ESTIMATING THE EXTREMAL INDEX, OR, CAN ONE ALLEVIATE THE THRESHOLD-SELECTION DIFFICULTY IN EXTREMAL INFERENCE?

JULIAN SUN AND GENNADY SAMORODNITSKY

Abstract. The question of selecting the number of upper order statistics to use in extremal inference or selecting the threshold above which we perform the extremal inference is one of the most common and most difficult questions in applications of extreme value theory. We propose an approach consisting of using multiple thresholds instead of a single threshold as a means to alleviate the consequences of making this choice. We apply this approach to the problem of estimating the extremal index and demonstrate its power both on simulated and real data.

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 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 (indeed, often an overwhelmingly 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.

This present work results from our belief 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".

Multiple thresholds in extremal inference have been used before, such as in Drees (2011), Laurini et al. (2003), and Süveges (2007). In Drees (2011), for example, estimates of the extremal index based on multiple thresholds were combined together in order to minimize the bias of the estimator. 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

2010 *Mathematics Subject Classification.* Primary 62G32. Secondary 60G70.

Key words and phrases. extremal inference, regular variation, threshold selection, extremal index, bias .

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

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.

In this paper we construct such a procedure for estimating the *extremal index*, a quantity designed to measure the amount of clustering of the extremes in a stationary sequence. Suppose that $X_1, X_2, \dots$ is a stationary sequence of random variables with a marginal distribution function F, and let $M_n = \max(X_1, \dots, X_n)$, $n = 1, 2, \dots$. 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 $n \to \infty$. Then θ is called the extremal index of the sequence $X_1, X_2, \dots$; it is automatically in the range $0 \leq \theta \leq 1$; see 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 $0 < \Gamma_1 < \Gamma_2 < \dots$ constitute a homogeneous Poisson process with intensity $\tau\theta$ on $[0, 1]$ which is independent of an 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 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 $X_1, X_2, \dots, X_n$ 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 $\max\{X_{i+1}, \dots, X_j\}$ 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 (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

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

with the convention that $f\big(\widehat{\tau_0/\tau_m}\big) = 0$. Note that when $m = 1$, (2.3) corresponds to (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

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

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

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

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,

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

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

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

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

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 $s = 1, \dots, m$.

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

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

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 $s = 1, \dots, m$.

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

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

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 $s = 1, \dots, m$.

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

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

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 $s, t = 1, \dots, m$.

Assumption $\mathbf{C_2'}$ For each $s, t = 1, \dots, m$,

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

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 $s, t = 1, \dots, m$.

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 $(\pi_j)_{j \ge 1}$ 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 $(\varpi_{s,t}(i,j))_{i\geq 1, j\geq 0}$ on $\mathbb{Z}_+ \times \mathbb{Z}_{\geq 0}$ such that for all $1 \leq s < t \leq m$,

$$P(N_{r_n}(u_n^s) = i, N_{r_n}(u_n^t) = j | M_{r_n} > u_n^s) \to \varpi_{s,t}(i,j),\ i \geq j \geq 0,\ i \geq 1,$$

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

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. Intuitively the mixing conditions Δ', C_1, C_1', C_2, C_2' only impose that the size of the blocks r_n must be "large enough". Many models, such as m-dependent sequences, or geometrically mixing sequences can easily satisfy those assumptions.

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 (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 $m_n = k_n$. For convenience, we denote

$$\widehat{M}_n(u_n^s) = \sum_{i=1}^{k_n} \mathbb{1}(M_{(i-1)r_n, ir_n} > u_n^s) \tag{3.3}$$

and

$$\widehat{\tau}_n(u_n^s) = \sum_{i=1}^{k_n r_n} \mathbb{1}(X_i > u_n^s)\,. \tag{3.4}$$

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 (2.4), and suppose that Assumption C_1 holds. Then for $1 \le s \le m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n}\mathrm{var}(\widehat{M}_{n,m_n}(u_n^s)) \to \tau_s\theta. \tag{3.5}$$

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

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

It follows from (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 (2.5) and Assumption C_1, so it remains to consider $I_{2,n}$. By (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 (2.5),

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

so we have both

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

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 (2.4), and suppose that Assumption C_2 holds. Then for $1 \leq s < t \leq m$, as $n \to \infty$,*

$$\frac{p_n}{m_n r_n} \mathrm{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{M}_{n,m_n}(u_n^t)) \to \tau_t \theta. \tag{3.6}$$

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 $1 \leq s < t \leq m$. 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 $1 \leq s \leq m$,*

$$\frac{p_n}{m_n r_n} \mathrm{var}(\widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_s \theta \sum_{j=1}^{\infty} j^2 \pi_j\,. \tag{3.7}$$

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

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

It follows from (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,

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

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,

$$\begin{aligned}E\big(N_{r_n} N_{r_n,2r_n}\big) = & E\big(N_{r_n-l_n} N_{r_n,2r_n}\big) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big) \\ \leq & (EN_{r_n})^2 + E(N_{r_n})^2 \rho_{n,l_n}(u_n^s) + E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big).\end{aligned}$$

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

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

as $n \to \infty$. Therefore,

$$\begin{aligned}k_n E\big(N_{r_n-l_n,r_n} N_{r_n,2r_n}\big) \leq & k_n \left(E(N_{l_n})^2\right)^{1/2} \left(E(N_{r_n})^2\right)^{1/2} \\ \leq & k_n E(N_{r_n})^2 \left(\frac{E(N_{l_n})^2}{E(N_{r_n})^2}\right)^{1/2} \to 0\end{aligned}$$

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 $1 \leq s < t \leq m$ 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 $1 \leq s < t \leq m$,*

$$\frac{p_n}{m_n r_n}\mathrm{cov}(\widehat{\tau}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_s \theta \sum_{i=1}^{\infty}\sum_{j=0}^{i} ij\varpi_{s,t}(i,j). \tag{3.8}$$

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 $1 \leq s \leq m$,*

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

Proof. Fix $1 \leq s \leq m$, we have

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

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,

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

by (2.4) and (2.5).

