3\}$, and with IMT-selected pairs $(u, K)$ as $\hat{\theta}^{K_{IMT}}$. Statistics (4.3) relating to the intervals and $K$-gaps estimators and corresponding to solutions of the discrepancy inequality (4.4) are denoted by asterisks in all tables. The intervals estimate with the threshold $u$ selected by the "plateau-finding" Algorithm 1 by Ferreira (2018a) is denoted as $\hat{\theta}^{IA1}$. This algorithm seems to be the best one for the intervals estimator among other algorithms proposed in Ferreira (2018a) according to the provided simulation study. For this algorithm we consider the bandwidth $d = \lfloor wn \rfloor$ with $w = 0.25$ and compute the moving average of $2d + 1$ successive points of $\hat{\theta}$. The value $w = 0.005$ used in Ferreira (2018a) demonstrates slightly worse accuracy uniformly for all processes and we do not show it in Tables 3 and 4.

The values in bold and italic bold correspond to the first and second best performances.

4.3 Conclusions and practical recommendations

We propose to select a threshold of the threshold-based intervals and $K$-gaps estimators and a tuning parameter of free-threshold procedures as solutions of the $\omega^2$ discrepancy equation, where the discrepancy value is equal to the mode of the $\omega^2$-statistic, i.e. to its most likelihood value.

On the first view, the intervals threshold-based estimator does not require another parameter to be specified apart of the threshold. The intervals estimator coupled with the discrepancy method works the same way as the $K$-gaps estimator. An additional regularization parameter such as the moving window size for the "plateau-finding" algorithm A1 (Ferreira 2018a) or the number of the largest order statistics $k$ is required to choose the threshold anyway. It is shown in our paper that there is a potential benefit in choosing $k$ jointly with the threshold.

It is proposed in Ferro and Segers (2003) to select the largest $C - 1 = \lfloor \theta L(u) \rfloor$ interexceedance times which are approximately independent intercluster times as associated with the threshold.
We follow a similar way, i.e. \( k = |\hat{\theta}_0 L| \) is used as one of the choices of \( k \).

The best 'ideal estimator' \( \hat{\theta}^{K_0} \) coupled with discrepancy method (4.2), where \( k \) is taken equal to \( |\theta L| \) and \( |\hat{\theta}_0 L| \), is presented in Tables 1 and 2. Fig. 1 shows that \( \hat{\theta}^{K_0} \) with \( k = |\theta L| \) outperforms other estimators in Tables 3-8 with the best RMSE. The results degrade if one selects the intervals estimate as \( \hat{\theta}_0 \). A small deviation from \( \theta \) does not worsen the best estimate much.

The discrepancy method is competitive with threshold choices such as the IMT and “plateau-finding” algorithms and it improves substantially the existing intervals and \( K \)-gaps estimates coupled with the mentioned adjustment methods. Fig. 2 corresponding to Tables 3 and 4 shows that the K-gaps estimator works better, if \( u \) is selected by the discrepancy method than by the IMT method. According to our simulation study the K-gaps estimator coupled with the IMT method demonstrates a slow convergence as the sample size increases. The IMT method requires more computation time due to a full search among pairs \((u, K)\). Generally, the K-gaps estimator works better than the intervals estimator both coupled with the discrepancy method. The intervals estimator coupled with algorithm A1 provides the RMSE similar to the discrepancy method coupled with both intervals and \( K \)-gaps estimates only for MM and ARMAX processes, see Fig. 2.

The discrepancy inequalities can be applied when the solutions of the discrepancy equation do not exist among the considered quantiles for given \( k \) and \( K \). This may slightly improve the RMSE and the absolute bias of both intervals and K-gaps estimates in comparison with the usage of the discrepancy equations, see Fig. 3. This property is due to a larger number of solutions.

Fig. 4 aims to compare the impact of the choice of \( k \). It shows that \( k = |\min(\hat{\theta}_0 L, \sqrt{L})| \) and \( k = \lfloor (\ln L)^2 \rfloor \) (both satisfy Theorem 3) provide similar values of the best RMSE. \( k = |\hat{\theta}_0 L| \) provides the best accuracy.

Fig. 5 aims to find the best measure from (4.3). Ratios \( \{RMSE(\theta_j)/\min_{i \in \{1, 2, 3\}}\ RMSE(\theta_i)\}, \ j \in \{1, 2, 3\} \) are compared. One may conclude that \( \hat{\theta}_1 \) provides consistently better accuracy than \( \hat{\theta}_2 \) and \( \hat{\theta}_3 \).

By the simulation study we recommend the \( K \)-gaps estimator coupled with the discrepancy method (4.4) with \( k = |\hat{\theta}_0 L| \) and an accurate pilot estimate \( \hat{\theta}_0 \), and the measure \( \hat{\theta}_1 \).

The impact of the heaviness of tail on the accuracy of the discrepancy method remains an open problem. Intuitively, the heaviness of the distribution tail may impact on the rate of convergence of the exceedance point process to a compound Poisson process and hence, on the convergence of the distribution of the discrepancy statistic to the limit distribution of the C-M-S statistic.
5 Application to real data

5.1 First example

Following Ferreira (2018a) we consider two data sets of the daily maximum temperatures (in 0.1 degrees Celsius) of July at Uccle (Belgium), from 1833 to 1999 and from 1900 to 1999 with sample sizes \( n \in \{5177, 3100\} \), respectively. The data are available at "http://lstat.kuleuven.be/Wiley/Data/ecad00045TX.txt". The extremal index of the smaller sample was shown to be ranged between 0.49 and 0.56 in Beirlant et al. (2004); a reduced-bias version of Nandagopalan’s runs estimator applied in Ferreira (2018a) has shown 0.41 and 0.57; and the wide range of estimators has shown 0.10 and 0.57 in Ferreira (2018a). We have analyzed the intervals and \( K \)-gaps estimators coupled with the discrepancy method based on the algorithm in Section 4. The \( K \)-gaps estimator with the IMT method and the intervals with "plateau-finding" Algorithm 1 with \( \omega = 0.3 \) were also applied here and in the next example. 'Kdis', 'K0dis' and 'Intdis' are calculated with \( k = \lfloor sL \rfloor \), where \( s \) was taken equal to the pilot intervals estimate \( \hat{\theta}_0 \) for each threshold value \( u \) or to values \( \{0.51, 0.56\} \) for \( n \in \{3100, 5177\} \), respectively, based on previous estimation of \( \theta \) and the ‘Kimt’ estimates. The discrepancy inequality method (4.4) was used. One may trust more \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) as well as 'Kdis', 'K0dis' estimates since they provide better results on the simulation. The results are shown in Table 9.

5.2 Second example

We use the data corresponding to Figure S18 in Raymond et al. (2020) and kindly provided by the authors, which represent daily-maximum dewpoint temperatures at station Dhahran, Saudi Arabia. This station is among several selected stations where a wet-bulb temperature (TW) has exceeded \( TW = 33^\circ C \) at least 5 times. The dates span from 1 Jan 1979 to 31 Dec 2017. The sample size is equal to \( n = 13866 \) due to missing observations. The estimated values of \( \theta \) are shown in Table 10.

6 Proofs

6.1 Proof of Theorem 2

Consider the conditional distribution of \( \omega_k^2(\theta) \) given \( U_{L-k,L} = t_k \). According to Lemma 2 and the condition \( \limsup_{n \to \infty} k/L < \theta \) the conditional joint distribution of the set of the order statis-
tics \( \{U_{i,L}\}_{i=L-k+1} \) asymptotically equals to the joint distribution of the set of order statistics \( \{U_{i,k}\}_{i=1}^{k} \) of a sample \( \{U_{i}\}_{i=1}^{k} \) from the uniform distribution on \([t_{k}, 1]\). Therefore, it holds
\[
\omega_{k}^{2}(\theta) = \frac{1}{(1-t_{k})^{2}} \left( \sum_{i=1}^{k} \left( U_{i,k}^{*} - t_{k} - \frac{i-0.5}{k}(1-t_{k}) \right)^{2} \right) + \frac{1}{12k}.
\]
Moreover, \( V_{i,k}^{*} = U_{i,k}^{*} - t_{k} \) are the order statistics of a sample \( \{V_{i}^{*}\} \) from the uniform distribution on \([0, 1-t_{k}]\). Hence, it follows
\[
\omega_{k}^{2}(\theta) = \frac{1}{(1-t_{k})^{2}} \left( \sum_{i=1}^{k} \left( V_{i,k}^{*} - \frac{i-0.5}{k}(1-t_{k}) \right)^{2} \right) + \frac{1}{12k}.
\]
Finally, \( W_{i,k}^{*} = V_{i,k}^{*}/(1-t_{k}) \) are the order statistics of a sample \( \{W_{i}^{*}\} \) from the uniform distribution on \([0, 1]\). Therefore, we get
\[
\omega_{k}^{2}(\theta) = \frac{d}{1} \left( \sum_{i=1}^{k} \left( W_{i,k}^{*} - \frac{i-0.5}{k} \right)^{2} \right) + \frac{1}{12k}.
\]
It is easy to see, that the last expression is the C-M-S statistic and it converges in distribution to the r.v. \( \xi \) with the cdf \( A_{1} \) independently of the value of \( t_{k} \).

