Extremes (2019) 22: 317–341
https://doi.org/10.1007/s10687-018-0337-5

Multiple thresholds in extremal parameter estimation

Julian Sun[1] · **Gennady Samorodnitsky**[2]

Received: 24 August 2017 / Revised: 18 September 2018 / Accepted: 21 September 2018 /
Published online: 3 October 2018

Abstract
Selecting the number of upper order statistics to use in extremal inference or selecting the threshold above which we perform the extremal inference is a common step in applications of extreme value theory. Not only is the selection itself difficult, but the large part of the sample below the threshold may potentially carry useful information. We propose an approach that takes an extremal parameter estimator and modifies it to allow for using multiple thresholds instead of a single one. We apply this approach to the problem of estimating the extremal index and demonstrate its power both on simulated and real data.

Keywords Extremal inference · Regular variation · Threshold selection · Extremal index · Bias

AMS 2000 Subject Classifications Primary 62G32; Secondary 60G70

1 Introduction

Many statistical procedures in extreme value theory depend on a choice of a threshold such that only the observations above that threshold are used for the inference. In the classical Hill estimator of the exponent of regular variation, this corresponds to

This research was partially supported by the ARO grant W911NF-12-10385 at Cornell University.

✉ Julian Sun
ys598@cornell.edu

Gennady Samorodnitsky
gs18@cornell.edu

[1] School of Operations Research and Information Engineering, Cornell University, Ithaca, NY, 14853, USA

[2] School of Operations Research and Information Engineering and Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA

choosing the number of the upper statistics used to construct the estimator, and in the standard "peaks over threshold" procedures, the term "threshold" even appears in the name; see e.g. de Haan and Ferreira (2006) and Resnick (2007). The inference results often depend on the threshold in a significant way, so a major effort has been invested in choosing the threshold "in the right way"; see e.g. Resnick and Stărică (1997), Drees and Kaufmann (1998), Dupuis (1998), Nguyen and Samorodnitsky (2012). A threshold-based extremal inference procedure discards the observations below the threshold, which in most cases amounts to discarding a larger part of the sample. This counterintuitive step reflects the underlying belief that the observations above the threshold carry information about the "tail" of the distribution, while those below the threshold carry information about the "center" of the distribution.

It is reasonable to assume that such a binary rule by necessity neglects a part of the information stored in the original sample that is relevant for extremal inference. An alternative to using a binary rule would be acknowledging that larger observations carry more information about the "extremes" than smaller observations do, but instead of discarding the latter completely, using them in the extremal inference, with a smaller weight. This idea can be implemented in a number of ways, the most natural of which is to use multiple "thresholds" instead of trying to select the "right" threshold. In this case it is more appropriate to talk about "levels" of observations that are weighted differently, rather than "thresholds". In this paper we apply this idea to estimating the extremal index (defined below), but the approach is more general than its application to the estimation of the extremal index. It can, in principle, be used in any extremal estimation problem, though the actual implementation may depend significantly on the problem.

A number of specific statistical algorithms for extremes have been proposed that avoid the problem of threshold selection entirely (such as Northrop 2015), or have the threshold be determined by something else (e.g. the block size, see Robert 2009). Multiple thresholds in extremal inference have been used as well. In Laurini and Tawn (2003) a two-threshold procedure, also for estimating the extremal index, is suggested. The role of the second, lower, threshold is to help separating between different exceedance clusters. In Drees (2011), on the other hand, estimates of the extremal index based on multiple thresholds were combined together in order to reduce the bias of the estimation.

Our idea is different. Since performing extremal inference based on a small number of observations tends to result in a high variance of the estimator, we view using multiple levels as a means to incorporate more observations into an estimator and to reduce the variance by doing so. However, incorporating smaller observations into extremal inference is likely to increase the bias of the resulting estimator, so one needs to find a way to cope with this problem. This approach can be applied to different extremal estimation problems. As mentioned earlier, in this paper we implement this idea in estimating the *extremal index*, a quantity designed to measure the amount of clustering of the extremes in a stationary sequence. Suppose that $X_1, X_2, \ldots$ is a stationary sequence of random variables with a marginal distribution function F, and let $M_n = \max(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$. Suppose there exists $\theta \geq 0$ with the following property: for every $\tau > 0$, there is a sequence (v_n) such that $n\bar{F}(v_n) \to \tau$ and $P(M_n \leq v_n) \to e^{-\theta\tau}$ as $n \to \infty$, where $\bar{F} = 1 - F$. Then θ is called the extremal index of the sequence $X_1, X_2, \ldots$; it is automatically in the range $0 \leq \theta \leq 1$; see Leadbetter et al. (1983) or Embrechts et al. (1997). The relation of the extremal index to extremal clustering is best observed by considering the exceedances of the stationary sequence over high thresholds. Let (v_n) be a sequence such that $n\bar{F}(v_n) \to \tau$ as $n \to \infty$ for some $\tau > 0$. Then under certain mixing conditions, the point processes of exceedances converge weakly in the space of finite point processes on [0, 1] to a compound Poisson process:

$$N_n = \sum_{i=1}^{n} \delta_{i/n} \mathbb{1}(X_i > v_n) \xrightarrow{d} N = \sum_{i=1}^{\infty} \xi_i \delta_{\Gamma_i}, \tag{1.1}$$

where δ_x is a point mass at x, the points $0 < \Gamma_1 < \Gamma_2 < \ldots$ constitute a homogeneous Poisson process with intensity $\tau\theta$ on [0, 1] which is independent of an independent and identically distributed (i.i.d.) positive integer-valued sequence $\{\xi_i\}$; see e.g. Hsing et al. (1988). The latter sequence is interpreted as the sequence of the extremal cluster sizes, and the extremal index θ is, under mild conditions, equal to the reciprocal of the expected cluster size $E\xi$. We will assume that the latter expectation is finite, and the extremal index is positive.

The problem of estimating the extremal index parameter is well-known in the literature; references include Hsing (1993), Smith and Weissman (1994), Ferro and Segers (2003), Northrop (2015), and Berghaus and Bücher (2017). The most common methods of estimation include the blocks method, the runs method, and the inter-exceedance method. In this paper we choose the blocks method in order to demonstrate an application of our idea for variance reduction using multiple levels.

The blocks method is based on the interpretation of the extremal index as the reciprocal of the expected cluster size of extremes. It is based on choosing a block size r_n much smaller than n and a level (or threshold) u_n. Split the n observations $X_1, X_2, \ldots, X_n$ into $k_n = \lfloor n/r_n \rfloor$ contiguous blocks of equal length r_n. The blocks estimator is then defined as the reciprocal of average number of exceedances of the level u_n per block among blocks with at least one exceedance. If $M_{i,j}$ denotes $\max\{X_{i+1}, \ldots, X_j\}$ for $i < j$ and $M_j = M_{0,j}$, then the blocks estimator has the form

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{k_n} \mathbb{1}\big(M_{(i-1)r_n, ir_n} > u_n\big)}{\sum_{i=1}^{k_n r_n} \mathbb{1}\big(X_i > u_n\big)}. \tag{1.2}$$

Assuming that $r_n\bar{F}(u_n) \to 0$ but $n\bar{F}(u_n) \to \infty$ as $n \to \infty$, and certain mixing conditions, this estimator has been shown to be consistent and asymptotically normal; see Hsing (1991) and Weissman and Novak (1998). In Section 2 we introduce a version of the blocks estimator using multiple thresholds (levels) and list the assumptions used in the paper. Section 3 considers the asymptotic behaviour of the various ingredients in our estimator. In Section 4 we prove a central limit theorem for the estimator. In Section 5 we both propose a procedure to reduce the bias of the estimator as well as present a simulation study and a case study.

2 The estimator

Let $X_1, \ldots, X_n$ be a stationary sequence of random variables with marginal distribution F, and an extremal index $\theta \in (0, 1]$. We now present a version of the blocks estimator (1.2) based on multiple levels. With a block size r_n and the number of blocks $k_n = \lfloor n/r_n \rfloor$ as before, we select now m levels $u_n^1 < \cdots < u_n^m := u_n$, and we view the highest level u_n^m as corresponding to the single level u_n in Eq. 1.2. The lower levels u_n^s, $s = 1, \ldots, m-1$ are used to reduce the variance of the estimator. The levels are chosen in an "asymptotically balanced" way. Specifically, it will be assumed that, as $n \to \infty$,

$$\frac{\bar{F}(u_n^s)}{\bar{F}(u_n^m)} \to \frac{\tau_s}{\tau_m},\ s = 1, \ldots, m \tag{2.1}$$

for some $\tau_1 > \cdots > \tau_m > 0$.

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuously differentiable positive decreasing function. We will use f as a weight function, and we would like to weigh the exceedances over the level u_n^s by $f(\tau_s/\tau_m)$. The fact that f is decreasing reflects our belief that higher exceedances provide more reliable information about the extremes. We will not assume that the numbers $\tau_1, \ldots, \tau_m$ are known ahead of time, so we will use, in practice, an estimator of the ratio τ_s/τ_m. Specifically, we will use

$$\widehat{\tau_s/\tau_m} = \frac{\sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)}{\sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^m)},\ s = 1, \ldots, m\,. \tag{2.2}$$

Then our version of the blocks estimator (1.2) based on multiple levels is

$$\widehat{\theta}_n(f) = \frac{\sum_{s=1}^m \left[f\left(\widehat{\tau_s/\tau_m}\right) - f\left(\widehat{\tau_{s-1}/\tau_m}\right)\right] \sum_{i=1}^{k_n} \mathbb{1}\left(M_{(i-1)r_n, ir_n} > u_n^s\right)}{\sum_{s=1}^m \left[f\left(\widehat{\tau_s/\tau_m}\right) - f\left(\widehat{\tau_{s-1}/\tau_m}\right)\right] \sum_{i=1}^{k_n r_n} \mathbb{1}\left(X_i > u_n^s\right)}\,, \tag{2.3}$$

with the convention that $f\left(\widehat{\tau_0/\tau_m}\right) = 0$. Note that when $m = 1$, Eq. 2.3 reduces to Eq. 1.2.