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

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

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

The asymptotic behaviour of $\operatorname{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t))$ 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 $1 \le s < t \le m$,*

$$\frac{p_n}{m_n r_n} \operatorname{cov}(\widehat{M}_{n,m_n}(u_n^s), \widehat{\tau}_{n,m_n}(u_n^t)) \to \tau_t. \tag{3.10}$$

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

$$\frac{p_n}{m_n r_n} \operatorname{cov}(\widehat{M}_{n,m_n}(u_n^t), \widehat{\tau}_{n,m_n}(u_n^s)) \to \tau_t \theta \psi_{s,t}, \tag{3.11}$$

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

Proof. As before,

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

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,

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

as $n \to \infty$ by (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, recall the following two definitions.

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

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 (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 (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 (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* $1 \le s \le t \le m$*,*

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

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

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

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

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

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

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

We will, actually, prove the statement

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

By (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 (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 $n \to \infty$,

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

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

$$Z_{n,1} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \bar{I}_i(u_n^1) + \mathrm{o}_p(1), \;\; Z_{n,2} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \bar{I}_i(u_n^2) + \mathrm{o}_p(1)$$

$$Z_{n,3} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \bar{J}_i(u_n^1) + \mathrm{o}_p(1), \;\; Z_{n,4} = \sqrt{\frac{p_n}{n}} \sum_{i=1}^{h_n} \bar{J}_i(u_n^2) + \mathrm{o}_p(1),$$

where

$$\bar{I}_i(u_n^1) = \sum_{j=(i-1)m_n}^{im_n-1} \left(\mathbb{1}(M_{(j-1)r_n, jr_n} > u_n^1) - P(M_{r_n} > u_n^1) \right),$$

$$\bar{I}_i(u_n^2) = \sum_{j=(i-1)m_n}^{im_n-1} \left(\mathbb{1}(M_{(j-1)r_n, jr_n} > u_n^2) - P(M_{r_n} > u_n^2) \right),$$

$$\bar{J}_i(u_n^1) = \sum_{j=(i-1)m_n r_n}^{im_n r_n-1} \left(\mathbb{1}(X_j > u_n^1) - \tau_1/p_n \right),$$

$$\bar{J}_i(u_n^2) = \sum_{j=(i-1)m_n r_n}^{im_n r_n-1} \left(\mathbb{1}(X_j > u_n^2) - \tau_2/p_n \right).$$

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

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

The variance of the second term is bounded by

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

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

$$\frac{p_n q_n}{n} \frac{m_n r_n}{p_n} \sim \frac{1}{h_n^*} \to 0$$

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

$$\frac{p_n q_n^2}{n}\frac{1}{h_n^*}\frac{m_n r_n}{p_n} \sim \frac{h_n}{(h_n^*)^3} \to 0$$

by the choice of h_n^*. Hence it follows that the variance of the second term in (4.9) converges to zero. Further, the variance of the third term in (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

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

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, $Y_{n,j}$, $j = 1, \dots, q_n$ 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 (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 (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

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

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

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

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

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

The first term above converges to zero as $n \to \infty$ by Proposition 3.1, while the second term converges to zero as $n \to \infty$ by Assumption C_1. Therefore, the convergence in (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 multilevel estimator (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 further to 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 $1/r_n$, $s=1,\dots,m$. 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:

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

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

Use the m levels $u_n^1,\dots,u_n^m$ and l values of block sizes $r_n^1,\dots,r_n^l$ 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 $s=1,\dots,m$, $i=1,\dots,l$, 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 are regression plane to the response variables $\widehat{\theta}_n(u_n^s,r_n^i)$ using the predictor variables $\big(\widehat{\tau}_n(u_n^s,r_n^i)/k_n^i\,, 1/r_n^i\big)$, $s=1,\dots,m$, $i=1,\dots,l$, 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

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

and

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

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 (5.2) as desired coefficients in (5.1). Alternatively, one could estimate that coefficient using different levels from the collection $u_n^1,\dots,u_n^m$.

Remark 5.1. Note that $\widehat{\beta}_0$ in (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 multilevel estimator (2.3) and its bias-corrected version (5.1) on simulated data and on S&P 500 Daily Log Returns.

5.1. Simulation Study.

We have drawn samples from ARMAX processes and MA(q) processes. The ARMAX process (X_i) is defined as follows. Let $Z_1,Z_2,\dots$ be a sequence of i.i.d. unit Fréchet random variables. For $0<\theta\le 1$, let $X_1 = Z_1/\theta$, and

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

It can be shown that the extremal index of such a sequence is θ.

The MA(q) process (X_i) is defined as

$$X_i = p_q Z_{i-q} + p_{q-1} Z_{i-q+1} + \cdots + p_1 Z_{i-1} + Z_i, \ i \geq 1\,, \tag{5.4}$$

with $0 < p_1, \ldots, p_q < 1$, and the noise sequence consisting of i.i.d. Pareto random variables $Z_{-(q-1)}, \ldots, Z_0, Z_1, \ldots$ with

$$P(Z_0 > x) = \begin{cases} 1, & \text{if } x < 1 \\ x^{-\alpha}, & \text{if } x \geq 1, \end{cases} \tag{5.5}$$

for some $\alpha > 0$. It is elementary that for this sequence the extremal index is

$$\theta = \frac{1}{1 + p_1^\alpha + \cdots + p_q^\alpha}\,,$$

see e.g. Scotto et al. (2003).