### 6.2 Proof of Theorem 3

Let \( \{E_{i}^{(L)}\}_{i=1}^{[\theta L]} \) be the sequence of r.v.s satisfying condition (3.1). Let us denote \( t_{i} = 1 - \theta e^{-\theta E_{i-1,i}}, \ t_{i} = 1 - \theta e^{-\theta E_{i-1,i}}, \ a_{i} = ((L-i) - (L-k) - 0.5)/k, \ 0 \leq i \leq k. \)

Turning back to (6.2) and (6.3), we consider the following difference
\[
\tilde{\omega}_{L}^{2}(\hat{\theta}_{n}) - \omega_{k}^{2}(\theta) = \left( \tilde{\omega}_{L}^{2}(\hat{\theta}_{n}) - \omega_{k}^{2}(\hat{\theta}_{n}) \right) + \left( \omega_{k}^{2}(\hat{\theta}_{n}) - \omega_{k}^{2}(\theta) \right)
\]
\[
= \left( \tilde{\omega}_{L}^{2}(\hat{\theta}_{n}) - \omega_{k}^{2}(\hat{\theta}_{n}) \right) + \left( \omega_{k}^{2}(\hat{\theta}_{n}) - \frac{1}{12k} \right) \left( 1 - \frac{(1 - \tilde{t}_{k})^{2}}{(1-t_{k})^{2}} \right)
\]
\[
+ \left( \omega_{k}^{2}(\hat{\theta}_{n}) - \frac{1}{12k} \right) \frac{(1 - \tilde{t}_{k})^{2}}{(1-t_{k})^{2}} - \left( \omega_{k}^{2}(\theta) - \frac{1}{12k} \right). \tag{6.1}
\]

The third term on the right-hand side of (6.1) is equal to
\[
\frac{1}{(1-t_{k})^{2}} \left( \sum_{i=0}^{k-1} (t_{i} - \tilde{t}_{k} - a_{i}(1-\tilde{t}_{k}))^{2} - \sum_{i=0}^{k-1} (t_{i} - t_{k} - a_{i}(1-t_{k}))^{2} \right).
\]
Using the relation \( x^2 - y^2 = (y - x)^2 - 2y(y - x) \), we obtain

\[
\frac{1}{(1 - t_k)^2} \left( \sum_{i=0}^{k-1} (t_i - \hat{t}_k - a_i(1 - \hat{t}_k))^2 - \sum_{i=0}^{k-1} (t_i - t_k - a_i(1 - t_k))^2 \right)
\]

\[
= \frac{1}{(1 - t_k)^2} \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k) (1 - a_i))^2
\]

\[
- \frac{2}{(1 - t_k)^2} \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k) (1 - a_i)) (t_i - t_k - a_i(1 - t_k)) = \frac{e^{\theta} - 1}{1 - t_k}.
\]

Thereby, we can rewrite

\[
\omega_k^2(\hat{\theta}_n) - \omega_k^2(\theta) = \left( \omega_k^2(\hat{\theta}_n) - \frac{1}{12k} \right) \left( 1 - \frac{1 - \hat{t}_k)^2}{(1 - t_k)^2} \right)
\]

\[
+ \frac{1}{(1 - t_k)^2} \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k) (1 - a_i))^2
\]

\[
- \frac{2}{(1 - t_k)^2} \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k) (1 - a_i)) (t_i - t_k - a_i(1 - t_k)).
\]

(6.2)

Let us find the asymptotics of the difference \( t_i - \hat{t}_i = \theta e^{-\theta E_{l-i,i}^{(L)}} - \hat{\theta}_n e^{-\hat{\theta}_n E_{l-i,i}^{(L)}} \), \( 0 \leq i \leq k \). We have

\[
\hat{\theta}_n e^{-\hat{\theta}_n E_{l-i,i}^{(L)}} - \theta e^{-\theta E_{l-i,i}^{(L)}} = (\hat{\theta}_n - \theta) e^{-\hat{\theta}_n E_{l-i,i}^{(L)}}
\]

\[
+ \theta e^{-\theta E_{l-i,i}^{(L)}} (e^{-\theta E_{l-i,i}^{(L)}}(\hat{\theta}_n - \theta) - 1).
\]

(6.3)

By (5.5) it holds

\[
\hat{\theta}_n - \theta = O_P \left( \frac{1}{\sqrt{m_n}} \right).
\]

(6.4)

We have \( U_{i+1,l} \overset{d}{=} e^{-E_{l-i,i}^{(L)}} \theta \leq e^{-E_{l-k,i}^{(L)}} \theta = U_{k+1,l}, \ 0 \leq i < k < l \), where \( \{U_{i,l}\} \) are the order statistics arising from a standard uniform distribution. By Lemma we get

\[
\frac{U_{k,l} - \frac{k-1}{l-1}}{\sqrt{\frac{(k-1)(l-k)}{l(l-1)}}} \overset{d}{\to} N(0, 1)
\]

as \( k \to \infty, L \to \infty, L - k \to \infty \). Since both

\[
\frac{k-1}{l-1} - \frac{k}{l} = \frac{-l-k}{l(l-1)} \quad \text{and} \quad \frac{l-k}{l(l-1)} \sqrt{\frac{(l-1)^3}{(l-k)(l-1)}} = \sqrt{\frac{(l-1)(l-k)}{(k-1)l^2}}
\]

23
are equivalent to \( o(1) \), then under the same conditions as in Lemma 1 and using Slutsky’s theorem, we obtain
\[
\frac{U_{k,l} - \frac{k}{l}}{\sqrt{k(l-k)}} \xrightarrow{d} N(0, 1).
\]
This result implies
\[
1 - t_k = \theta e^{-E_{l-k,t}^{(L)}} = \frac{k}{L}(1 + o_P(1)),
\]
and \( E_{l-i,l}^{(L)} = \ln(l/i)/\theta(1 + o_P(1)) \) as \( i \leq k, i \to \infty \) holds. Using the condition \((\ln L)^2 = o(m_n)\), we have \( E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta) = o_P(1) \) and
\[
e^{-E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta)} - 1 = -E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta)(1 + o_P(1)).
\]
Then the first term on the right-hand side in (6.5) is asymptotically smaller than the second one. Hence, from (6.3), (6.5) and (6.6) we obtain
\[
\hat{t}_i - t_i = -\frac{i}{L} E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta)(1 + o(1)) = O_P\left( \frac{i \ln(i/L)}{L \sqrt{m_n}} \right).
\]
Therefore, the asymptotics of the expression \( t_i - \hat{t}_i - (t_k - \hat{t}_k)(1 - a_i) \) is the following
\[
\frac{i}{L} E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta)(1 + o_P(1)) = \frac{i}{L} E_{l-k,l}^{(L)}(\hat{\theta}_n - \theta)(1 + o_P(1))
\]
\[
= -\frac{i}{L} E_{l-k,l}^{(L)} - E_{l-i,l}^{(L)}(\hat{\theta}_n - \theta)(1 + o_P(1)) = O_P\left( \frac{i \ln(k/i)}{L \sqrt{m_n}} \right),
\]
since (6.4) and
\[
E_{l-i,l}^{(L)} - E_{l-k,l}^{(L)} \xrightarrow{d} \frac{\ln(k/i)}{\theta}(1 + o_P(1))
\]
hold. Note that the maximum of the function \( f(x) = \ln(a/x)x \), where \( a \) is a positive constant, is achieved in the point \( x_0 = a/e \), so \( i \ln(k/i) \leq k/e \), \( 1 \leq i \leq k \). Hence, from (6.5) and (5.1) the asymptotics of the second term on the right-hand side of (6.2) is given by
\[
\frac{1}{(1 - t_k)^2} \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k)(1 - a_i))^2 = O_P\left( \frac{k}{m_n} \right) = o_P(1)
\]
24
due to (3.6). Now we estimate the asymptotics of the third summand on the right-hand side of (6.2). An appeal to (6.7) and the Cauchy-Schwarz inequality gives us the following

\[
\frac{2}{(1-t_k)^2} \left| \sum_{i=0}^{k-1} (t_i - \hat{t}_i - (t_k - \hat{t}_k) (1 - a_i)) (t_i - t_k - a_i(1 - t_k)) \right| \\
\leq \max_i \left( \frac{i}{L} (E_{l-i,i}^{(L)} - E_{l-i,k}^{(L)}) |\hat{\theta}_n - \theta| \right) \cdot \frac{2}{(1-t_k)^2} \\
\cdot \sum_{i=0}^{k-1} \left| 1 - \theta \exp(-E_{l-i,i}^{(L)} - t_k - \frac{k-i-0.5}{k}(1-t_k)) \right| (1 + o_P(1)) \\
\leq O_P \left( \frac{k}{L \sqrt{m_n}} \right) \frac{2\sqrt{k}}{(1-t_k)^2} \\
\cdot \sqrt{\sum_{i=0}^{k-1} \left( 1 - \theta \exp(-E_{l-i,i}^{(L)} - t_k - \frac{k-i-0.5}{k}(1-t_k)) \right)^2} \\
= O_P \left( \frac{k^{3/2}}{L \sqrt{m_n}} \right) \frac{2}{(1-t_k)^2} \sqrt{\omega_k^2(\theta) - \frac{1}{12k}} = O_P \left( \frac{k^{3/2}}{L \sqrt{m_n}} \cdot \frac{L}{k} \right) \\
= O_P \left( \frac{\sqrt{k}}{\sqrt{m_n}} \right) = o_P(1),
\]

where the last two strings follow from (3.6) and (6.5) and since \( \omega_k^2(\theta) - \frac{1}{12k} = O_P(1) \) holds from Theorem 2.