Consistency and asymptotic normality of this estimator depend, as they do for all other related estimators, on certain mixing-type assumptions. Different sets of such conditions are available in literature. We explain next the conditions that we will use in this paper. These are based on the setup in Hsing et al. (1988). For $1 \le i \le j \le n$, and levels w_n, w_n', let $\mathscr{B}_i^j(w_n, w_n')$ denote the σ-field generated by the events $\{X_d \le w_n\}$ and $\{X_d \le w_n'\}$ for $i \le d \le j$. For $n \ge 1$ and $1 \le l \le n-1$ define

$$\begin{aligned}&\alpha_{n,l}(w_n, w_n') = \max(|P(A \cap B) - P(A)P(B)|\\ &: A \in \mathscr{B}_1^k(w_n, w_n'),\, B \in \mathscr{B}_{k+l}^n(w_n, w_n'),\, 1 \le k \le n-l)\end{aligned}$$

and write $\alpha_{n,l}(w_n) = \alpha_{n,l}(w_n, w_n)$. Similarly, one uses the maximal correlation coefficient

$$\begin{aligned}&\rho_{n,l}(w_n, w_n') = \max(\mathrm{corr}(X, Y) : X \in L^2(\mathscr{B}_1^k(w_n, w_n')),\\ &Y \in L^2(\mathscr{B}_{k+l}^n(w_n, w_n')),\, 1 \le k \le n-l)\,,\end{aligned}$$

where $L^2(\mathscr{F})$ denotes the space of $\mathscr{F}$-measurable square-integrable random variables. Again, we write $\rho_{n,l}(w_n) = \rho_{n,l}(w_n, w_n)$. Trivially,

$$\rho_{n,l}(w_n, w_n') \geq 4\alpha_{n,l}(w_n, w_n').$$

The sequence $\{X_i\}$ is said to satisfy the condition $\Delta(\{w_n\})$ if $\alpha_{n,l_n}(w_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(n)$. If $\{p_n\}$ is a sequence of integers and $\alpha_{p_n,l_n}(w_n) \to 0$ as $n \to \infty$ for some sequence $\{l_n\}$ with $l_n = o(p_n)$, then we will say that $\{X_i\}$ satisfies the condition $\Delta_{\{p_n\}}(\{w_n\})$.

As mentioned earlier, the condition that $r_n \bar{F}(u_n) \to 0$ but $n\bar{F}(u_n) \to \infty$ as $n \to \infty$ is usually required for asymptotic consistency results. This implicitly uses the traditional assumption that $r_n = o(n)$ as $n \to \infty$. It will be convenient to introduce a specific sequence of the integers $\{p_n\}$, which is an intermediate growth sequence between the sequence of the block size $\{r_n\}$ and the sequence of the sample sizes $\{n\}$. Specifically, let

$$p_n \bar{F}(u_n^s) \to \tau_s, \ s = 1, \ldots, m. \tag{2.4}$$

According to Eq. 2.1 one such sequence is $p_n = \lceil \tau_m (\bar{F}(u_n))^{-1} \rceil, n = 1, 2, \ldots$.

The following assumptions on the stationary sequence $\{X_i\}$ will used throughout this paper, not necessarily all in the same place. Some of the assumptions form stronger versions of other assumptions.

Assumption Δ' There is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \alpha_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

Assumption $\mathbf{C_1}$ For each $s = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s) = o(r_n)$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

Assumption $\mathbf{C_1'}$ For each $s = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s) = o(r_n^{1/2})$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$ for each $s = 1, \ldots, m$.

Assumption $\mathbf{C_2}$ For each $s, t = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s, u_n^t) = o(r_n)$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s, u_n^t) \to 0$ as $n \to \infty$ for each $s, t = 1, \ldots, m$.

Assumption $\mathbf{C}_2'$ For each $s, t = 1, \ldots, m$,

$$\sum_{l=1}^{n} \rho_{n,l}(u_n^s, u_n^t) = o(r_n^{1/2})$$

as $n \to \infty$, and there is a sequence $l_n = o(r_n)$ such that $p_n r_n^{-1} \rho_{n,l_n}(u_n^s, u_n^t) \to 0$ as $n \to \infty$ for each $s, t = 1, \ldots, m$. The next group of assumptions deals with convergence of certain counting processes. Let $N_{p_n}^{(u)}$ be the point process on [0, 1] with points $(j/p_n : 1 \le j \le p_n, X_j > u_n)$. Furthermore, for $w > 0$ we write $N_k(w) = \sum_{i=1}^{k} \mathbb{1}(X_i > w)$.

Assumption P $N_{p_n}^{(u)}$ converges weakly in the space of finite point processes on [0, 1].

Assumption $\mathbf{D_1}$ There exists a probability distribution $(\pi_j)_{j\ge1}$ on the positive integers such that for all $1 \le s \le m$,

$$P(N_{r_n}(u_n^s) = j | M_{r_n} > u_n^s) \to \pi_j, \ j \ge 1,$$
$$E[N_{r_n}^2(u_n^s) | M_{r_n} > u_n^s] \to \sum_{j=1}^{\infty} j^2 \pi_j < \infty.$$

Assumption $\mathbf{D_2}$ There exist probability distributions $(\varpi_{s,t}(i, j))_{i\ge1, j\ge0}$ on $\mathbb{Z}_+ \times \mathbb{Z}_{\ge0}$ such that for all $1 \le s < t \le m$,

$$P(N_{r_n}(u_n^s) = i, N_{r_n}(u_n^t) = j | M_{r_n} > u_n^s) \to \varpi_{s,t}(i, j), \ i \ge j \ge 0, \ i \ge 1,$$
$$E\big[N_{r_n}(u_n^s) N_{r_n}(u_n^t) | M_{r_n} > u_n^s\big] \to \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij \varpi_{s,t}(i, j) < \infty.$$

Remark 2.1 It is clear that Assumption Δ' is implied by Assumption C_1 which is, in turn, implied both by Assumption C_1' and by Assumption C_2. Further, it follows by Theorem 4.1 of Hsing et al. (1988) that the first part of Assumption D_1 is implied by Assumptions Δ' and P. Note that Assumptions C_1, C_2 and D_1 are identical to those posed (Robert et al. 2009).

Remark 2.2 The mixing conditions $\Delta', C_1, C_1', C_2, C_2'$ are conditions relating the rate of decay of mixing and correlation coefficients and the size of the blocks r_n, which must be "large enough". In many models the mixing and correlation coefficients decay very fast (e.g. m-dependent sequences, or geometrically mixing sequences), and so there is a great latitude in choosing the block sizes.

If Assumption Δ' holds, then it follows from Theorem 5.1 and Lemma 2.3 of Hsing et al. (1988) that

$$P(M_{r_n} > u_n^s) \sim \tau_s \theta r_n / p_n \tag{2.5}$$

as $n \to \infty$ for $s = 1, \ldots, m$. If we denote

$$\theta_n(f) = \theta_n(\tau_1, \ldots, \tau_m, f) = \frac{p_n}{r_n} \cdot \frac{\sum_{s=1}^{m}(f(\tau_s/\tau_m) - f(\tau_{s-1}/\tau_m))P(M_{r_n} > u_n^s)}{\sum_{s=1}^{m}(f(\tau_s/\tau_m) - f(\tau_{s-1}/\tau_m))\tau_s}, \tag{2.6}$$

then $\theta_n(f) \to \theta$ as $n \to \infty$.

Another immediate conclusion from Eq. 2.5 is that if Assumptions Δ', D_1 and D_2 hold, then for $1 \leq s < t \leq m$,

$$P(N_{r_n}(u_n^s) = i | M_{r_n} > u_n^t) \to \frac{\tau_s}{\tau_t}\big(\pi_i - \varpi_{s,t}(i,0)\big),\ i \geq 1,$$

$$E\big[N_{r_n}(u_n^s)|M_{r_n} > u_n^t\big] \to \psi_{s,t} := \frac{\tau_s}{\tau_t}\sum_{i=1}^{\infty} i\big(\pi_i - \varpi_{s,t}(i,0)\big). \tag{2.7}$$

3 Preliminary results

The estimator (2.3) is composed of several extremal statistics. In this section we will take a close look at these and related statistics and derive their asymptotic variances and covariances. The derivations are similar to those in Robert et al. (2009). Let $m_n \to \infty$ be a sequence of positive integers such that $m_n r_n \leq n$ for all n. For each level $u_n^s,\ s = 1, \ldots, m$

$$\widehat{M}_{n,m_n}(u_n^s) = \sum_{i=1}^{m_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{3.1}$$

and

$$\widehat{\tau}_{n,m_n}(u_n^s) = \sum_{i=1}^{m_n r_n} \mathbb{1}(X_i > u_n^s). \tag{3.2}$$

Note that the estimator (2.3) uses these statistics with $m_n = k_n$.