We first test the effect of the estimators (2.3) and (5.1) on the ARMAX model using values of $\theta = 0.25, 0.5, 0.75$, and a finite sample length of $n = 10000$. 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 $m = 1, \ldots, 20$, 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 higher 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 $\widehat{\theta}(f)$ and $\widehat{\theta}^b(f)$ on the basis of their bias, standard error, and mean squared error. The results computed from 5000 simulated sequences are displayed in Figure 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 bias increases while the standard error decreases as more levels of observations are incorporated into the estimator. Overall we see an increase in the mean squared error as m increases. As for the bias-corrected version of the estimator $\widehat{\theta}^b(f)$, the bias of $\widehat{\theta}^b(f)$ seems to be largely insensitive to the choice of m (compared to that of $\widehat{\theta}(f)$), while the standard error is decreasing with increasing levels m, resulting in a decreasing trend in the mean squared error. Comparatively, the bias correction we employed seems to work very well. The estimator $\widehat{\theta}^b(f)$ exhibits a much smaller bias than $\widehat{\theta}(f)$, while roughly doubling the standard error. Overall, $\widehat{\theta}^b(f)$ still achieves a much better mean squared error compared to $\widehat{\theta}(f)$, at all levels m considered.

Recall that the traditional blocks estimator (1.2) is equivalent to the multilevel estimator $\widehat{\theta}(f)$ when the number of levels $m = 1$. Namely, simulation results of (1.2) are displayed as the leftmost points of dotted lines in Figure 1. In this context, we see that our approach of incorporating more observations at lower levels does indeed improve the estimator in the sense of an improved mean squared error.

A similar exercise is done for the MA(q) model, where the coefficients in (5.4) are designed to achieve values of $\theta = 0.25, 0.5, 0.75$, respectively. Using the same settings as the ARMAX model for $\widehat{\theta}(f)$ and $\widehat{\theta}^b(f)$, we plot the results of the simulation in Figure 2.

We see that despite the different models, all the previous observations we made regarding the bias, standard error, and mean squared error of the estimators on the ARMAX model still hold on the MA(q) model. The major difference between the rows in each figure is the magnitude of the errors, which depend on the model itself as well as the value of θ. This leads us to believe that the performance of $\widehat{\theta}^b(f)$ is quite robust against different models and different values of θ.

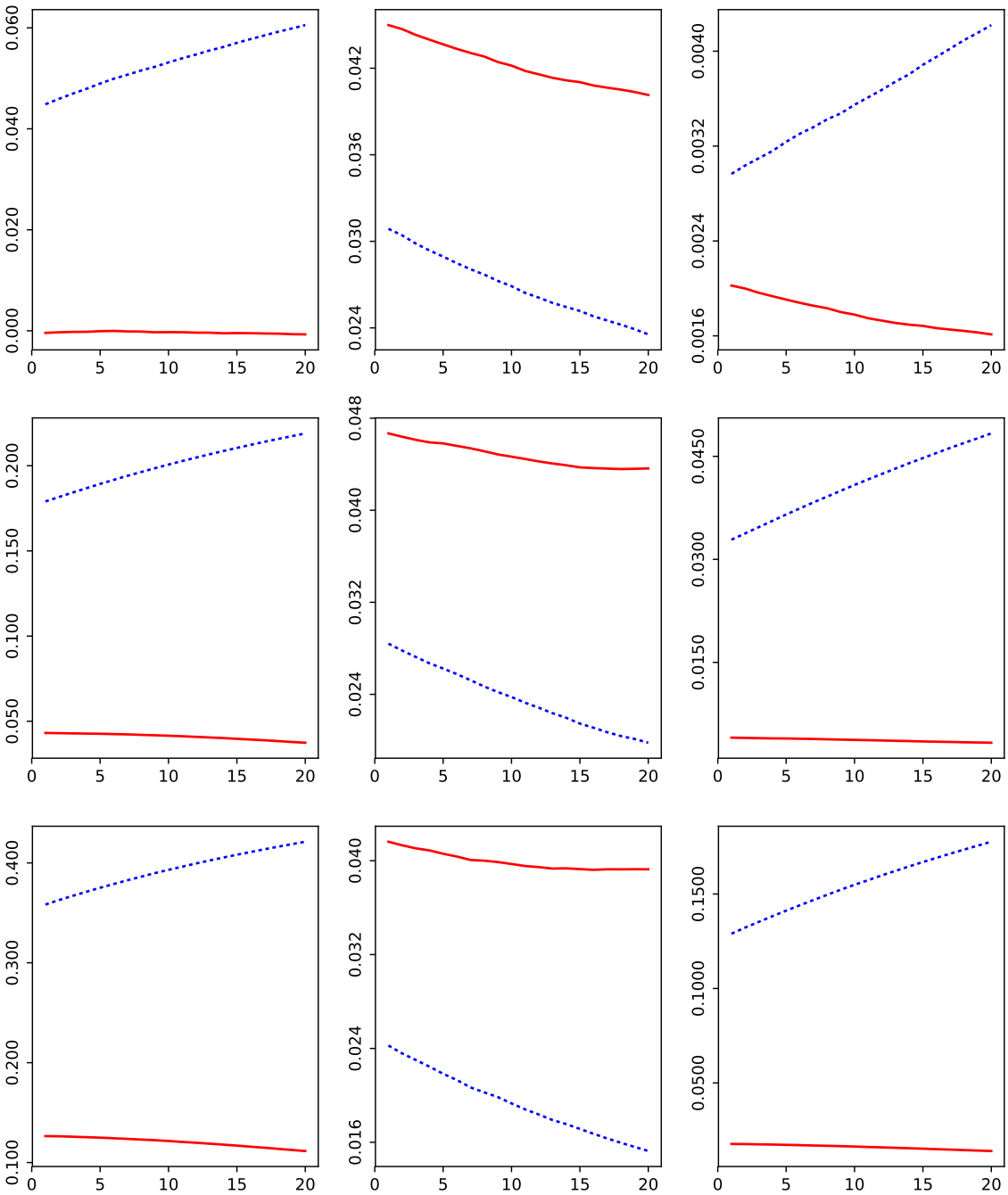

FIGURE 1. Bias (left column), standard error (center column), and mean squared error (right column) for the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used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