Thus, by (6.1) the sum of the second and the third terms in (6.2) is equal to

\[
\left( \omega_k^2(\hat{\theta}_n) - \frac{1}{12k} \right) \frac{(1 - \hat{t}_k)^2}{(1-t_k)^2} - \left( \omega_k^2(\theta) - \frac{1}{12k} \right) = o_P(1). \quad (6.9)
\]

Now we derive that the asymptotic of the first term on the right-hand side of (6.2) is \( O_P\left( \frac{\ln L}{\sqrt{m_n}} \right) \).

Let us prove the following

\[
\frac{(1 - \hat{t}_k)^2}{(1-t_k)^2} = 1 = o_P(1).
\]

Using (6.3), (6.5) and (6.7), we obtain

\[
\frac{(1 - \hat{t}_k)^2}{(1-t_k)^2} - 1 = \frac{(\hat{\theta}_n e^{-\hat{\theta}_n E_{l-k,l}^{(L)}})^2}{(\theta e^{-\theta E_{l-k,l}^{(L)}})^2} - 1 = \left( \hat{\theta}_n e^{-\hat{\theta}_n E_{l-k,l}^{(L)}} - \theta e^{-\theta E_{l-k,l}^{(L)}} \right) \\
\cdot \left( \hat{\theta}_n e^{-\hat{\theta}_n E_{l-k,l}^{(L)}} - \theta e^{-\theta E_{l-k,l}^{(L)}} + 2\theta e^{-\theta E_{l-k,l}^{(L)}}/(\theta e^{-\theta E_{l-k,l}^{(L)}})^2 \right) \\
= O_P \left( \frac{k \ln L}{L \sqrt{m_n}} \right) \left( O_P \left( \frac{k \ln L}{L \sqrt{m_n}} \right) + O_P \left( \frac{k}{L} \right) \right) = O_P \left( \frac{\ln L}{\sqrt{m_n}} \right) = o_P(1). \quad (6.10)
\]
It remains to show that $\omega^2_k(\hat{\theta}_n) - \frac{1}{12k} = O_P(1)$. It follows from (6.10) that $\frac{(1-t_k)^2}{(1-t_k)^2} = O_P(1)$. Dividing the expression (6.9) by $\frac{(1-t_k)^2}{(1-t_k)^2}$, we obtain again the expression, that is equal to $o_P(1)$. Namely, we get
\[
\left( \omega^2_k(\hat{\theta}_n) - \frac{1}{12k} \right) - \left( \omega^2_k(\theta) - \frac{1}{12k} \right) \frac{(1-t_k)^2}{(1-t_k)^2} = \left( \omega^2_k(\hat{\theta}_n) - \frac{1}{12k} \right) - \left( \omega^2_k(\theta) - \frac{1}{12k} \right) = o_P(1).
\]
(6.11)

Using (6.10), we obtain that the second term on the left-hand side of (6.11) is $o_P(1)$, hence
\[
\omega^2_k(\hat{\theta}_n) - \frac{1}{12k} = O_P \left( \omega^2_k(\theta) - \frac{1}{12k} \right) = O_P(1)
\]
holds. Therefore, the first term in (6.2) is $o_P(1)$.

It remains to prove, that the first term in (6.1) is $o_P(1)$. We have
\[
\tilde{\omega}^2_L(\hat{\theta}_n) - \omega^2_k(\hat{\theta}_n) = \sum_{i=0}^{k-1} \left( \frac{i+0.5}{k} - y_i \right)^2 - \sum_{i=0}^{k-1} \left( \frac{i+0.5}{k} - e_i \right)^2
\]
\[
= \sum_{i=0}^{k-1} \left( \frac{2i+1}{k} - e_i - y_i \right) (e_i - y_i),
\]
where $e_i = e_i(\hat{\theta}_n) = \exp \left( -\hat{\theta}_n(E_{l-i,l}^{(L)} - E_{l-k,l}^{(L)}) \right)$ and $y_i = y_i(\hat{\theta}_n) = \exp \left( -\hat{\theta}_n(Y_{l-i,l} - Y_{l-k,l}) \right), \ i \in \{0, \ldots, k-1\}$. It easily follows from (3.4), (3.6) and (6.4), that for all $i$, $0 \leq i \leq k-1$,
\[
e_i - y_i = \exp \left( -\hat{\theta}_n(E_{l-i,l}^{(L)} - E_{l-k,l}^{(L)}) \right) - \exp \left( -\hat{\theta}_n(E_{l-i,l}^{(L)} - E_{l-k,l}^{(L)} + o_P(1/\sqrt{k})) \right) = o_P(1/\sqrt{k}).
\]

Further, from the latter, (3.6), (6.4) and (6.8) we obtain for all $i$, $0 \leq i \leq k-1$,
\[
2\frac{i+0.5}{k} - e_i - y_i = o_P(1/\sqrt{k}).
\]

Thus, we derive
\[
\tilde{\omega}^2_L(\hat{\theta}_n) - \omega^2_k(\hat{\theta}_n) = \sum_{i=0}^{k-1} o_P(1/k) = o_P(1),
\]
(6.12)
the required result.

### 6.3 Proof of Theorem 4

Note that formula (6.12) in the proof of Theorem 3
\[
\tilde{\omega}^2_L(\theta) - \omega^2_k(\theta) = o_P(1)
\]
(6.13)
is valid after the replacement \( \hat{\theta}_n \) by \( \theta \). Simplifying (3.2), we obtain

\[
\bar{\omega}^2_L(\theta) = \sum_{i=0}^{k-1} \left( \exp \left( -\theta (Y_{L-i,L} - Y_{L-k,L}) \right) - i + 0.5 \right)^2 + \frac{1}{12k}.
\]

Let us consider the difference \( \bar{\omega}^2_L(\hat{\theta}_n) - \bar{\omega}^2_L(\theta) \). We have

\[
\bar{\omega}^2_L(\hat{\theta}_n) - \bar{\omega}^2_L(\theta) = \sum_{i=0}^{k-1} y_i (d_i - 1) \left( y_i (d_i + 1) - 2 \frac{i + 0.5}{k} \right),
\]

where \( y_i = y_i(\theta) \) is taken as in the proof of Theorem 3 and \( d_i = \exp(- (\hat{\theta}_n - \theta)(Y_{L-i,L} - Y_{L-k,L})) \). It follows from (3.4) and (6.8), that

\[
y_i - i + 0.5 \frac{1}{k} = y_i + o_P \left( \frac{1}{\sqrt{k}} \right) = O_P \left( \frac{1}{\sqrt{k}} \right)
\]

and

\[
y_i d_i - i + 0.5 \frac{1}{k} = y_i (d_i - 1) + O_P \left( \frac{1}{\sqrt{k}} \right) = \left( \frac{i}{k} + O_P \left( \frac{1}{\sqrt{k}} \right) \right) (d_i - 1) + O_P \left( \frac{1}{\sqrt{k}} \right).
\]