We first consider the asymptotic variance of $\widehat{M}_{n,m_n}(u_n^s)$.

Proposition 3.1 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Let (p_n) be as in* Eq. 2.4*, and suppose that Assumption C_1 holds. Then for $1 \leq s \leq m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n}\mathrm{var}(\widehat{M}_{n,m_n}(u_n^s)) \to \tau_s\theta. \tag{3.3}$$

Proof Fix $1 \le s \le m$ and write out the variance:

$$\begin{aligned}
\mathrm{var}(\widehat{M}_{n,m_n}(u_n^s)) &= \sum_{i=1}^{m_n} \mathrm{var}(\mathbb{1}(M_{(i-1)r_n,ir_n} > u_n^s)) \\
&\quad +2 \sum_{1\le i<j\le m_n} \mathrm{cov}(\mathbb{1}(M_{(i-1)r_n,ir_n} > u_n^s), \mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^s)) \\
&= m_n P(M_{r_n} > u_n^s)(1 - P(M_{r_n} > u_n^s)) \\
&\quad +2(m_n-1)\big(P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) - (P(M_{r_n} > u_n^s))^2\big) \\
&\quad +2\sum_{v=2}^{m_n-1} (m_n - v)\mathrm{cov}(\mathbb{1}(M_{r_n} > u_n^s), \mathbb{1}(M_{vr_n,(v+1)r_n} > u_n^s)) \\
&:= I_{1,n} + I_{2,n} + I_{3,n} \,.
\end{aligned}$$

It follows from Eq. 2.5 that

$$\frac{p_n}{m_n r_n} I_{1,n} \to \tau_s \theta$$

as $n \to \infty$. Furthermore,

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{3,n} &\le 2\,\frac{p_n}{r_n}\mathrm{var}(\mathbb{1}(M_{r_n} > u_n^s)) \sum_{v=2}^{m_n-1} \rho_{n,(v-1)r_n}(u_n^s) \\
&\le 2\,\frac{p_n}{r_n}\mathrm{var}(\mathbb{1}(M_{r_n} > u_n^s)) \frac{1}{r_n}\sum_{l=1}^{n} \rho_{n,l}(u_n^s) \to 0
\end{aligned}$$

by Eq. 2.5 and Assumption C_1, so it remains to consider $I_{2,n}$. By Eq. 2.5 we only need to show that

$$p_n r_n^{-1} P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) \to 0.$$

Note that

$$\begin{aligned}
P(M_{r_n} > u_n^s, M_{r_n,2r_n} > u_n^s) &\le P(M_{r_n-l_n} > u_n^s, M_{r_n,2r_n} > u_n^s) + P(M_{l_n} > u_n^s) \\
&\le P(M_{r_n-l_n} > u_n^s) P(M_{r_n} > u_n^s) + \alpha_{n,l_n}(u_n^s) + P(M_{l_n} > u_n^s) \\
&\le P(M_{r_n} > u_n^s)^2 + \alpha_{n,l_n}(u_n^s) + P(M_{l_n} > u_n^s).
\end{aligned}$$

Since $l_n = o(r_n)$, and $p_n r_n^{-1}\alpha_{n,l_n}(u_n^s) \to 0$, there is an intermediate sequence l_n' with $l_n = o(l_n')$ and $l_n' = o(r_n)$, such that $p_n(l_n')^{-1}\alpha_{n,l_n}(u_n^s) \to 0$. Then as in Eq. 2.5,

$$p_n(l_n')^{-1} P(M_{l_n'} > u_n^s) \to \tau_s \theta,$$

so we have both

$$p_n r_n^{-1} P(M_{l_n} > u_n^s) \le p_n(l_n')^{-1}(l_n' r_n^{-1}) P(M_{l_n'} > u_n^s) \to 0$$

and

$$p_n r_n^{-1} \alpha_{n,l_n}(u_n^s) \to 0\,.$$

Therefore, the result follows. □

The asymptotic covariance of $\widehat{M}_{n,m_n}(u_n^s)$ and $\widehat{M}_{n,m_n}(u_n^t)$ for $s \neq t$ can be obtained in an identical way (with a slightly different assumption). The proof is omitted.

Proposition 3.2 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Let (p_n) be as in* Eq. 2.4*, and suppose that Assumption C_2 holds. Then for $1 \le s < t \le m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{M}_{n,m_n}(u_n^t)) \to \tau_t \theta. \tag{3.4}$$

Now we find the variance and covariance of $\widehat{\tau}_{n,m_n}(u_n^s)$ and $\widehat{\tau}_{n,m_n}(u_n^t)$ for $1 \le s < t \le m$. We start with the variance.

Proposition 3.3 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumptions C_1 and D_1 hold. Then as $n \to \infty$, for $1 \le s \le m$,*

$$\frac{p_n}{m_n r_n} \mathrm{var}(\widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_s \theta \sum_{j=1}^{\infty} j^2 \pi_j\,. \tag{3.5}$$

Proof We proceed as in Proposition 3.1. Using the notation $N_{a,b}(w) = \sum_{a<i\le b} \mathbb{1}(X_i > w)$ for integers $0 \le a < b$, we obtain for a fixed $1 \le s \le m$,

$$\begin{aligned}
\mathrm{var}(\widehat{\tau}_{n,m_n}(u_n^s)) &= \mathrm{var}(N_{m_n r_n}(u_n^s)) \\
&= \sum_{i=1}^{m_n} \mathrm{var}\big(N_{(i-1)r_n, i r_n}(u_n^s)\big) \\
&\quad +2 \sum_{1 \le i < j \le m_n} \mathrm{cov}\big(N_{(i-1)r_n, i r_n}, N_{(j-1)r_n, j r_n}\big) \\
&= m_n \mathrm{var}(N_{r_n}) + 2(m_n - 1)\mathrm{cov}\big(N_{r_n}, N_{r_n, 2r_n}\big) \\
&\quad +2 \sum_{v=2}^{m_n-1} (m_n - v)\mathrm{cov}\big(N_{r_n} N_{v r_n, (v+1) r_n}\big) \\
&:= I_{1,n} + I_{2,n} + I_{3,n}\,.
\end{aligned}$$

It follows from Eq. 2.5 and Assumption D_1 that

$$\begin{aligned}
\frac{p_n}{m_n r_n} I_{1,n} &\sim \frac{p_n}{r_n} P(M_{r_n} > u_n^s) E[N_{r_n}^2(u_n^s) | M_{r_n} > u_n^s] \\
&- \frac{p_n}{r_n} \big(P(M_{r_n} > u_n^s)\big)^2 \big(E[N_{r_n}(u_n^s) | M_{r_n} > u_n^s]\big)^2 \\
&\to \tau_s \theta \sum_{j=1}^{\infty} j^2 \pi_j
\end{aligned}$$

as $n \to \infty$. Furthermore,

$$\frac{p_n}{m_n r_n} I_{3,n} \le 2\,\frac{p_n}{r_n}\mathrm{var}(N_{r_n}) \sum_{v=2}^{m_n-1} \rho_{n,(v-1)r_n}(u_n^s)$$
$$\le 2\,\frac{p_n}{r_n} P(M_{r_n} > u_n^s) E[N_{r_n}^2(u_n^s)|M_{r_n} > u_n^s] \frac{1}{r_n} \sum_{l=1}^{n} \rho_{n,l}(u_n^s) \to 0$$

by Assumptions C_1 and D_1. As far as $I_{2,n}$ is concerned, we only need to show that

$$p_n r_n^{-1} E\big(N_{r_n} N_{r_n,2r_n}\big) \to 0.$$

However,

$$E\big(N_{r_n} N_{r_n,2r_n}\big) = E\big(N_{r_n-l_n} N_{r_n,2r_n}\big) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big)$$
$$\le (EN_{r_n})^2 + E(N_{r_n})^2 \rho_{n,l_n}(u_n^s) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big).$$

By Assumptions C_1 and D_1 and the above calculation, both $k_n(EN_{r_n})^2 \to 0$ and $k_n E(N_{r_n})^2 \rho_{n,l_n}(u_n^s) \to 0$ as $n \to \infty$. Furthermore, by stationarity it is clear that

$$\frac{E(N_{r_n})^2}{E(N_{l_n})^2} \ge \lfloor r_n/l_n \rfloor \to \infty$$

as $n \to \infty$. Therefore,

$$k_n E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big) \le k_n \left(E(N_{l_n})^2\right)^{1/2} \left(E(N_{r_n})^2\right)^{1/2}$$
$$\le k_n E(N_{r_n})^2 \left(\frac{E(N_{l_n})^2}{E(N_{r_n})^2}\right)^{1/2} \to 0$$

as $n \to \infty$. This completes the proof. □

The asymptotic covariance between $\widehat{\tau}_{n,m_n}(u_n^s)$ and $\widehat{\tau}_{n,m_n}(u_n^t)$ for $1 \le s < t \le m$ can be found in the same way. Once again, we omit the proof.

Proposition 3.4 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumptions C_2 and D_2 hold. Then as $n \to \infty$, for $1 \le s < t \le m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{\tau}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_s \theta \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij\varpi_{s,t}(i, j). \tag{3.6}$$

We now address the asymptotic covariances between $\widehat{\tau}$ and $\widehat{M}$. We start with the "diagonal" case.