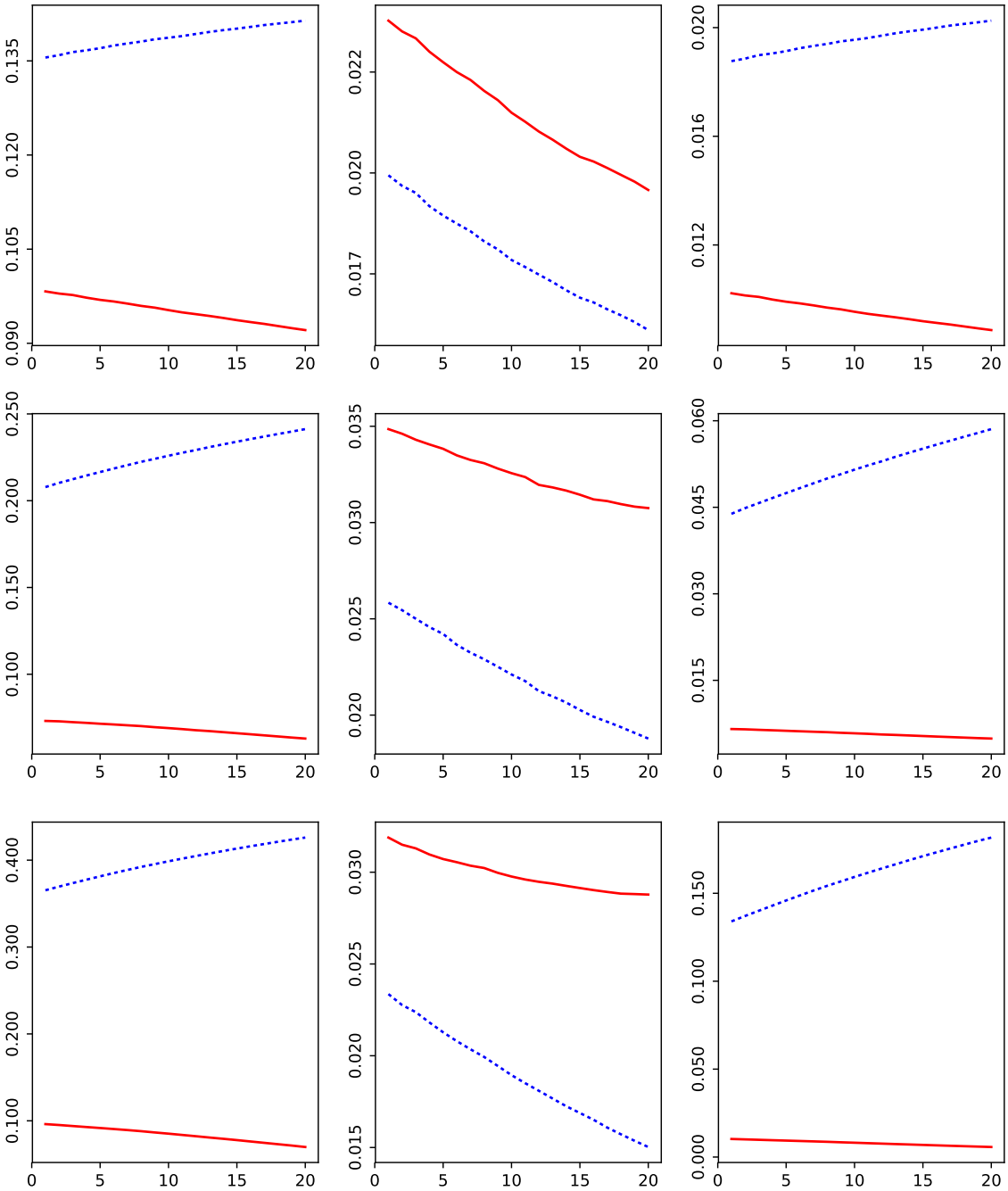

FIGURE 2. Bias (left column), standard error (center column), and mean squared error (right column) for the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from MA(q) models with $\theta = 0.25, 0.5, 0.75$ (top to bottom).

When we vary the value of the highest threshold u_m, or when we vary the choice of the block size r_n, the plots of the resulting errors still look similar to those of Figures 1 and 2, and we do not display those here. Since from the above observations, it seems that the "best" estimator is the bias-corrected estimator $\widehat{\theta}^b(f)$ that uses the largest amount of data ($m = 20$ levels) for inference, we will focus further studies on the outputs of this estimator alone.

We next look at the effects of the varying the quantile value of the highest threshold $u_m = u_{20}$ on the estimator $\widehat{\theta}^b(f)$, with different values of θ. Again our study uses the ARMAX model with values of $\theta = 0.25, 0.5, 0.75$, sample lengths of $n = 10000$, block sizes of $r_n = 200$, and a weight function of $f(x) = e^{-x}$. We now use a value of $m = 20$ levels (and $m' = 12$ levels for regression), but we choose 3 different sets of levels $u_n^1, \ldots, u_n^m$ (as well as 3 different sets of levels $\bar{u}_n^1, \ldots, \bar{u}_n^m$ for regression) to evaluate $\widehat{\theta}^b(f)$ for each ARMAX sequence. For the ith set, we choose u_n^s to be the $(50i+2(m-s)+1)$-th largest order statistic of the sequence, and $\bar{u}_n^{s'}$ being the $(50i+5(m'-s')-9)$-th largest order statistic in the sequence, $1 \le i \le 3$, $1 \le s \le m$, $1 \le s' \le m'$. The number of blocks l and block sizes r_n^i for the regression are again chosen as $l = 25$, with $r_n^i = 10i$, $1 \le i \le l$.