Thus, the latter two equations imply

\[
\bar{\omega}^2_L(\hat{\theta}_n) - \bar{\omega}^2_L(\theta) = \sum_{i=0}^{k_{n_\alpha}-1} i^2 (d_i - 1)^2 \left( 1 + O_P \left( \frac{1}{\sqrt{k_{n_\alpha}}} \right) \right)
\]

under the condition that \( \sqrt{k_{n_\alpha}} |d_i - 1| \rightarrow \infty \) as \( n \rightarrow \infty \) holds. Indeed, in terms of the subsequence \( \{k_{n_\alpha}\}_{s > 1} \), from (3.4) and (6.8) it follows

\[
d_i - 1 = \exp\{-(\hat{\theta}_{n_\alpha} - \theta)(Y_{L-s,L} - Y_{L-k_{n_\alpha},L})\} - 1 = \frac{\ln(k_{n_\alpha}/i)}{\theta} O_P(\hat{\theta}_{n_\alpha} - \theta)
\]

for \( \alpha > 0 \) and \( d_i - 1 = \Omega_P(1) \) for \( \alpha = 0 \). Here, \( \xi_n := \Omega_P(\eta_n) \) means that \( |\xi_n/\eta_n| \xrightarrow{P} \infty \) holds as \( n \rightarrow \infty \). From the latter, (6.13), (6.14) and Theorem 2 we finally obtain

\[
\bar{\omega}^2_L(\hat{\theta}_{n_\alpha}) - \bar{\omega}^2_L(\theta) = k_{n_\alpha} O_P(\hat{\theta}_{n_\alpha} - \theta)^2
\]

and

\[
\bar{\omega}^2_L(\hat{\theta}_{n_\alpha}) = \left( \bar{\omega}^2_L(\hat{\theta}_n) - \bar{\omega}^2_L(\theta) \right) + (\bar{\omega}^2_L(\theta) - \omega^2_{k_{n_\alpha}}(\theta)) + \omega^2_{k_{n_\alpha}}(\theta)
\]

\[
= \Omega_P(k_{n_\alpha}^{1-2\alpha}) + o_P(1) = \Omega_P(k_{n_\alpha}^{1-2\alpha}),
\]

the required result.
7 Acknowledgements

The authors were partly supported by the Russian Foundation for Basic Research (grant No. 19-01-00090).

References

[1] Ancona–Navarrete, M.A., Tawn, J. A. (2000). A comparison of Methods for Estimating the Extremal Index. *Extremes* 3:1 5–38.

[2] Balakrishnan, N., Rao, C. R., eds. (1998). *Handbook of Statistics* 16. Elsevier Science B.V.

[3] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, Chichester, West Sussex.

[4] Berghaus, B., Bücher, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *The Annals of Statistics* 46(5) 2307–2335.

[5] Bolshev, L.N., Smirnov, N.V. (1965). *Tables of Mathematical Statistics*, Nauka, Moscow (in Russian)

[6] Chernick, M.R., Hsing, T., McCormick, W.P. (1991). Calculating the extremal index for a class of stationary sequences. *Advances in Applied Probability* 23 835–850.

[7] Drees, H. (2011). Bias correction for estimators of the extremal index. Preprint, arXiv: 1107.0935.

[8] de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer.

[9] Ferreira, M. (2018a). Heuristic tools for the estimation of the extremal index: a comparison of methods. *REVSTAT – Statistical Journal* 16:1 115–136.

[10] Ferreira, M. (2018b). Analysis of estimation methods for the extremal index. *Electronic Journal of Applied Statistical Analysis* 11:1 296–306.

[11] Ferro, C.A.T., Segers, J. (2003). Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society Series B*. 65 545–556.
[12] Fukutome, S., Liniger, M.A., Süveges, M. (2015). Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland. *Theoretical and Applied Climatology* **120** 403–416.

[13] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis* **32** 177–203.

[14] Hsing, T., Huesler, J., Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* **78** 97–112.

[15] Kobzar, A.I. (2006). *Applied mathematical statistics for engineers and scientists*. Fizmatlit, Moscow (In Russian).

[16] Laurini, F., Tawn, J.A. (2012). The extremal index for GARCH(1,1) processes. *Extremes* **15** 511–529.

[17] Leadbetter, M.R., Lingren, G., Rootzén, H. (1983). *Extremes and Related Properties of Random Sequence and Processes*. ch.3, Springer, New York.

[18] Markovich, N.M. (1989). Experimental analysis of nonparametric probability density estimates and of methods for smoothing them. *Automation and Remote Control* **50** 941–948.

[19] Markovich, N.M. (2007). *Nonparametric Analysis of Univariate Heavy–Tailed data: Research and Practice*. Wiley, Chichester, West Sussex.

[20] Markovich, N.M. (2014). Modeling clusters of extreme values. *Extremes* **17:1** 97–125.

[21] Markovich, N.M. (2015). Nonparametric estimation of extremal index using discrepancy method. In: Proceedings of the X International conference “System identification and control problems” SICPRO–2015 Moscow January 26–29, V.A. Trapeznikov Institute of Control Sciences. 160–168. ISBN 978-5-91450-162-1

[22] Markovich, N.M. (2016). Erratum to: modeling clusters of extreme values. *Extremes* **19:1** 139–142.

[23] Markovich, N.M. (2017). Clusters of extremes: modeling and examples. *Extremes* **20** 519–538.

[24] Martynov, G. V. (1978). *The omega square tests*. Nauka, Moscow (In Russian).
[25] Northrop, P.J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* 18:4 585–603.

[26] Ramsay, C.M. (2008). The Distribution of Sums of I.I.D. Pareto Random Variables with Arbitrary Shape Parameter. *Communications in Statistics - Theory and Methods* 37:14 2177–2184.

[27] Raymond, C., Matthews, T., Horton, R.M. (2020). The emergence of heat and humidity too severe for human tolerance. *Science Advances* 6(19) eaaw1838

[28] Robert, C.Y. (2009a). Asymptotic distributions for the intervals estimators of the extremal index and the cluster–size probabilities. *Journal of Statistics Planning and Inference* 139 3288–3309.

[29] Robert, C.Y. (2009b). Inference for the limiting cluster size distribution of extreme values. *The Annals of Statistics* 37 271–310.

[30] Robert, C.Y., Segers, J., Ferro, C.A.T. (2009). A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics* 3 993–1020.

[31] Smirnov, N.V. (1937). *On the $\omega^2$-distribution of von Mises*. Matematicheskij Sbornik. 2:5 973–993. (In Russian) (French abstract)

[32] Smirnov, N.V. (1949). Limit distributions for the terms of a variational series. In Russian: Trudy Matematicheskogo Instituta imeni V.A. Steklova. 25 Translation: *American Mathematical Society Translations* (1952). 11 82–143.

[33] Sun, J., Samorodnitsky, G. (2010). Estimating the extremal index, or, can one avoid the threshold-selection difficulty in extremal inference? Technical Report, Cornell University.

[34] Sun, J., Samorodnitsky, G. (2019). Multiple thresholds in extremal parameter estimation. *Extremes* 22 317–341.

[35] Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes* 10 41–55.

[36] Süveges, M., Davison, A.C. (2010). Model misspecification in peaks over threshold analysis. *The Annals of Applied Statistics* 4:1 203–221.

[37] Vapnik, V. N., Markovich, N.M. and Stefanyuk, A.R. (1992). Rate of convergence in $L_2$ of the projection estimator of the distribution density. *Automation and Remote Control* 53 677–686.
[38] Weissman, I., Novak, S.Yu. (1978). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference* 66 281–288.
Table 1: The root mean squared error of $\hat{\theta}_0^k (k = |sL|)$, $\hat{\theta}_0$ is a pilot intervals estimate.

| $\text{RMSE}$ | $10^{4}$/$\theta$ | $\text{ARMAX}$ | $\text{ARu}^+$ | $\text{ARu}^-$ | $\text{MA}(2)$ | $\text{ARc}$ | $\text{AR}(2)$ | $\text{GARCH}$ |
|---------------|------------------|----------------|----------------|----------------|----------------|------------|------------|--------------|
| $s = \theta$  | $\hat{\theta}_0^1$ | 9.018  11       | 8.879  10       | 9.489  19      | 15  19         | 12  10     | 5.709     | 9.104   16    |
| $s = \theta$  | $\hat{\theta}_0^2$ | 8.701  11       | 8.741  9.985    | 9.042  19      | 14  19         | 12  10     | 5.709     | 9.081   16    |
| $s = \theta$  | $\hat{\theta}_0^3$ | 9.866  12       | 9.264  11       | 10  19         | 15  19         | 12  10     | 5.709     | 9.141   16    |
| $s = \theta$  | $\hat{\theta}_0^4$ | 3.419  5.357    | 2.438  5.016    | 4.292  9.341   | 7.617  7.434   | 5.507  9.527 | 0.605     | 4.696   19    |
| $s = \theta$  | $\hat{\theta}_0^5$ | 1.244  1.959    | 1.073  1.859    | 1.681  5.412   | 3.712  2.938   | 2.643  9.569 | 0.446     | 2.667   17    |
| $s = \theta$  | $\hat{\theta}_0^6$ | 14  17          | 14  16          | 14  16         | 16  19         | 14  10     | 5.709     | 11  22 |
| $s = \theta$  | $\hat{\theta}_0^1$ | 161  213        | 154  196        | 251  893       | 304  386       | 196  376   | -  357     | -        |
| $s = \theta$  | $\hat{\theta}_0^2$ | 159  213        | 150  195        | 249  894       | 304  387       | 198  377   | -  357     | -        |
| $s = \theta$  | $\hat{\theta}_0^3$ | 168  216        | 165  202        | 256  893       | 303  386       | 196  376   | -  357     | -        |
| $s = \theta$  | $\hat{\theta}_0^4$ | 106  159        | 89  150         | 299  931       | 525  329       | 229  540   | 14  401  413 |
| $s = \theta$  | $\hat{\theta}_0^5$ | 126  157        | 103  145        | 415  1072      | 690  361       | 324  734   | 156  424  404 |
| $s = \theta$  | $\hat{\theta}_0^6$ | 348  443        | 293  445        | 419  989       | 511  350       | 335  463   | 86  446  474 |