Proposition 3.5 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumption C_1' holds. Then as $n \to \infty$, for $1 \le s \le m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_s. \tag{3.7}$$

Proof Fix $1 \le s \le m$, we have

$$\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^s)) = \sum_{i=1}^{m_n} \sum_{j=1}^{m_n r_n} \mathrm{cov}(\mathbb{1}(M_{(i-1)r_n, ir_n} \le u_n^s), \mathbb{1}(X_j \le u_n^s)).$$

We split the sum into two pieces, $I_{1,n} + I_{2,n}$, depending on whether $(i-1)r_n < j \le ir_n$ or not. By stationarity,

$$\begin{aligned} \frac{p_n}{m_n r_n} I_{1,n} &\sim \frac{p_n}{r_n} \sum_{i=1}^{r_n} \mathrm{cov}(\mathbb{1}(M_{r_n} \le u_n^s), \mathbb{1}(X_i \le u_n^s)) \\ &\sim p_n P(X_1 > u_n^s) P(M_{r_n} \le u_n^s) \to \tau_s \end{aligned}$$

by Eqs. 2.4 and 2.5.

Furthermore, we can bound $I_{2,n}$ as follows:

$$|I_{2,n}| \le 2m_n \sqrt{\mathrm{var}(\mathbb{1}(M_{r_n} \le u_n^s))\mathrm{var}(\mathbb{1}(X_1 \le u_n^s))} \sum_{l=1}^{n} \rho_{n,l}(u_n^s),$$

and the fact that $(p_n/(m_n r_n)) I_{2,n} \to 0$ as $n \to \infty$ follows from Eqs. 2.4, 2.5 and Assumption C_1'. □

The asymptotic behaviour of $\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t))$ with $1 \le s < t \le m$ is similar to the "diagonal" case. The proof of the next proposition is similar to the argument in Proposition 3.5 (once we use the appropriate assumption), and is omitted.

Proposition 3.6 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumption C_2' holds. Then as $n \to \infty$, for $1 \le s < t \le m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_t. \tag{3.8}$$

Finally, we consider the asymptotic behaviour of $\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s))$ with $1 \le s < t \le m$.

Proposition 3.7 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumptions Δ', D_1 and D_2 hold. Then as $n \to \infty$, for $1 \le s < t \le m$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_t \theta \psi_{s,t}, \tag{3.9}$$

where $\psi_{s,t}$ is defined in Eq. 2.7.

Proof As before,

$$\mathrm{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s)) = \sum_{i=1}^{m_n} \sum_{j=1}^{m_n r_n} \mathrm{cov}\big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^t), \mathbb{1}(X_j > u_n^s)\big).$$

Once again we split the sum into two pieces, $I_{1,n} + I_{2,n}$, depending on whether $(i-1)r_n < j \le ir_n$ or not. By stationarity,

$$\begin{aligned}\frac{p_n}{m_n r_n} I_{1,n} &\sim \frac{p_n}{r_n}\sum_{i=1}^{r_n} \mathrm{cov}(\mathbb{1}(M_{r_n} > u_n^t), \mathbb{1}(X_i > u_n^s)) \\ &= \frac{p_n}{r_n}\sum_{i=1}^{r_n} P(M_{r_n} > u_n^t, X_i > u_n^s) - p_n P(M_{r_n} > u_n^t)P(X_1 > u_n^s) \\ &= \frac{p_n}{r_n} E[N_{r_n}(u_n^s)|M_{r_n} > u_n^t]P(M_{r_n} > u_n^t) - p_n P(M_{r_n} > u_n^t)P(X_1 > u_n^s) \\ &\to \tau_t \theta \psi_{s,t}\end{aligned}$$

as $n \to \infty$ by Eqs. 2.4, 2.5 and 2.7. Since $I_{2,n} \to 0$ as before, the proof of the proposition is complete. □

4 A central limit theorem for the multilevel estimator

In this section we establish the asymptotic normality of our multilevel estimator (2.3). We start by checking the consistency of the estimator. For notational convenience we restate definitions (3.1) and (3.2), with $m_n = k_n$:

$$\widehat{M}_n(u_n^s) = \widehat{M}_{n,k_n}(u_n^s) = \sum_{i=1}^{k_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{4.1}$$

and

$$\widehat{\tau}_n(u_n^s) = \widehat{\tau}_{n,k_n}(u_n^s) = \sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)\,. \tag{4.2}$$

Proposition 4.1 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Suppose that Assumptions C_1 and D_1 hold. Then as $n \to \infty$,*

$$\widehat{\theta}_n(f) \to_P \theta. \tag{4.3}$$

Proof Note that for $1 \le s \le m$, by Eq. 2.5,

$$E\left(\frac{p_n}{n}\widehat{M}_n(u_n^s)\right) = \frac{k_n p_n}{n} P(M_{r_n} > u_n^s) \to \tau_s \theta$$

as $n \to \infty$. Since $\mathrm{var}\big((p_n/n)\widehat{M}_n(u_n^s)\big) \to 0$ by Proposition 3.1, it follows that $(p_n/n)\widehat{M}_n(u_n^s) \to_P \tau_s\theta$ as $n \to \infty$.

Similarly, by Eq. 2.4 and Proposition 3.3 we have $(p_n/n)\widehat{\tau}_n(u_n^s) \to_P \tau_s$ as $n \to \infty$ for $1 \le s \le m$. In particular,

$$\widehat{\tau_s/\tau_m} \to_P \tau_s/\tau_m \quad \text{for } 1 \le s \le m,$$

and the result follows. □

The next theorem is the main result of this section. It establishes asymptotic normality of the estimator (2.3). It requires an assumption on the rate of convergence in Eq. 2.5. We assume that, as $n \to \infty$,

$$\sqrt{n/p_n}\big[(p_n/r_n)P(M_{r_n} > u_n^s) - \tau_s\theta\big] \to 0,\ \ 1 \le s \le m\,. \tag{4.4}$$

Such an assumption is sometimes associated with a sufficiently large block size r_n; see e.g. Robert et al. (2009).

Under the notation of Assumptions D_1 and D_2 we denote

$$\mu_2 := \sum_{j=1}^{\infty} j^2\pi_j,$$

$$\mu_{s,t} := \sum_{i=1}^{\infty}\sum_{j=0}^{i} ij\varpi_{s,t}(i,j),\ \ 1 \le s < t \le m.$$

Theorem 4.2 *Let $\{X_i\}$ be a stationary sequence with extremal index θ. Assume that Assumptions C_1', C_2, C_2', D_1 and D_2 hold. Assume further (4.4). Then as $n \to \infty$,*

$$\sqrt{n/p_n}\big(\widehat{\theta}_n(f) - \theta\big) \to_d \mathcal{N}(0, \sigma^2), \tag{4.5}$$

where $\sigma^2 = \mathbf{h}^T\mathbf{\Sigma}\mathbf{h}$, with a $(2m)\times(2m)$ covariance matrix $\mathbf{\Sigma}$ and a $2m$-dimensional vector $\mathbf{h}$ defined as follows: for $1 \le s \le t \le m$,

$$\sigma_{s,t} = \tau_t\theta,$$
$$\sigma_{m+s,m+t} = \tau_s\theta\mu_{s,t},$$
$$\sigma_{s,m+t} = \tau_t,$$
$$\sigma_{t,m+s} = \tau_t\theta\psi_{s,t}\,,$$

where $\mu_{s,s}$ is taken to be μ_2 for each s, while $\psi_{s,t}$ is defined by Eq. 2.7 *for $s < t$ and taken to be $1/\theta$ if $s = t$. Furthermore,*

$$h_s = \frac{f\big(\tau_s/\tau_m\big) - f\big(\tau_{s-1}/\tau_m\big)}{\sum_{t=1}^m \big(f\big(\tau_t/\tau_m\big) - f\big(\tau_{t-1}/\tau_m\big)\big)\tau_t},\quad 1 \le s \le m,$$

$$h_{m+s} = -\frac{\big(f\big(\tau_s/\tau_m\big) - f\big(\tau_{s-1}/\tau_m\big)\big)\theta}{\sum_{t=1}^m \big(f\big(\tau_t/\tau_m\big) - f\big(\tau_{t-1}/\tau_m\big)\big)\tau_t},\quad 1 \le s \le m,$$

where we set $\tau_0 = \infty$ and $f(\infty) = 0$.