The results of this exercise are displayed in Figure 3. Again the results are computed from 5000 simulated sequences. Since this paper uses many similar assumptions and methods to those in Robert et al. (2009), we also included the results for the sliding blocks estimator from that study.

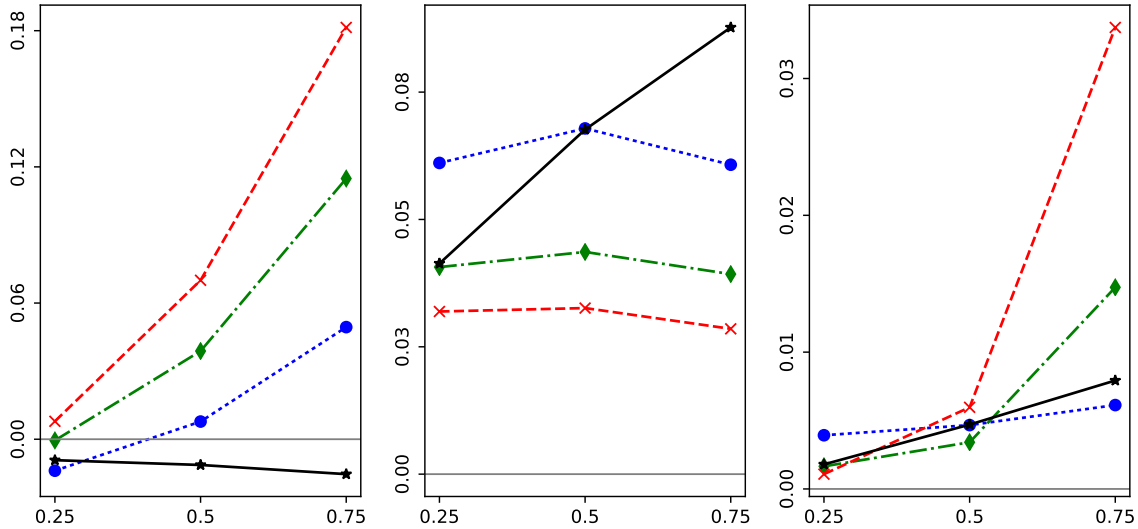

FIGURE 3. Bias (left), standard error (center), and mean squared error (left) for the bias corrected multilevel estimator $\widehat{\theta}^b(f)$ when the highest level of threshold include the largest 50 (dotted line with circle markers), the largest 100 (dot-dash line with diamond markers), and the largest 150 (dash line with 'x' markers) order statistics plotted against the value of θ. Results are also shown for the sliding blocks estimator (solid line with star markers). Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$.

From the plot it seems that among the four estimators, there is no clear winner in terms of the performance. Among the estimators $\widehat{\theta}^b(f)$ using different sets of levels $u_n^1, \ldots, u_n^m$, it seems that for a smaller value of θ, using a lower value for the highest threshold u_n^m results in a lower mean squared error. However, as the value of θ increases, we see an alarming spike in the overall error. Whereas

when using a higher value of u_n^m - meaning that a fewer total of observations are incorporated into the estimator $\widehat{\theta}^b(f)$ - the bias corrected multilevel estimator performs very well for a large value of θ, but its performance deteriorates (compared to using smaller values of u_n^m) when the value of θ decreases. Looking at the bias and standard error plots on the left (and following our intuition), using a higher threshold level u_n^m - hence using less data overall - results in a lower bias but higher standard error, and vice versa. The bias seems to go up more quickly than the variance, resulting in the shape of the mean squared error plot.

Next let us study the effect of the choice of block size r_n on the performance of the estimator $\widehat{\theta}^b(f)$. Even though the choice of r_n was not a main contention of our study, it is still an important feature in the blocks estimator. For this we again will only use the ARMAX model with $\theta = 0.25, 0.5, 0.75$. The results for MA(q) models are similar. We will use the same setting as before, with $n = 10000$, $f(x) = e^{-x}$, $m = 20$, and choosing u_n^s to be equal to the $(101 + 2(m - s))$-th largest order statistic of the sequence, $1 \leq s \leq m$. When computing coefficients for the bias-reduced estimator (5.1), we again 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 test the performance of $\widehat{\theta}^b(f)$ using block sizes of $r_n = 40, 50, \ldots, 200$. The bias and standard errors from 5000 simulated sequences are displayed in Figure 4.

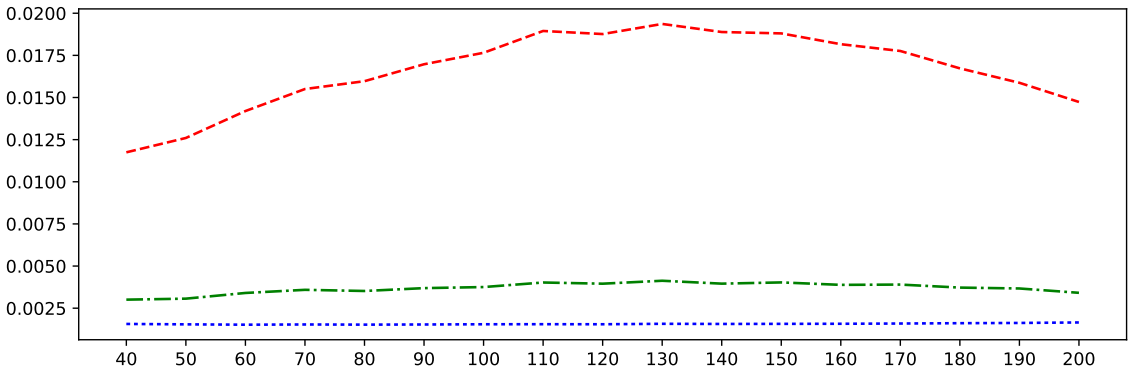

FIGURE 4. Mean squared error for the bias corrected multilevel estimator $\widehat{\theta}^b(f)$ for true values of θ being 0.25 (dotted line), 0.5 (dot-dash line), and 0.75 (dash line) plotted against the choice of r_n, the size of the blocks used in the estimator. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$.