$n = 10^5$

| $s = \theta$  | $\hat{\theta}_0^1$ | 152  147        | 138  151        | 162  189       | 199  211       | 173  254   | -  199  106 |
| $s = \theta$  | $\hat{\theta}_0^2$ | 134  134        | 125  133        | 146  177       | 186  191       | 155  244   | -  186  106 |
| $s = \theta$  | $\hat{\theta}_0^3$ | 214  180        | 181  202        | 207  210       | 232  251       | 216  274   | -  231  109 |
| $s = \theta$  | $\hat{\theta}_0^4$ | 60  59          | 59  51          | 61  60         | 63  79         | 65  82    | 65  66  85 |
| $s = \theta$  | $\hat{\theta}_0^5$ | 13  16          | 12  15          | 16  16         | 18  22         | 16  27    | 0.87683  19  44 |
| $s = \theta$  | $\hat{\theta}_0^6$ | 441  334        | 480  423        | 446  323       | 417  400       | 447  428   | 815  437  236 |

$n = 5000$

$\hat{\theta}_0^1$ to $\hat{\theta}_0^6$ follow similar patterns as $\hat{\theta}_0^1$ to $\hat{\theta}_0^6$ with slight variations in the numbers.

32
Table 2: The absolute bias of $\hat{\theta}_K^0$ ($k = |sL|$), $\hat{\theta}_0$ is a pilot intervals estimate.

| Bias | $10^4/\rho$ | MM | ARMAX | $AR_u^+$ | $AR_u^-$ | $MA(2)$ | ARc | AR(2) | GARCH |
|------|-------------|-----|--------|----------|----------|---------|-----|-------|-------|
| $s = \theta$ | $\hat{\theta}_K^{\text{dis}}$ | 5.824 | 5.427 | 6.299 | 2.789 | 0.859 | 9.880 | 5.880 | $\hat{\theta}_K^{\text{dis}}$ |
| $s = \hat{\theta}_0$ | 16 | 12 | 18 | 28 | 76 | 57 | $\hat{\theta}_K^{\text{dis}}$ |
| $n = 10^5$ | | | | | | | | |
| | | | | | | | | |
| $s = \theta$ | $\hat{\theta}_K^{\text{dis}}$ | 101 | 75 | 142 | 56 | 8.9597 | 405 | 436 | $\hat{\theta}_K^{\text{dis}}$ |
| $s = \hat{\theta}_0$ | 111 | 97 | 145 | 58 | 3.3883 | 410 | 202 | $\hat{\theta}_K^{\text{dis}}$ |
| $n = 5000$ | | | | | | | | |
| $s = \theta$ | $\hat{\theta}_K^{\text{dis}}$ | 365 | 365 | 365 | 365 | 1.19 | 1.19 | 1.19 | $\hat{\theta}_K^{\text{dis}}$ |
| $s = \hat{\theta}_0$ | 119 | 81 | 81 | 81 | 49 | 49 | 49 | $\hat{\theta}_K^{\text{dis}}$ |
| $n = 100,000$ | | | | | | | | |


Table 3: The root mean squared error \((k = |\hat{\theta}_0|L|)\), \(\hat{\theta}_0\) is a pilot intervals estimate.

| \(\text{RMSE} \times 10^{-1/\theta}\) | MM | ARMAX | \(ARu^+\) | \(ARu^-\) | \(MA(2)\) | \(ARc\) | AR(2) | \(GARCH\) |
|--------------------------------------|----|--------|----------|----------|----------|--------|--------|--------|
| \(n = 10^5\)                        |    |        |          |          |          |        |        |        |
| \(\theta_1\)                        | 147| 215   | 159     | 211     | 230     | 887    | 287    | 383    | 199    | \(\hat{\theta}_0\) | 400  | - 305 | -        |
| \(\hat{\theta}_2\)                  | 146| 213   | 158     | 211     | 230     | 889    | 287    | 384    | 201    | 402    | - 305 | -        |
| \(\hat{\theta}_3\)                  | 154| 222   | 164     | 215     | 235     | 886    | 287    | 383    | 203    | 400    | - 305 | -        |
| \(\hat{\theta}_1^*\)                | 103| 160   | 89      | 151     | 292     | 928    | 516    | 328    | 229    | 542    | 17    | 400  413  |
| \(\hat{\theta}_2^*\)                | 123| 156   | 99      | 151     | 410     | 1070   | 690    | 354    | 324    | 745    | 155   | 420  405  |
| \(\hat{\theta}_3^*\)                | 354| 442   | 294     | 425     | 413     | 957    | 516    | 351    | 336    | 467    | 67    | 443  470  |
| \(\hat{\theta}_1^{K_{dis}}\)        | 133| 204   | 145     | 190     | 195     | 772    | 209    | 349    | 141    | 423    | - 390 | 375    |
| \(\hat{\theta}_2^{K_{dis}}\)        | 777| 844   | 648     | 940     | 799     | 777    | 208    | 1396   | 799    | 420    | - 519 | 382    |
| \(\hat{\theta}_3^{K_{dis}}\)        | 895| 702   | 651     | 821     | 398     | 763    | 207    | 1115   | 927    | 637    | - 477 | 389    |
| \(\hat{\theta}_1^{K_{dis^2}}\)      | 136| 264   | 100     | 237     | 150     | 631    | 220    | 105    | 229    | 390    | 12    | 394   464  |
| \(\hat{\theta}_2^{K_{dis^2}}\)      | 127| 226   | 70      | 207     | 249     | 788    | 518    | 940    | 152    | 1944   | 34    | 1231 3929 |
| \(\hat{\theta}_3^{K_{dis^2}}\)      | 652| 897   | 238     | 1013    | 690     | 1300   | 963    | 345    | 798    | 958    | 75    | 353   573  |
| \(\hat{\theta}_{K_{int}}\)          | 217| 569   | 69      | 498     | 173     | 844    | 2501   | 401    | 309    | 466    | 33    | 3630 4028 |
| \(\hat{\theta}_{IA1}\)              | 116| 122   | 95      | 113     | 447     | 1193   | 1756   | 399    | 387    | 977    | 233   | 693   580  |
| \(n = 5000\)                         |    |        |          |          |          |        |        |        |        |        |        |        |
| \(\theta_1\)                        | 565| 938   | 506     | 818     | 783     | 1431   | 1364   | 394    | 533    | 816    | - 900 | 1497  |
| \(\hat{\theta}_2\)                  | 557| 913   | 476     | 787     | 748     | 1401   | 1337   | 396    | 537    | 810    | - 897 | 1497  |
| \(\hat{\theta}_3\)                  | 633| 1013  | 620     | 903     | 879     | 1490   | 1416   | 394    | 593    | 850    | - 910 | 1497  |
| \(\hat{\theta}_1^*\)                | 359| 496   | 350     | 464     | 715     | 1294   | 1276   | 315    | 352    | 760    | 606   | 835   955  |
| \(\hat{\theta}_2^*\)                | 352| 466   | 291     | 450     | 808     | 1587   | 1666   | 395    | 557    | 1179   | 422   | 724   870  |
| \(\hat{\theta}_1^{K_{dis}}\)        | 1635| 1564  | 1377    | 1652    | 1902    | 1644   | 1754   | 713    | 1505   | 1656   | 1455  | 1610  2036 |
| \(\hat{\theta}_2^{K_{dis}}\)        | 480| 917   | 496     | 772     | 787     | 1186   | 1820   | 427    | 406    | 807    | - 1690| 1877  |
| \(\hat{\theta}_3^{K_{dis}}\)        | 1525| 1880  | 982     | 1793    | 1836    | 1863   | 2672   | 2981   | 1218   | 2020   | - 1855| 2929  |
| \(\hat{\theta}_1^{K_{dis^2}}\)      | 1624| 1993  | 1286    | 1815    | 1692    | 1686   | 2348   | 1775   | 1924   | 2349   | - 1737| 2337  |
| \(\hat{\theta}_2^{K_{dis^2}}\)      | 320| 605   | 299     | 507     | \textbf{453} | \textbf{592} | 641    | \textbf{213} | 404 | 754 | 72 | 1528 | 1491 |
| \(\hat{\theta}_3^{K_{dis^2}}\)      | \textbf{252} | 548     | \textbf{199} | 487     | 535     | \textbf{866} | 2529   | 423    | \textbf{335} | \textbf{488} | \textbf{25} | 3684 | 3787  |
| \(\hat{\theta}_{K_{int}}\)          | 824| 1007  | 931     | 871     | 1086    | 927    | 916    | 555    | 714    | 929    | 2321 | 1106 1324 |
| \(\hat{\theta}_{IA1}\)              | 247| 588   | \textbf{188} | 525     | \textbf{293} | \textbf{869} | 2518   | 418    | \textbf{325} | \textbf{474} | \textbf{25} | 3680 | 3900  |
| \(\hat{\theta}_{IA1}\)              | 385| 513   | 319     | 478     | 694     | 1388   | 1985   | 394    | 514    | 1114   | 676   | 980   1077 |
Table 4: The absolute bias ($k = |\hat{\theta}_0 L|$), $\hat{\theta}_0$ is a pilot intervals estimate.