Proof The argument is similar to that used in Theorem 4.2 of Robert et al. (2009). Notice that

$$\widehat{\theta}_n(f) = h\big((p_n/n)\widehat{M}_n(u_n^1), \ldots, (p_n/n)\widehat{M}_n(u_n^m), (p_n/n)\widehat{\tau}_n(u_n^1), \ldots, (p_n/n)\widehat{\tau}_n(u_n^m)\big),$$
$$\theta = h\big(\tau_1\theta, \ldots, \tau_m\theta, \tau_1, \ldots, \tau_m\big)\,,$$

where $h:\ [0,\infty)^m \times (0,\infty)^m \to [0,\infty)$ is defined by

$$h(x_1,\ldots,x_m,y_1,\ldots,y_m) = \frac{\sum_{s=1}^m\big(f(y_s/y_m) - f(y_{s-1}/y_m)\big)x_s}{\sum_{s=1}^m\big(f(y_s/y_m) - f(y_{s-1}/y_m)\big)y_s}\,.$$

Here and for the remainder of the proof we use the convention $y_0 = \infty$ and $f(\infty) = 0$. Since

$$\nabla h\big(\tau_1\theta, \ldots, \tau_m\theta, \tau_1, \ldots, \tau_m\big) = \mathbf{h}\,,$$

by the delta method we only need to prove that

$$\sqrt{n/p_n}\begin{pmatrix} (p_n/n)\widehat{M}_n(u_n^1) - \tau_1\theta \\ \vdots \\ (p_n/n)\widehat{M}_n(u_n^m) - \tau_m\theta \\ (p_n/n)\widehat{\tau}_n(u_n^1) - \tau_1 \\ \vdots \\ (p_n/n)\widehat{\tau}_n(u_n^m) - \tau_m \end{pmatrix} \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})\,. \tag{4.6}$$

We will, actually, prove the statement

$$\sqrt{n/p_n}\begin{pmatrix} (p_n/n)\big[\widehat{M}_n(u_n^1) - k_n P(M_{r_n} > u_n^1)\big] \\ \vdots \\ (p_n/n)\big[\widehat{M}_n(u_n^m) - k_n P(M_{r_n} > u_n^m)\big] \\ (p_n/n)\widehat{\tau}_n(u_n^1) - \tau_1 \\ \vdots \\ (p_n/n)\widehat{\tau}_n(u_n^m) - \tau_m \end{pmatrix} \to_d \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})\,. \tag{4.7}$$

By Eq. 4.4 this will imply (4.6).

We present an argument for the case $m = 2$. The argument for larger values of m is only notationally different. Denote by $Z_{n,i}$, $i = 1, 2, 3, 4$ the 4 entries in the vector in the left hand side of Eq. 4.7. By the Cramér-Wold device it suffices to show that for any $\mathbf{a} = (a_1, a_2, a_3, a_4)^T \in \mathbb{R}^4$, as $n \to \infty$,

$$a_1 Z_{n,1} + a_2 Z_{n,2} + a_3 Z_{n,3} + a_4 Z_{n,4} \to_d \mathcal{N}(0, \mathbf{a}^T \boldsymbol{\Sigma}\mathbf{a}). \tag{4.8}$$

Denote $m_n = \lfloor n/p_n \rfloor$ and let $h_n = \lfloor k_n/m_n \rfloor$ and write

$$\begin{aligned} Z_{n,1} &= \sqrt{\frac{p_n}{n}}\sum_{i=1}^{h_n} \bar{I}_i(u_n^1) + o_p(1),\;\; Z_{n,2} = \sqrt{\frac{p_n}{n}}\sum_{i=1}^{h_n} \bar{I}_i(u_n^2) + o_p(1) \\ Z_{n,3} &= \sqrt{\frac{p_n}{n}}\sum_{i=1}^{h_n} \bar{J}_i(u_n^1) + o_p(1),\;\; Z_{n,4} = \sqrt{\frac{p_n}{n}}\sum_{i=1}^{h_n} \bar{J}_i(u_n^2) + o_p(1), \end{aligned}$$

where

$$\bar{I}_i(u_n^1) = \sum_{j=(i-1)m_n}^{im_n-1} \big(\mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^1) - P(M_{r_n} > u_n^1)\big),\ \bar{I}_i(u_n^2)$$
$$= \sum_{j=(i-1)m_n}^{im_n-1} \big(\mathbb{1}(M_{(j-1)r_n,jr_n} > u_n^2) - P(M_{r_n} > u_n^2)\big),\ \bar{J}_i(u_n^1)$$
$$= \sum_{j=(i-1)m_nr_n}^{im_nr_n-1} \big(\mathbb{1}(X_j > u_n^1) - \tau_1/p_n\big),\ \bar{J}_i(u_n^2)$$
$$= \sum_{j=(i-1)m_nr_n}^{im_nr_n-1} \big(\mathbb{1}(X_j > u_n^2) - \tau_2/p_n\big).$$

Let $h_n^* \to \infty$ be a sequence of integers with $(h_n^*)^2 = o(h_n)$, $h_n = o\big((h_n^*)^3\big)$. Partition the set $\{1, \ldots, h_n\}$ into subsets of length h_n^* of consecutive integers, with two adjacent such subsets separated by a singleton. The number of subsets of length h_n^* is then $q_n = \lfloor (h_n+1)/(h_n^*+1) \rfloor$. We have

$$\sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \bar{I}_i(u_n^1) = \sqrt{\frac{p_n}{n}} \sum_{j=1}^{q_n} \sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1} \bar{I}_i(u_n^1) + \sqrt{\frac{p_n}{n}} \sum_{j=1}^{q_n-1} \bar{I}_{j(h_n^*+1)}(u_n^1) + \sqrt{\frac{p_n}{n}} \sum_{i=q_n(h_n^*+1)}^{h_n} \bar{I}_i(u_n^1). \tag{4.9}$$

The variance of the second term is bounded by

$$\frac{p_nq_n}{n}\mathrm{var}(\bar{I}_1(u_n^1)) + \frac{p_nq_n^2}{n}\rho_{n,h_n^*r_n}(u_n^1)\mathrm{var}(\bar{I}_1(u_n^1)).$$

By Proposition 3.1 the first entry above does not exceed a constant multiple of

$$\frac{p_nq_n}{n}\frac{m_nr_n}{p_n} \sim \frac{1}{h_n^*} \to 0$$

since $h_n^* \to \infty$. Since Assumption C_1 is in force,

$$\rho_{n,h_n^*r_n}(u_n^1) = \frac{1}{h_n^*r_n} h_n^*r_n\rho_{n,h_n^*r_n}(u_n^1) \le \frac{1}{h_n^*r_n} \sum_{l=1}^{n} \rho_{n,l}(u_n^1) = \mathrm{o}\left(\frac{1}{h_n^*}\right).$$

Therefore, the second entry above does not exceed a constant multiple of

$$\frac{p_nq_n^2}{n}\frac{1}{h_n^*}\frac{m_nr_n}{p_n} \sim \frac{h_n}{(h_n^*)^3} \to 0$$

by the choice of h_n^*. Hence it follows that the variance of the second term in Eq. 4.9 converges to zero. Further, the variance of the third term in Eq. 4.9 is, apart from a multiplicative constant, bounded by

$$\frac{p_n(h_n^*)^2}{n}\mathrm{var}(\bar{I}_1(u_n^1)) \sim \frac{p_n(h_n^*)^2}{n}\frac{m_n r_n}{p_n} \sim \frac{(h_n^*)^2}{h_n} \to 0,$$

once again by the choice of h_n^*. Therefore, we can write

$$Z_{n,1} = \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\left(\sqrt{\frac{p_n q_n}{n}}\sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1}\bar{I}_i(u_n^1)\right) + o_p(1) =: \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j,1} + o_p(1).$$

Similarly,

$$Z_{n,2} = \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\left(\sqrt{\frac{p_n q_n}{n}}\sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1}\bar{I}_i(u_n^2)\right) + o_p(1) =: \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j,2} + o_p(1),$$

$$Z_{n,3} = \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\left(\sqrt{\frac{q_n}{h_n}}\sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1}\bar{J}_i(u_n^1)\right) + o_p(1) =: \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j,3} + o_p(1),$$

$$Z_{n,4} = \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\left(\sqrt{\frac{q_n}{h_n}}\sum_{i=(j-1)(h_n^*+1)+1}^{j(h_n^*+1)-1}\bar{J}_i(u_n^2)\right) + o_p(1) =: \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j,4} + o_p(1).$$

Writing $\xi_{n,j} = a_1\xi_{n,j,1} + a_2\xi_{n,j,2} + a_3\xi_{n,j,3} + a_4\xi_{n,j,4}$, we conclude that

$$a_1 Z_{n,1} + a_2 Z_{n,2} + a_3 Z_{n,3} + a_4 Z_{n,4} = \frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j} + o_p(1).$$

Notice that for fixed n the elements of the stationary sequence defining each pair of $\xi_{n,i}$ and $\xi_{n,j}$, $i \neq j$, are separated by at least $h_n^* r_n$ entries. Furthermore, by Assumptions C_1 and C_2,

$$\rho_{n,h_n^* r_n}(u_n^1, u_n^2) = o(1/h_n) = o(1/q_n)\,.$$

Since for any real θ

$$\left|E\exp\left\{i\theta\frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n}\xi_{n,j}\right\} - \prod_{j=1}^{q_n}E\exp\left\{i\theta\frac{1}{\sqrt{q_n}}\xi_{n,j}\right\}\right|$$
$$\leq \sum_{k=1}^{q_n}\left|E\exp\left\{i\theta\frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n-k+1}\xi_{n,j}\right\} - E\exp\left\{i\theta\frac{1}{\sqrt{q_n}}\sum_{j=1}^{q_n-k}\xi_{n,j}\right\}\right.$$
$$\left. E\exp\left\{i\theta\frac{1}{\sqrt{q_n}}\xi_{n,q_n-k+1}\right\}\right| \leq q_n\rho_{n,h_n^* r_n}(u_n^1, u_n^2)$$

up to a multiplicative constant, the statement (4.8) will follow once we prove that

$$\frac{1}{\sqrt{q_n}} \sum_{j=1}^{q_n} Y_{n,j} \to_d \mathcal{N}(0, \mathbf{a}^T \mathbf{\Sigma} \mathbf{a}) , \tag{4.10}$$

where for each n, $Y_{n,j}$, $j = 1, \ldots, q_n$ are i.i.d. random variables with the same law as $\xi_{n,1}$. Since Propositions 3.1 – 3.7 tell us that $\mathrm{var}(\xi_{n,1}) \to \mathbf{a}^T \mathbf{\Sigma} \mathbf{a}$ as $n \to \infty$, by the Lindeberg-Feller central limit theorem the convergence in Eq. 4.10 will follow once we check that for any $\varepsilon > 0$,

$$E\big(\xi_{n,1}^2 \mathbb{1}(|\xi_{n,1}| > \varepsilon q_n^{1/2})\big) \to 0$$

as $n \to \infty$, which reduces to showing that

$$E\big(\xi_{n,1,i}^2 \mathbb{1}(|\xi_{n,1,j}| > \varepsilon q_n^{1/2})\big) \to 0 \tag{4.11}$$

for each $\varepsilon > 0$ and each pair $i, j = 1, 2, 3, 4$. We will check (4.11) for $i = j = 1$. All other combinations of i, j can be treated in a similar way. If $\widehat{M}_n^*(u_n^1)$ is defined by Eq. 3.1 with m_n replaced by $m_n h_n^*$, then we have to check that

$$\frac{p_n q_n}{n} E\left((\widehat{M}_n^*(u_n^1))^2 \mathbb{1}\big(|\widehat{M}_n^*(u_n^1)| > \varepsilon\sqrt{n/p_n}\big)\right) \to 0 \, .$$