As we can see from Figure 4, there is no discernable trend or pattern for the performance of $\widehat{\theta}^b(f)$ when we vary the block size for 40 to 200. Overall a choice of smaller values of r_n seem to lead to a better mean squared error. To take a closer look let us plot the bias and standard error for the ARMAX model with $\theta = 0.5$ against the choice of r_n, as shown in Figure 5.

From the plot we see that the standard error remains relatively constant, whereas the bias goes up and down. It seems that a block size of 40 would produce the best mean squared error - and possibly a smaller block size would be even better. However, the difference in the mean squared error or bias between a block size of $r_n = 40$ and a block size of $r_n = 200$ is very small compared to the actual value of $\theta = 0.5$. Since it is not the goal of this study to achieve an "optimal" block size, we do not study this behavior further.

Next let us study the effect of the weight function f. To do this, we use the same setting as the previous exercise, but we fix the block size to again be $r_n = 200$, and we use a second weight

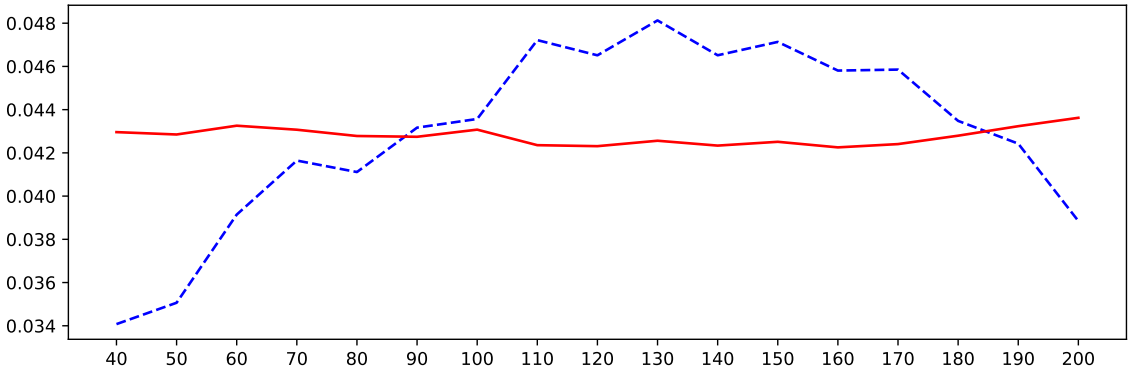

FIGURE 5. Bias (dotted line) and standard error (solid line) for the bias corrected multilevel estimator $\widehat{\theta}^b(f)$ plotted against the choice of r_n, the size of the blocks used in the estimator. Data are simulated from an ARMAX model with $\theta = 0.5$.

function, $f'(x) = 1/x^{20}$. For the values relevant, f' decreases at a much faster rate than f. We compare the performance of the estimators $\widehat{\theta}^b(f)$ and $\widehat{\theta}^b(f')$. The results are presented in Figure 6.

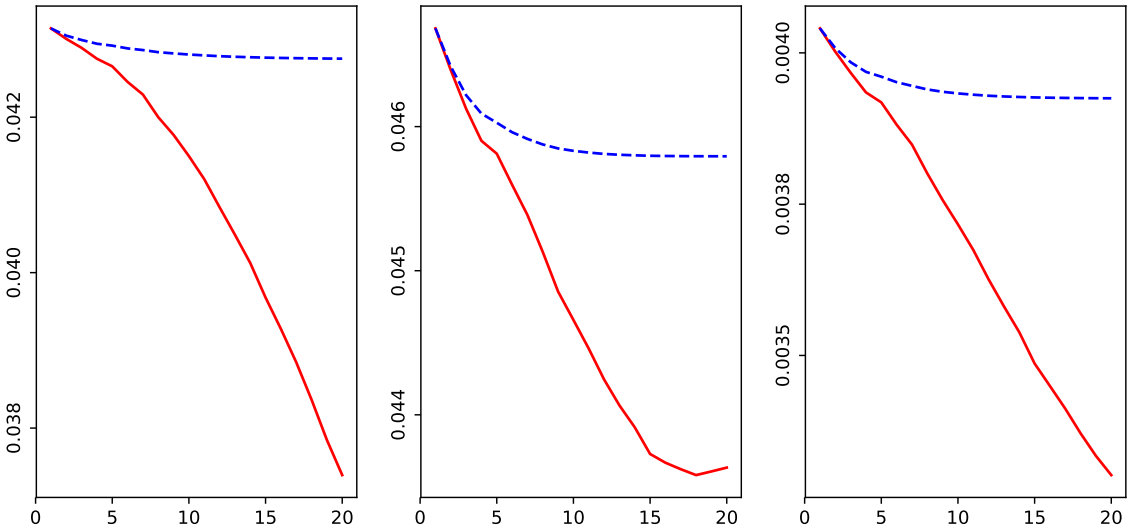

FIGURE 6. Bias (left), standard error (center) and mean squared error (right) for the bias corrected multilevel estimators $\widehat{\theta}^b(f)$ (solid line) and $\widehat{\theta}^b(f')$ (dotted line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 0.5$.