| $10^4 \theta$ | MM     | ARMAX  | $ARu^+$ | $ARu^-$ | $MA(2)$ | ARc   | AR(2)  | GARCH |
|-------------|---------|---------|---------|---------|---------|-------|---------|-------|
| 0.5         | 0.8     | 0.25    | 0.75    | 0.5     | 0.8     | 0.75  | 0.96    | 0.5   | 2/3    | 0.3   | 0.25  | 0.447 |

$n = 10^5$

| $\hat{\theta}_1$ | 18     | 2.515  | 20     | 21     | 142    | 872   | 203    | 380   | 132   | 333   | -     | 217   | -     |
| $\hat{\theta}_2$ | 16     | 2.498  | 16     | 20     | 142    | 874   | 203    | 381   | 134   | 335   | -     | 217   | -     |
| $\hat{\theta}_3$ | 20     | 8.680  | 24     | 21     | 141    | 870   | 203    | 379   | 131   | 331   | -     | 217   | -     |
| $\hat{\theta}_1^*$ | 29     | 6.723  | 28     | 14     | 259    | 915   | 469    | 313   | 200   | 513   | 17    | 354   | 184   |
| $\hat{\theta}_2^*$ | 73     | 43     | 72     | 55     | 396    | 1064  | 669    | 336   | 300   | 719   | 155   | 386   | 198   |
| $\hat{\theta}_3^*$ | 11     | 42     | 66     | 63     | 132    | 858   | 254    | 253   | 62    | 261   | 67    | 320   | 170   |
| $\hat{\theta}_1^{K\text{disf}}$ | 24     | 66     | 18     | 43     | 29     | 757   | 48     | 341   | 60    | 383   | -     | 323   | 14    |
| $\hat{\theta}_2^{K\text{disf}}$ | 133    | 138    | 172    | 145    | 93     | 763   | 44     | 175   | 161   | 380   | -     | 284   | 79    |
| $\hat{\theta}_3^{K\text{disf}}$ | 188    | 133    | 181    | 135    | 1.3951 | 747   | 50     | 185   | 234   | 423   | -     | 271   | 92    |
| $\hat{\theta}_1^{K\text{disf}}^*$ | 103    | 212    | 40     | 194    | 30     | 620   | 43     | 90    | 217   | 368   | 12    | 364   | 346   |
| $\hat{\theta}_2^{K\text{disf}}^*$ | 115    | 204    | 51     | 190    | 56     | 786   | 3497   | 404   | 137   | 448   | 34    | 1085  | 3020  |
| $\hat{\theta}_3^{K\text{disf}}^*$ | 225    | 488    | 34     | 485    | 141    | 136   | 306    | 321   | 359   | 636   | 75    | 222   | 249   |

$n = 5000$

| $\hat{\theta}_1$ | 144    | 336    | 100    | 255    | 339    | 1284  | 851    | 373   | 46    | 242   | -     | 345   | 11    |
| $\hat{\theta}_2$ | 103    | 290    | 54     | 205    | 317    | 1245  | 811    | 376   | 8.581 | 265   | -     | 345   | 11    |
| $\hat{\theta}_3$ | 185    | 391    | 148    | 305    | 360    | 1322  | 891    | 370   | 81    | 221   | -     | 344   | 11    |
| $\hat{\theta}_1^*$ | 99     | 142    | 109    | 112    | 560    | 1236  | 1150   | 272   | 170   | 631   | 606   | 665   | 382   |
| $\hat{\theta}_2^*$ | 82     | 14     | 51     | 35     | 681    | 1534  | 1574   | 385   | 450   | 1094  | 422   | 636   | 378   |
| $\hat{\theta}_3^*$ | 812    | 779    | 719    | 754    | 992    | 1192  | 1021   | 114   | 735   | 658   | 1455  | 1047  | 842   |
| $\hat{\theta}_1^{K\text{disf}}$ | 16     | 218    | 58     | 98     | 115    | 1048  | 1260   | 356   | 41    | 109   | -     | 1186  | 1254  |
| $\hat{\theta}_2^{K\text{disf}}$ | 437    | 158    | 290    | 269    | 454    | 727   | 926    | 421   | 305   | 332   | -     | 853   | 379   |
| $\hat{\theta}_3^{K\text{disf}}$ | 467    | 165    | 467    | 238    | 341    | 838   | 858    | 80    | 715   | 564   | -     | 635   | 154   |
| $\hat{\theta}_1^{K\text{disf}}^*$ | 169    | 343    | 34     | 309    | 108    | 480   | 467    | 135   | 343   | 680   | 72    | 1458  | 1342  |
| $\hat{\theta}_2^{K\text{disf}}^*$ | 192    | 492    | 43     | 443    | 145    | 852   | 2529   | 423   | 290   | 429   | 25    | 3320  | 3781  |
| $\hat{\theta}_3^{K\text{disf}}^*$ | 129    | 146    | 295    | 101    | 304    | 534   | 157    | 110   | 25    | 306   | 2321  | 582   | 241   |

| $\hat{\theta}_1^{K\text{kimf}}$ | 2.636  | 4.447  | 3.448  | 3.598  | 5.124  | 1.988 | 1.836  | 1.642 | 2.544 | 5.331 | 1.492 | 2.527 | 7.127 |
| $\hat{\theta}_1^{I\text{Al}}$ | 36     | 29     | 31     | 20     | 523    | 1301  | 1919   | 385   | 361   | 1033  | 676   | 884   | 876   |
Table 5: The root mean squared error \( k = \min(\hat{\theta}_0 L, \sqrt{L}) \), \( \hat{\theta}_0 \) is a pilot intervals estimate.