While proving Proposition 3.1 we decomposed the variance of $\widehat{M}_n^*(u_n^1)$ into a sum of two terms, the second of which is of a smaller order than the first one. Therefore, we only need to prove that

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} E\Bigg[\Big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\Big)^2$$
$$\mathbb{1}\left(\left|\sum_{j=1}^{m_n h_n^*} \big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\right)\Bigg] \to 0$$

and, since $n/p_n \to \infty$, by changing $\varepsilon > 0$ to a smaller positive number, we only need to show that

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} P\left(M_{(i-1)r_n, ir_n} > u_n^1, \left|\sum_{|j-i|\geq 2} \big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\right) \to 0 \, .$$

Note that the expression in the left hand side above can be bounded by

$$\frac{p_n q_n}{n} \sum_{i=1}^{m_n h_n^*} P\big(M_{(i-1)r_n, ir_n} > u_n^1\big) P\left(\left|\sum_{|j-i|\geq 2} \big(\mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^1) - P(M_{r_n} > u_n^1)\big)\right| > \varepsilon\sqrt{n/p_n}\right)$$
$$+ \frac{p_n q_n}{n} m_n h_n^* \alpha_{n, r_n}(u_n^1) \, .$$

The first term above converges to zero as $n \to \infty$ by Proposition 1, while the second term converges to zero as $n \to \infty$ by Assumption C_1. Therefore, the convergence in Eq. 4.10 has been established. □

Remark 4.3 Note that without the assumption (4.4) what Theorem 4.2 proves is that

$$\sqrt{n/p_n}\big(\widehat{\theta}_n(f) - \theta_n(f)\big) \to_d \mathcal{N}(0, \sigma^2).$$

The difference $\theta_n(f) - \theta$ is then responsible for the bias of our estimator.

5 Testing the estimator

This section is devoted to testing the effect of using multiple thresholds in the blocks estimator as in Eq. 2.3, both on simulated data and real data. As in many cases of extremal inference, we should address the question of the bias of the estimator; see, in particular, Remark 4.3. One approach of tackling the bias is to build a simple model for it and then estimate it from the data. Following (Drees 2011), and to further account for the effect of block size r_n, we assume that the main terms in the bias of $\widehat{M}_n(u_n^s)/\widehat{\tau}_n(u_n^s)$ as an estimator of θ are linear in τ_s/k_n and $1/r_n$, $s = 1, \ldots, m$. Since we estimate τ_s by a scaled version of the statistics $\widehat{\tau}_n(u_n^s)$, it is natural to use the following bias-corrected version of the multilevel estimator:

$$\widehat{\theta}_n^b(f) = \frac{\sum_{s=1}^m \big[f\big(\widehat{\tau_s/\tau_m}\big) - f\big(\widehat{\tau_{s-1}/\tau_m}\big)\big](\widehat{M}_n(u_n^s) - \widehat{\beta}_1 \frac{\widehat{\tau}_n(u_n^s)^2}{k_n} - \widehat{\beta}_2 \frac{\widehat{\tau}_n(u_n^s)}{r_n})}{\sum_{s=1}^m \big[f\big(\widehat{\tau_s/\tau_m}\big) - f\big(\widehat{\tau_{s-1}/\tau_m}\big)\big]\widehat{\tau}_n(u_n^s)}, \tag{5.1}$$

where $\widehat{\beta}_1, \widehat{\beta}_2$ are coefficients estimated from the data. We simply use linear regression as follows.

Use the m levels $u_n^1, \ldots, u_n^m$ and l values of block sizes $r_n^1, \ldots, r_n^l$ to compute the values of $\widehat{M}_n(u_n^s, r_n^i)$, $\widehat{\tau}_n(u_n^s, r_n^i)$ and $\widehat{\theta}_n(u_n^s, r_n^i) = \widehat{M}_n(u_n^s, r_n^i)/\widehat{\tau}_n(u_n^s, r_n^i)$ for $s = 1, \ldots, m$, $i = 1, \ldots, l$, where $\widehat{M}_n(u_n^s, r_n^i)$, $\widehat{\tau}_n(u_n^s, r_n^i)$ respectively denote the quantities $\widehat{M}_n(u_n^s)$ and $\widehat{\tau}_n(u_n^s)$ evaluated using block size r_n^i. Now fit a regression plane to the response variables $\widehat{\theta}_n(u_n^s, r_n^i)$ using the predictor variables $\big(\widehat{\tau}_n(u_n^s, r_n^i)/k_n^i\,, 1/r_n^i\big)$, $s = 1, \ldots, m$, $i = 1, \ldots, l$, where $k_n^i = \lfloor \frac{n}{r_n^i} \rfloor$. Specifically, we use the least squares coefficients

$$\big(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2\big)^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\boldsymbol{\theta}}_n\,, \tag{5.2}$$

where

$$\widehat{\boldsymbol{\theta}}_n = \Big(\widehat{\theta}_n(u_n^1, r_n^1)\ \ldots\ \widehat{\theta}_n(u_n^m, r_n^l)\Big)^T,$$

and

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \widehat{\tau}_n(u_n^1, r_n^1)/k_n^1 & \widehat{\tau}_n(u_n^2, r_n^1)/k_n^1 & \ldots & \widehat{\tau}_n(u_n^m, r_n^l)/k_n^l \\ 1/r_n^1 & 1/r_n^1 & \ldots & 1/r_n^l \end{pmatrix}^T,$$

where $\widehat{\boldsymbol{\theta}}_n$ is a vector of length ml and $\mathbf{X}$ is a matrix of dimension $ml \times 3$. We use $\widehat{\beta}_1, \widehat{\beta}_2$ in Eq. 5.2 as desired coefficients in Eq. 5.1. Alternatively, one could estimate the coefficients using levels different from the collection $u_n^1, \ldots, u_n^m$.

Remark 5.1 A limiting theory for the bias-corrected estimator $\widehat{\theta}_n^b(f)$ can be developed, but it requires a number of additional assumptions and a fairly long argument.

The main idea is similar to that in Drees (2011), and it relies on concentrating on the "leading terms" in the bias. In order to keep the paper readable we have chosen not to include this theory here. It can be found in Sun (2018).

Remark 5.2 Note that $\widehat{\beta}_0$ in Eq. 5.2 is itself an estimator for θ. We have not studied its statistical properties, but it performs well on simulated data.

In the sequel we test the blocks estimator with multiple thresholds (2.3) and its bias-corrected version (5.1) on simulated data and on S&P 500 Daily Log Returns. As it is invariably done in practice, we use random thresholds given by different order statistics of the observations. In a similar situation, it was shown in Corollary 2.4 of Drees (2011) that, under certain continuity assumptions, this has no effect on the asymptotic distribution of the estimator.

5.1 Simulation Study

We have drawn samples from ARMAX processes. Specifically, we use the ARMAX(1) process (X_i) is defined as follows. Let $Z_1, Z_2, \dots$ be a sequence of i.i.d. unit Fréchet random variables with shape 1. For $0 < \theta \leq 1$ a stationary sequence is obtained by letting $X_1 = Z_1/\theta$, and

$$X_i = \max((1-\theta)X_{i-1}, Z_i),\ i \geq 2\,. \tag{5.3}$$

It can be shown that the extremal index of such a sequence is θ; see e.g. Chapter 10 of Beirlant et al. (2006).

We first test the performance of the estimators (2.3) and Eq. 5.1 on the ARMAX model using values of $\theta = 0.25, 0.5, 0.75$, and a sample of length $n = 10000$. For the estimator, we have chosen a block size of $r_n = 200$, and a weight function of $f(x) = e^{-x}$. We run the experiments for $m = 1, \dots, 20$, and for each fixed m we choose u_n^s to be equal to the $(101+2(m-s))$-th largest order statistic of the sequence, $1 \leq s \leq m$. That is, each level incorporates 2 more observations above it than the level immediately above it does. When computing coefficients for the bias-reduced estimator (5.1), we use $m' = 12$, with $\bar{u}_n^s$ being the $(91 + 5(m' - s))$-th largest order statistic of the sequence, $1 \leq s \leq m'$, and $l = 25$, with $r_n^i = 10i$, $1 \leq i \leq l$.