We have seen from Figure 1 that the bias, standard error, and mean squared error for $\widehat{\theta}^b(f)$ are all decreasing against the increase of m, number of levels used in the estimator. It is then to be expected that when we use a weight function f' that is much more rapidly decreasing than f - meaning that the relative weight for lower levels of observations used in $\widehat{\theta}^b(f')$ would be less than those of $\widehat{\theta}^b(f)$ - we should still expect the bias, standard error, and mean square error to be decreasing against the increase of m, but at a slower rate. This phenomena is exactly displayed

in Figure 6. However, exactly to what extend the weight function f influences the estimator is a topic not studied in this paper, and warrants further investigations.

Finally, let us go about testing the "robustness" of this estimator. First we will test the performance of the estimator $\widehat{\theta}^b(f)$ on an ARMAX model with $\theta = 1$. In light of the observations made regarding Figure 3 - that the performance of $\widehat{\theta}^b(f)$ deteriorates when θ is large and u_n^m is small, we make some changes to our simulation settings. We still use $n = 10000$, $r_n = 200$, and $f(x) = e^{-x}$. However for each fixed $m = 1, \ldots, 20$, we will set u_n^s to be the $(26+m-s)$-th largest order statistic for each sequence, $1 \leq s \leq m$. As for computing the coefficients for the bias corrected estimator, we use $m' = 12$, with $\bar{u}_n^s$ being the $(16 + 3(m' - s))$-th largest order statistic in the sequence, and $l = 25$, with $r_n^i = 10i$, $1 \leq i \leq l$. The results of the simulations are displayed in Figure 7.

We observe that as we increase the number of levels m, the bias of the estimator $\widehat{\theta}^b(f)$ briefly increases before trending downwards, whereas the standard error follows a downward trend with a slight uptick from values of $m = 18$ to $m = 20$. Overall the bias and standard errors are very small, so the bias corrected estimator $\widehat{\theta}^b(f)$ seems to still perform quite well with a value of $\theta = 1$.

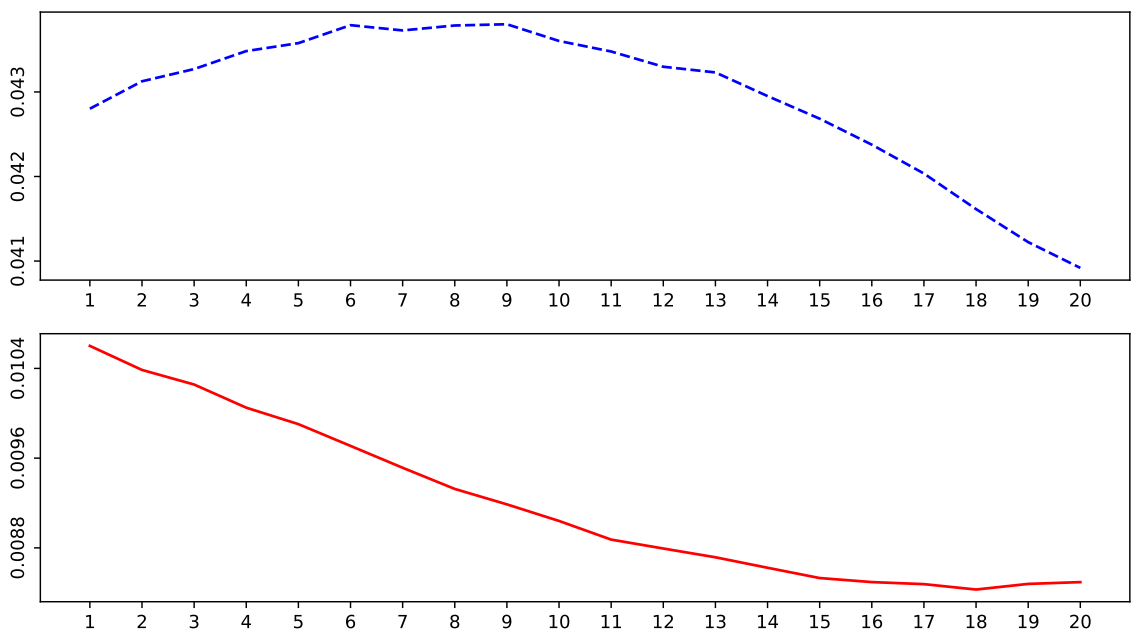

FIGURE 7. Bias (top, dotted line) and standard error (bottom, solid line) for the bias corrected multilevel estimator $\widehat{\theta}^b(f)$ plotted against the choice of m, number of levels used in the estimators. Data are simulated from an ARMAX model with $\theta = 1$.

Lastly, we will test the estimator $\widehat{\theta}^b(f)$ against a smaller value of n. We choose a value of $n = 5000$. In this case, we use ARMAX models with $\theta = 0.25, 0.5, 0.75$, and set $r_n = 100$, and $f(x) = e^{-x}$. We again run the experiments for $m = 1, \ldots, 20$, and for each fixed m we choose u_n^s to be equal to the $(101+m-s)$-th largest order statistic of the 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 $(91 + 3(m' - s))$-th largest order statistic of the sequence, $1 \leq s \leq m'$, and $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$.

Again the results from 5000 simulated sequences are displayed in Figure 8.