| RMSE | MM | ARMAX | \( AR_u^+ \) | \( AR_u^- \) | \( MA(2) \) | ARc | AR(2) | GARCH |
|------|----|-------|------------|------------|-----------|-----|-------|--------|
| \( 10^4/\theta \) | 0.5 | 0.8 | 0.25 | 0.75 | 0.5 | 0.8 | 0.75 | 0.96 | 0.5 | 2/3 | 0.3 | 0.25 | 0.447 |
| \( n = 10^5 \) | | | | | | | | | | | | | |
| \( \hat{\theta}_1 \) | 135 | 171 | 120 | 164 | 405 | 1059 | 1241 | \textbf{338} | 348 | 846 | 14 | 611 | 479 |
| \( \hat{\theta}_2 \) | 146 | 169 | 120 | 162 | 471 | 1150 | 1524 | 340 | 414 | 1193 | 137 | 721 | 700 |
| \( \hat{\theta}_3 \) | 191 | 245 | 171 | 246 | 383 | 997 | 1029 | 343 | 324 | 467 | 245 | 533 | 427 |
| \( \hat{\theta}_1^* \) | 101 | 118 | \textbf{85} | 115 | 388 | 1116 | 1310 | \textbf{329} | 331 | 834 | 151 | 620 | 416 |
| \( \hat{\theta}_2^* \) | 155 | \textbf{130} | 116 | \textbf{142} | 597 | 1527 | 2374 | 370 | 499 | 991 | 358 | 953 | 1124 |
| \( \hat{\theta}_3^* \) | 359 | 444 | 298 | 433 | 399 | 964 | 518 | 349 | 348 | 701 | 67 | 431 | 476 |
| \( \hat{\theta}_1^{K\text{dis}} \) | 384 | 738 | 155 | 670 | \textbf{188} | \textbf{384} | 975 | 672 | 537 | 1015 | \textbf{19} | 382 | 322 |
| \( \hat{\theta}_2^{K\text{dis}} \) | 3222 | 4566 | 1672 | 4317 | 3221 | 4457 | 4252 | 4334 | 3158 | 4232 | 3000 | 1697 | 3024 |
| \( \hat{\theta}_3^{K\text{dis}} \) | 2821 | 4372 | 1398 | 4061 | 2517 | 4411 | 4273 | 4596 | 3255 | 4385 | \textbf{19} | 1589 | 3192 |
| \( \hat{\theta}_1^{K\text{dis}*} \) | 673 | 1361 | 207 | 1232 | 381 | 591 | 379 | 642 | 865 | 1577 | 380 | 273 | 195 |
| \( \hat{\theta}_2^{K\text{dis}*} \) | 220 | 572 | \textbf{67} | 501 | 237 | 484 | 2501 | 404 | 315 | 517 | 34 | 2531 | 4176 |
| \( \hat{\theta}_3^{K\text{dis}*} \) | 223 | 536 | 194 | 449 | 240 | \textbf{565} | 397 | 5588 | 303 | 596 | 75 | \textbf{312} | 398 |
| \( n = 5000 \) | | | | | | | | | | | | | |
| \( \hat{\theta}_1 \) | 535 | 636 | 471 | 625 | 813 | 1260 | 1499 | 424 | 507 | \textbf{786} | - | 943 | 992 |
| \( \hat{\theta}_2 \) | 499 | 620 | 413 | 593 | 795 | 1328 | 1754 | 428 | 560 | 1203 | - | 1000 | 1122 |
| \( \hat{\theta}_3 \) | 918 | 1008 | 823 | 1004 | 1152 | 1302 | 1460 | 490 | 828 | 1655 | - | 1093 | 1242 |
| \( \hat{\theta}_1^* \) | \textbf{365} | \textbf{499} | 343 | \textbf{483} | 714 | 1290 | 1445 | \textbf{313} | \textbf{357} | 936 | 606 | 897 | 891 |
| \( \hat{\theta}_2^* \) | 368 | \textbf{496} | 292 | \textbf{458} | 815 | 1604 | 2260 | \textbf{389} | 577 | 1104 | 422 | 1120 | 1359 |
| \( \hat{\theta}_3^* \) | 1668 | 1555 | 1359 | 1697 | 1887 | 1662 | 1716 | 721 | 1570 | 1043 | 1455 | 1649 | 2030 |
| \( \hat{\theta}_1^{K\text{dis}} \) | 598 | 1147 | 354 | 1028 | \textbf{454} | \textbf{633} | \textbf{746} | 954 | 725 | 1283 | 77 | \textbf{844} | \textbf{808} |
| \( \hat{\theta}_2^{K\text{dis}} \) | 3437 | 5398 | 1705 | 5075 | 3412 | 5268 | 5266 | 5742 | 3543 | 4689 | 3000 | 2133 | 4032 |
| \( \hat{\theta}_3^{K\text{dis}} \) | 3054 | 5158 | 1399 | 4744 | 2832 | 5335 | 5137 | 5513 | 3163 | 4776 | \textbf{11} | 1859 | 3553 |
| \( \hat{\theta}_1^{K\text{dis}*} \) | 792 | 1909 | \textbf{276} | 1712 | 582 | 1253 | 929 | 1044 | 959 | 1821 | 93 | \textbf{836} | \textbf{377} |
| \( \hat{\theta}_2^{K\text{dis}*} \) | 251 | 585 | \textbf{188} | 509 | \textbf{288} | \textbf{885} | 2522 | 423 | \textbf{341} | \textbf{498} | \textbf{25} | 3528 | 3775 |
| \( \hat{\theta}_3^{K\text{dis}*} \) | 793 | 958 | 981 | 876 | 1090 | 905 | \textbf{819} | 554 | 725 | 929 | 2321 | 1095 | 1404 |

36
Table 6: The absolute bias ($k = \lfloor \min(\hat{\theta}_0 L, \sqrt{L}) \rfloor$), $n = 10^5$.

| $\theta$ | $10^3/\hat{\theta}$ | $n = 10^5$ | $n = 5000$ |
|----------|----------------------|-------------|-------------|
| $\hat{\theta}_1$ | 61 36 | 41 30 | 356 1036 | 541 1132 |
| $\hat{\theta}_2$ | 84 48 | 60 49 | 423 1122 | 605 1114 |
| $\hat{\theta}_3$ | 35 15 | 21 5.4013 | 281 957 | 605 1114 |
| $\hat{\theta}_1^*$ | 59 37 | 42 43 | 376 1111 | 548 1222 |
| $\hat{\theta}_2^*$ | 133 76 | 97 97 | 589 1522 | 682 1522 |
| $\hat{\theta}_3^*$ | 38 35 | 45 42 | 119 865 | 163 842 |
| $\hat{\theta}_1^{K_{dis}}$ | 347 677 | 122 615 | 80 197 | 250 251 |
| $\hat{\theta}_2^{K_{dis}}$ | 2177 2874 | 1143 2717 | 2012 1932 | 2501 4049 |
| $\hat{\theta}_3^{K_{dis}}$ | 1840 2909 | 853 2669 | 1405 2460 | 2250 2460 |
| $\hat{\theta}_1^{K_{dis}^*}$ | 669 1355 | 200 1227 | 372 577 | 365 476 |
| $\hat{\theta}_2^{K_{dis}^*}$ | 217 570 | 52 499 | 163 842 | 2501 4049 |
| $\hat{\theta}_3^{K_{dis}^*}$ | 151 494 | 18 405 | 74 161 | 250 251 |

| $\hat{\theta}_1$ | 101 75 | 105 105 | 541 1132 | 1331 201 |
| $\hat{\theta}_2$ | 15 6.55145.8965 22 | 538 1203 | 1601 206 | 317 1139 |
| $\hat{\theta}_3$ | 268 203 | 263 210 | 605 1114 | 1068 199 |
| $\hat{\theta}_1^*$ | 111 125 | 99 150 | 548 1222 | 1384 260 |
| $\hat{\theta}_2^*$ | 54 26 | 50 29 | 682 1549 | 2249 381 |
| $\hat{\theta}_3^*$ | 925 697 | 706 897 | 977 1310 | 975 102 |
| $\hat{\theta}_1^{K_{dis}}$ | 476 967 | 101 881 | 148 135 | 418 |
| $\hat{\theta}_2^{K_{dis}}$ | 2435 3873 | 1157 3633 | 2266 2983 | 3242 |
| $\hat{\theta}_3^{K_{dis}}$ | 2044 3748 | 689 3427 | 1649 3572 | 3578 3254 |
| $\hat{\theta}_1^{K_{dis}^*}$ | 775 1891 | 143 1696 | 522 1201 | 868 779 |
| $\hat{\theta}_2^{K_{dis}^*}$ | 197 549 | 36 473 | 181 874 | 2522 423 |
| $\hat{\theta}_3^{K_{dis}^*}$ | 153 103 | 305 55 | 289 555 | 104 131 |
Table 7: The root mean squared error ($k = (\ln L)^2$)).