We compare the estimators $\widehat{\theta}(f)$, $\widehat{\theta}^b(f)$ and the plain blocks estimator on the basis of their bias, standard error, and root mean squared error. The results computed from 5000 simulated sequences are displayed in Fig. 1. Looking from the top row to the bottom row along the varying values of $\theta = 0.25, 0.5, 0.75$, we see that the plots tell a similar story. As expected for the multilevel estimator $\widehat{\theta}(f)$, the magnitude of the bias increases while the standard error decreases as more levels of observations are incorporated into the estimator. The bias of the bias-corrected version of the estimator, $\widehat{\theta}^b(f)$, seems to be largely insensitive to the choice of m, with a decreasing trend both in the standard error and in the root mean squared error. Overall, $\widehat{\theta}^b(f)$ achieves a much better root mean squared error compared to $\widehat{\theta}(f)$, for all levels m considered. It also outperforms the plain blocks estimator in the sense of an improved root mean squared error.

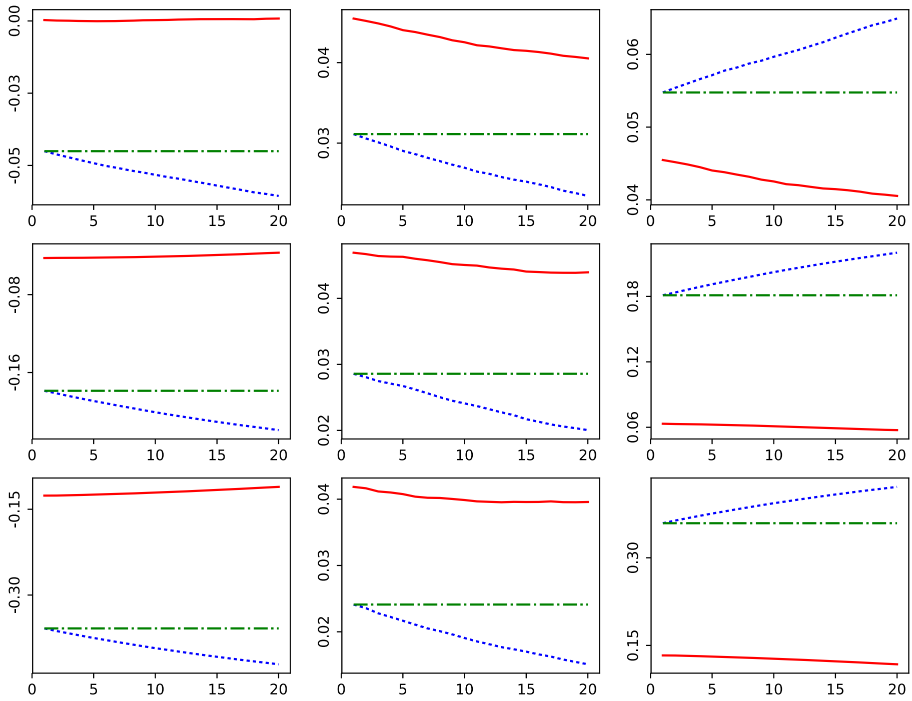

Fig. 1 Bias (left column), standard error (center column), and root mean squared error (right column) for the blocks estimator (1.2) (dot-dash line), the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom)

We have performed the same analysis for other models, for different values of the highest threshold and for different block sizes. We have also analyzed the non-clustering case $\theta = 1$. Invariably, the qualitative structure seen on Fig. 1 remained the same. In the remaining experiments in this section we will, therefore, focus on the best performing bias-corrected estimator $\widehat{\theta}^b(f)$ that uses the largest amount of data ($m = 20$ levels).

Our next experiment addresses the effect of the choice of block size r_n on the performance of the estimator $\widehat{\theta}^b(f)$. We again use the ARMAX model with $\theta = 0.25, 0.5, 0.75$ as before. We test the performance of $\widehat{\theta}^b(f)$ using block sizes of $r_n = 40, 50, \ldots, 200$. The root mean squared errors from 5000 simulated sequences are displayed in Fig. 2. We see that the choice of the block size does not have a major effect on the root mean squared error. We have also looked at the effect of the block size on the bias and standard error separately (not shown). Once again, the standard error is largely insensitive to the choice of the block size. The bias does vary with the block size, but remains invariably small in the absolute value, leading to the root mean squared errors displayed in Fig. 2.

In the next experiment we fix the the block size to $r_n = 200$ and study the effect of the choice of the weight function. In the setting of the previous experiments we

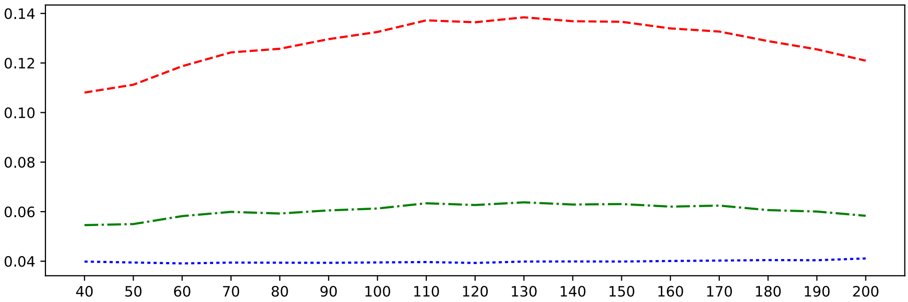

Fig. 2 Root mean squared error for the estimator $\widehat{\theta}^b(f)$ for true values of θ being 0.25 (dotted line), 0.5 (dot-dash line), and 0.75 (dash line) plotted against the choice of r_n, the size of the blocks used in the estimator. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$

use a second weight function, $f_1(x) = 1/x^{20}$ along with the original weight function f. In the relevant range f_1 decreases at a much faster rate than f. We compare the performance of the estimators $\widehat{\theta}^b(f)$ and $\widehat{\theta}^b(f_1)$. The results are presented in Fig. 3.

As in Fig. 1 the magnitude of the bias, the standard error, and the root mean squared error of the estimators are all decreasing when m, the number of levels used in the estimator, increases. The phenomenon displayed in Fig. 3 demonstrates that the faster decay of the weight function f_1 compared to f leads to smaller contributions from additional levels to the efficiency of the estimator. However, the exact overall effect of the weight function on the estimator is a topic not studied in detail in this paper. It warrants further investigation.

In the previous experiments we have used samples of size $n = 10000$. Sometimes extremal inference has to be performed on data sets of a smaller size, so we have

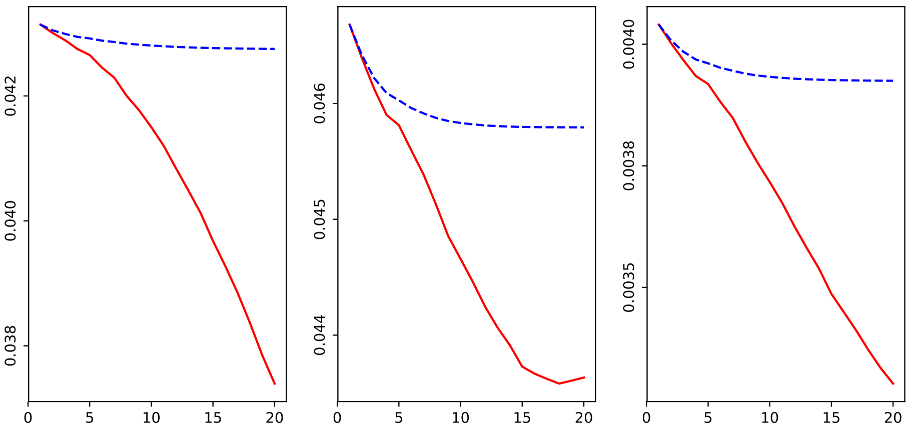

Fig. 3 Bias (left), standard error (center) and root mean squared error (right) for the bias corrected multilevel estimators $\widehat{\theta}^b(f)$ (solid line) and $\widehat{\theta}^b(f_1)$ (dotted line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 0.5$

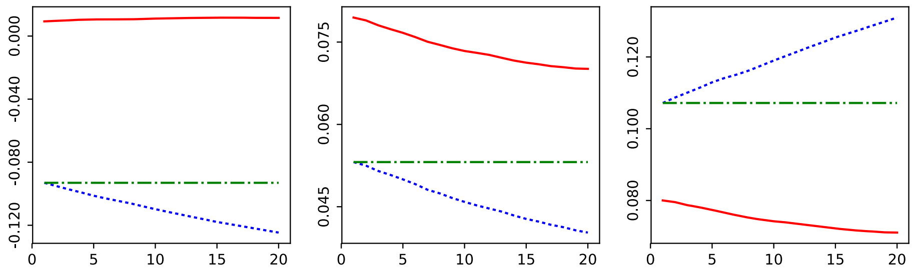

Fig. 4 Bias (left), standard error (center), and root mean squared error (right) for the naive blocks estimator (1.2) (dot-dash line), the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 0.5$

repeated our experiment leading to Fig. 1 for samples of size $n = 5000$. We only display the results for the ARMAX model with $\theta = 0.5$. We use $r_n = 100$, and $f(x) = e^{-x}$. Once again, we experiment with $m = 1, \ldots, 20$ levels, and for each fixed m we choose u_n^s to be equal to the $(51 + m - s)$-th largest order statistic of the

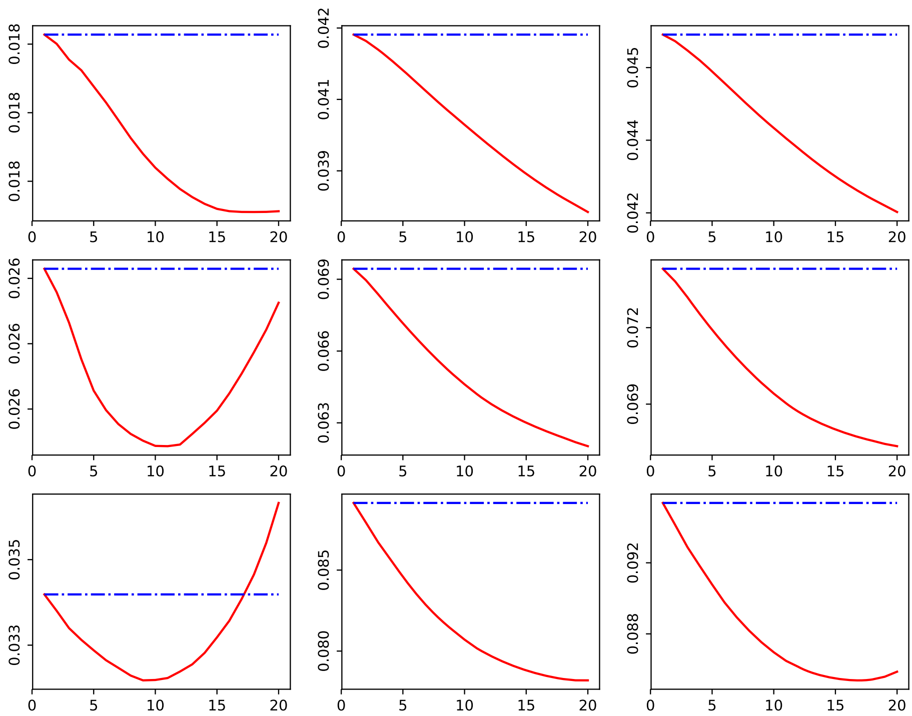

Fig. 5 Bias (left column), standard error (center column), and root mean squared error (right column) for the sliding blocks estimator (dot-dash line), the multilevel sliding blocks estimator (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom)

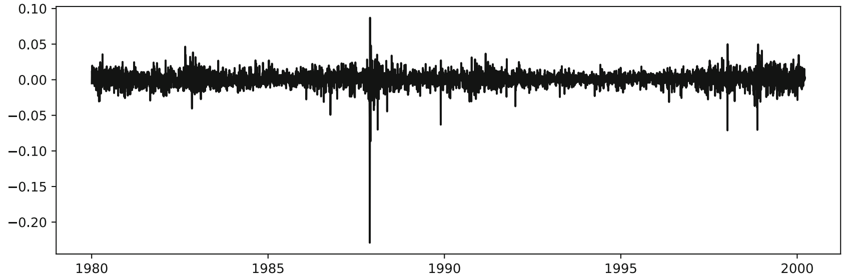

Fig. 6 Daily Log Returns for S&P 500 from 1980 - 1999

sequence, $1 \leq s \leq m$. When computing coefficients for the bias-reduced estimator, we use $m' = 12$, with $\bar{u}_n^s$ being the $(41 + 3(m' - s))$-th largest order statistic of the sequence, $1 \leq s \leq m'$, and $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$. The results from 5000 simulated sequences are displayed, in Fig. 4. As expected, the smaller sample size leads to some deterioration in the quality of the estimation in comparison with the large sample size used in Fig. 1, but the comparison of the estimators and the lessons derived from both figures remain the same.

Finally, we experiment with constructing a multiple threshold version of an estimator different from the plain blocks estimator. We have chosen the sliding blocks estimator of Robert et al. (2009). We use the ARMAX models with $\theta = 0.25, 0.5, 0.75$. For each simulated sequence, we first compute the optimal threshold as described in Robert et al. (2009), then choose $m = 1, \ldots, 20$, where u_n^m corresponds to the level of the optimal threshold, and for each $1 \leq s \leq m$, the level corresponding to u_n^s incorporates 10 more observations than the level immediately above it. The results from 5000 simulated sequences are displayed in Fig. 5. Once again we see that the root mean squared error is almost invariably decreasing with increasing number of levels m.

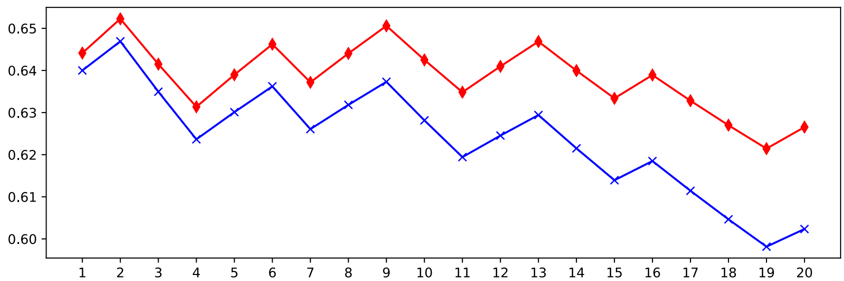

Fig. 7 The values of the multilevel estimator $\widehat{\theta}(f)$ ('x' marker) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (diamond marker) plotted against the choice of m, number of levels used in the estimators, for the negative daily log returns of S&P 500

5.2 S&P 500 Daily Log Returns

We now use the estimators developed in this paper to estimate the extremal index of the losses among the daily log returns for S&P 500 during the ten-year period between 1 January 1990 and 31 December 1999. The log returns themselves are plotted in Fig. 6.

There are $n = 5055$ returns in this data set, and the negative of their values form our sample. We choose $m = 1, \ldots, 20$ and u_n^s to be the $51 + (m - s)$-th largest order statistic, $1 \leq s \leq m$. We choose the block size $r_n = 40$, resulting in $k_n = 126$ blocks. For the weight function we use $f(x) = e^{-x}$. When computing the bias-corrected estimator we use Eq. 5.2 with $m' = 12$ levels, $\bar{u}_n^s$ being the $41 + 3(m' - s)$-th largest order statistic in the sample, $1 \leq s \leq m'$, and set $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$.

The plots of the two estimators are shown above as a function of the number of levels m. We have also evaluated the variability of the estimators by performing a block-level bootstrap. We have not presented the resulting pointwise 1-standard error confidence intervals on Fig. 7 since this makes the structure of the pointwise estimators harder to see, but the order of magnitude of these intervals is [0.5, 0.8].

Acknowledgments We are enormously grateful to the Associate Editor and three anonymous referees for their comments that helped us to sharpen both the focus and the presentation of the paper.

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels J.: Statistics of extremes: theory and applications, Wiley (2006)
Berghaus, Bücher: Weak convergence of a pseudo maximum likelihood estimator for the extremal index, arXiv:1608.01903 (2017)
de Haan, L., Ferreira, A.: Extreme Value Theory: an Introduction. Springer, New York (2006)
Drees, H.: Bias correction for estimators of the extremal index, arXiv:1107.0935 (2011)
Drees, H., Kaufmann, E.: Selecting the optimal sample fraction in univariate extreme value estimation. Stoch. Process. Appl. **75**, 149–172 (1998)
Dupuis, D.: Exceedances over high thresholds: a guide to threshold selection. Extremes **1**, 251–261 (1998)
Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events for insurance and finance. Springer, Berlin (1997)
Ferro, C., Segers, J.: Inference for clusters of extreme values. J. R. Stat. Soc. Ser. B Methodol. **65**, 545–556 (2003)
Hsing, T.: Estimating the parameters of rare events. Stoch. Process. Appl. **37**, 117–139 (1991)
Hsing, T.: Extremal index estimation for a weakly dependent stationary sequence. Ann. Stat. **21**, 2043–2071 (1993)
Hsing, T., Hüsler, J., Leadbetter, M.: On the exceedance point process for a stationary sequence. Probab. Theory Relat. Fields **78**, 97–112 (1988)
Laurini, F., Tawn, J.A.: New estimators for the extremal index and other cluster characteristics. Extremes **6**, 189–211 (2003)
Leadbetter, M., Lindgren, G., Rootzén, H.: Extremes and related properties of random sequences and processes. Springer, New York (1983)
Nguyen, T., Samorodnitsky, G.: Tail inference: where does the tail begin? Extremes **15**, 437–461 (2012)
Northrop, P.: An efficient semiparametric maxima estimator of the extremal index. Extremes **18**, 585–603 (2015)
Resnick, S.: Heavy-tail phenomena: probabilistic and statistical modeling. Springer, New York (2007)
Resnick, S., Stărică, C.: Smoothing the Hill estimator. Adv. Appl. Probab. **20**, 271–293 (1997)

Robert, C.: Inference for the limiting cluster size distribution of extreme values. Ann. Stat. **37**, 271–310 (2009)

Robert, C., Segers, J., Ferro, C.: A sliding block estimator for the extremal index. Electron. J. Stat. **3**, 993–1020 (2009)

Smith, R., Weissman, I.: Estimating the extremal index. J. R. Stat. Soc. Ser. B Methodol. **56**, 515–528 (1994)

Sun, J.: Modelling and Inference for Extremal Evnts: Methods and Techniques, Ph.D. thesis, Cornell University, available at http://hdl.handle.net/1813/58726 (2018)

Weissman, I., Novak, S.: On blocks and runs estimators of the extremal index. J. Stat. Planning and Inference **66**, 281–288 (1998)