| RMSE | MM  | ARMAX | ARu$^+$ | ARu$^-$ | MA(2) | ARc  | AR(2) | GARCH |
|------|-----|-------|---------|---------|-------|------|-------|-------|
| $10^{4}/\theta$ | 0.5 | 0.8  | 0.25  | 0.75  | 0.5  | 0.8  | 0.75  | 0.96  | 0.5 | 2/3 | 0.3  | 0.25  | 0.447 |
| $n = 10^5$ |   |      |       |        |       |      |       |       |       |      |     |      |       |       |
| $\theta_1$ | **147** | 167  | 126  | 176  | 396  | 383  | 1310 | 373  | 345  | 834  | 83  | 601  | 453  |       |
| $\theta_2$ | 158 | 159  | 131  | 173  | 468  | 463  | 1669 | 387  | 409  | 982  | 358 | 720  | 693  |       |
| $\theta_3$ | 218 | 261  | 172  | 257  | 370  | **353** | 1039 | **367** | 319  | 706  | 245 | 519  | 406  |       |
| $\hat{\theta}_1$ | **98** | **118** | **85** | **117** | 386 | 1122 | 1444 | 378  | 334  | 846  | 151 | 619  | 416  |       |
| $\hat{\theta}_2$ | 151 | **134** | 113  | **141** | 599  | 1534 | 2500 | 400  | 504  | 1197 | 358 | 948  | 1131 |       |
| $\hat{\theta}_3$ | 343 | 478  | 293  | 421  | 412  | 980  | 515  | **359** | 336  | **439** | 67  | 444  | 455  |       |
| $\hat{\theta}_{1}^{Kdis}$ | 366 | 710  | 155  | 675  | **186** | 374  | 1018 | 508  | 527  | 1032 | 150 | 338  |       | **311** |
| $\hat{\theta}_{2}^{Kdis}$ | 3148 | 4925 | **1659** | 4403 | 3336 | 4512 | 3895 | 5259 | 3208 | 4254 | 3000 | 1636 | 2962 |       |
| $\hat{\theta}_{3}^{Kdis}$ | 3003 | 4608 | **1439** | 4441 | 2598 | 4428 | 4004 | 5216 | 3255 | 4597 | **19** | 1559 | 3034 |       |
| $\hat{\theta}_{1}^{Kdis}$ | 696 | 1368 | **210** | 1255 | 398 | 598  | **356** | 1248 | 880  | 1602 | 380  | 230  | 181  |       |
| $\hat{\theta}_{2}^{Kdis}$ | 220 | **573** | **71** | 500  | 362 | 844  | 2501 | 401  | 315  | **558** | **34** | 2512 | 4101 |       |
| $\hat{\theta}_{3}^{Kdis}$ | **225** | **526** | **183** | **458** | 243 | 263  | 462  | 498  | 297  | 584  | 75  | **304** | 392  |       |
| $n = 5000$ |   |      |       |        |       |      |       |       |       |      |     |      |       |       |
| $\theta_1$ | 493 | 683  | 399  | 652  | 808 | 1301 | **1619** | **382** | 535 | 898  | -   | 905 | 988  |       |
| $\theta_2$ | 484 | 662  | 348  | 641  | 784 | 1384 | 1910 | 407  | 593 | 1087 | -   | 953 | 1132 |       |
| $\theta_3$ | 796 | 1026 | 624  | 997  | 1091 | 1365 | 1544 | 486  | 814 | 1037 | -   | 1011 | 1146 |       |
| $\hat{\theta}_1$ | 357 | **480** | **340** | **472** | 689 | 1274 | 1631 | **325** | **367** | **764** | 559 | 909 | 908  |       |
| $\hat{\theta}_2$ | **353** | **482** | 295  | **432** | 790 | 1611 | 2441 | 399  | 554 | 1211 | 422 | 1138 | 1358 |       |
| $\hat{\theta}_3$ | 1572 | 1568 | 1295 | 1735 | 1871 | 1658 | 1754 | 716  | 1547 | 1617 | 1455 | 1618 | 1979 |       |
| $\hat{\theta}_{1}^{Kdis}$ | 486 | 915  | 351  | 822  | **504** | 557  | 914  | 673  | 596 | 1123 | -   | **710** | **825** |       |
| $\hat{\theta}_{2}^{Kdis}$ | 3214 | 4704 | 1543 | 4618 | 3238 | 4581 | 4483 | 5477 | 3290 | 4262 | -   | 1746 | 3571 |       |
| $\hat{\theta}_{3}^{Kdis}$ | 2979 | 4482 | 1718 | 4283 | 2765 | 4564 | 4559 | 5671 | 3218 | 4425 | -   | 1894 | 3264 |       |
| $\hat{\theta}_{1}^{Kdis}$ | 767 | 1706 | **263** | 1535 | 549 | 976  | **702** | 1715 | 943 | 1757 | **93** | **726** | **406** |       |
| $\hat{\theta}_{2}^{Kdis}$ | **257** | 589  | **194** | 516  | **397** | **874** | 2518 | 418  | **345** | **502** | **25** | 3299 | 3817 |       |
| $\hat{\theta}_{3}^{Kdis}$ | 812 | 1002 | 943  | 913  | 1058 | **921** | **897** | 573  | 720 | 860  | 2321 | 1253 | 1424 |       |
Table 8: The absolute bias \( (k = \lfloor (\ln L)^2 \rfloor) \).

| Bias \( \times 10^4 \) | MM  | ARMAX | \( AR_u^+ \) | \( AR_u^- \) | \( MA(2) \) | \( AR_c \) | \( AR(2) \) |
|----------------|-----|-------|-------------|-------------|----------|--------|--------|
| \( n = 10^5 \) |     |       |             |             |          |        |        |
| \( \theta_1 \) | 67  | 37    | 45 39      | 354 343     | 1212 361 | 309 798| 83 569 |
| \( \theta_2 \) | 95  | 56    | 69 60      | 428 425     | 1543 378 | 377 944| 358 683 |
| \( \theta_3 \) | 42  | 8.2991| 15 17      | 277 253     | 861 338  | 232 622| 245 454 |
| \( \theta_1^* \) | 57  | 30    | 42 42      | 371 1117    | 1440 377 | 326 841| 151 613 |
| \( \theta_2^* \) | 126 | 78    | 95 96      | 589 1530    | 2500 400 | 498 1193| 944 1124|
| \( \theta_3^* \) | 56  | 48    | 42 44      | 120 872     | 262 235  | 82 242 | 67 316 |
| \( \theta_1^{K_{dis}} \) | 331 | 653   | 122 615    | 69 208      | 486 321  | 492 966| 150 249 |
| \( \theta_2^{K_{dis}} \) | 2084| 3275  | 1128 2821  | 2169 2006   | 22 2675  | 2200 2978| 3000 521 |
| \( \theta_3^{K_{dis}} \) | 2022| 3119  | 891 3047   | 1478 2458   | 1895 3226| 2359 3575| 19 802 |
| \( \theta_1^{K_{dis}}^* \) | 693 | 1362  | 203 1249   | 388 584     | 337 1239 | 878 1598| 380 191 |
| \( \theta_2^{K_{dis}}^* \) | 217 | 571   | 56 499     | 149 843     | 2501 401 | 314 482 | 34 167 |
| \( \theta_3^{K_{dis}}^* \) | 159 | 483   | 17 417     | 87 211      | 265 484  | 258 556 | 75 234 |
| \( n = 5000 \) |     |       |             |             |          |        |        |
| \( \theta_1 \) | 91  | 147   | 112 128    | 553 1173    | 1470 268 | 144 629| - 781 |
| \( \theta_2 \) | 17  | 41    | 28 12      | 543 1255    | 1767 331 | 299 890| - 827 |
| \( \theta_3 \) | 255 | 290   | 235 326    | 619 1106    | 1212 210 | 79 260 | - 789 |
| \( \theta_1^* \) | 105 | 121   | 103 146    | 529 1274    | 1580 294 | 175 674| 559 816 |
| \( \theta_2^* \) | 67  | 20    | 55 26      | 662 1560    | 2435 398 | 445 1148| 422 1065|
| \( \theta_3^* \) | 814 | 722   | 761 840    | 959 1277    | 1088 89  | 762 536| 1455 830|
| \( \theta_1^{K_{dis}} \) | 322 | 731   | 47 610     | 26 133 421  | 425 947  | 479 253|
| \( \theta_2^{K_{dis}} \) | 2140| 3018  | 944 3037   | 2001 2049   | 543 2890 | 2277 2926| - 359 |
| \( \theta_3^{K_{dis}} \) | 1871| 2941  | 1072 2757  | 1477 2572   | 2638 3702| 2215 3232| - 952 |
| \( \theta_1^{K_{dis}}^* \) | 747 | 1680  | 139 1511   | 477 943     | 645 1699 | 933 1748| 93 648 |
| \( \theta_2^{K_{dis}}^* \) | 203 | 55    | 39 478     | 158 863     | 2518 418 | 305 454 | 25 1489|
| \( \theta_3^{K_{dis}}^* \) | 195 | 173   | 352 91     | 239 559     | 106 136  | 77 203 | 2321 378 | 237 39
Table 9: Extremal index estimates for Uccle data.

| n   | 'Kimt' | 'IntA1' | 'Intdis' | 'Kdis' | 'K0dis' |
|-----|--------|---------|----------|--------|--------|
|     | s = \theta_0 s = 0.51 s = \theta_0 s = 0.51 s = \theta_0 s = 0.51 |
| 3100 | 0.5133 | 0.4625  | 0.5329   | 0.5741 | 0.5670 | 0.5879 | 0.5383 | 0.5148 |
|     | \hat{\theta}_1^*  | \hat{\theta}_2^*  | \hat{\theta}_3^*  |
|     | 0.4199 | 0.4637  | 0.7244   | 1      | &nbsp; | &nbsp; | &nbsp; | &nbsp; |
| 5177 | 0.5695 | 0.4392  | 0.4655   | 0.4837 | 0.5632 | 0.6251 | 0.4741 | 0.5691 |
|     | \hat{\theta}_1^*  | \hat{\theta}_2^*  | \hat{\theta}_3^*  |
|     | 0.4184 | 0.4919  | 0.7024   | 0.6524 | 0.6524 |

Table 10: Extremal index estimates for dewpoint temperatures data.

| n   | 'Kimt' | 'IntA1' | 'Intdis' | 'Kdis' | 'K0dis' |
|-----|--------|---------|----------|--------|--------|
|     | k = \lceil \theta_0 L \rceil |
| 13886 | 0.4753 | 0.1541  | \hat{\theta}_1^*  | \hat{\theta}_2^*  | \hat{\theta}_3^*  |
|     | 0.2489 | 0.2003  | 0.4092   | 0.3178 | 0.4765 | 0.2016 | 0.2749 | 0.5181 |