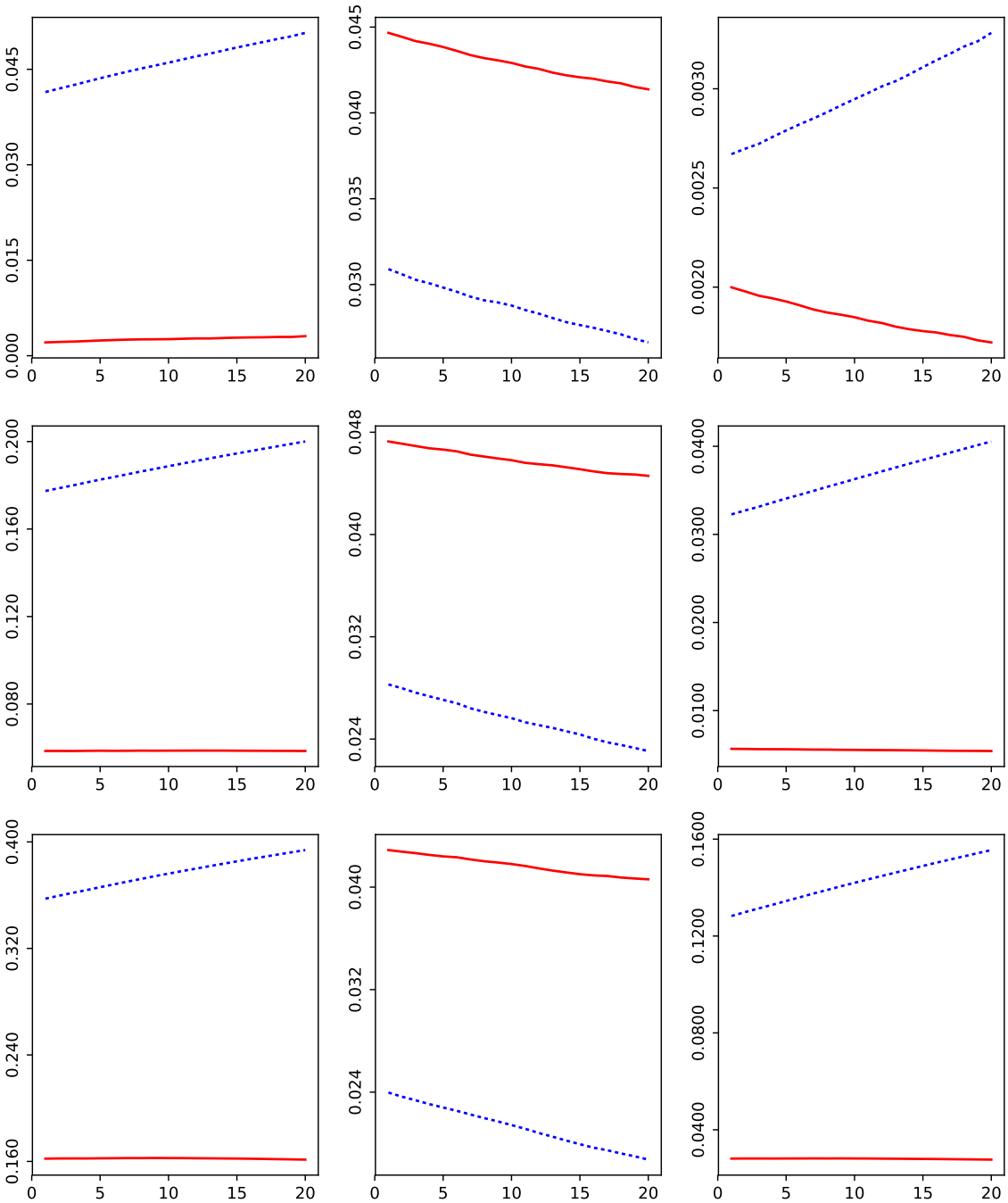

FIGURE 8. Bias (left column), standard error (center column), and mean squared error (right column) for the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom).

Again we are seeing a similar pattern to that is displayed in Figure 1, that the estimator $\widehat{\theta}^b(f)$ outperforms the traditional blocks estimator (1.2) in terms of mean squared error, and that the mean squared error is decreasing with increasing number of levels m. However, one caveat in this case is that for the value of $\theta = 0.75$, the bias is in fact quite large, as can be seen in the bottom left plot of Figure 8. Even for number of levels $m = 20$, the bias is still at around 0.16.

Recalling the conclusion from the analysis of Figure 3, that a choice of larger values of the threshold u_n^m tends work better for larger values of θ, let us repeat the above experiment, but this time for fixed n choose u_n^s to be equal to the $(21+m-s)$-th largest order statistic of the sequence, $1 \le s \le m$. When computing coefficients for the bias-reduced estimator, we use $m' = 12$, with $\bar{u}_n^s$ being the $(11 + 3(m' - s))$-th largest order statistic of the sequence, $1 \le s \le m'$.

The results are displayed in Figure 9. From the plots here we see some drastic differences against previous plots of similar type. While at the $\theta = 0.75$ level, the estimator $\widehat{\theta}^b(f)$ is still performing well, it as a different story at the $\theta = 0.25$ and $\theta = 0.5$ level. Comparatively, standard errors of $\widehat{\theta}^b(f)$ remains higher than those of $\widehat{\theta}(f)$ at each level of θ, just as before. However, at $\theta = 0.25$, the estimator $\widehat{\theta}^b(f)$ seems to "over-correct", leading to a larger absolute bias. On the other hand, at $\theta = 0.5$, the bias of $\widehat{\theta}^b(f)$ is on par with that of $\widehat{\theta}(f)$ for smaller values of m, and is better for larger values of m. However, that does not make up for that larger standard error. In both cases for $\theta = 0.25$ and $\theta = 0.5$, the bias corrected estimator $\widehat{\theta}^b(f)$ does not outperform the traditional blocks estimator (1.2).

On the other hand, at the $\theta = 0.75$ level, we observe a much smaller value of bias (and mean squared error) for the estimator $\widehat{\theta}^b(f)$, compared to when we set $u_n^m = 100$. Hence what we infer from this case as well as the analysis for Figure 3, is that for each θ, as we increase the value of the threshold at the highest level, u_n^m, the performance of the bias-corrected estimator would eventually be worse than that of blocks estimator (1.2). However, for sequences with larger values of θ, the bias corrected estimator seems to have more tolerance for the increasing value of the level u_n^m, and in fact it seems to perform better in terms of mean squared error with larger values of u_n^m. This is an interesting behavior whose cause could be a topic for further study.

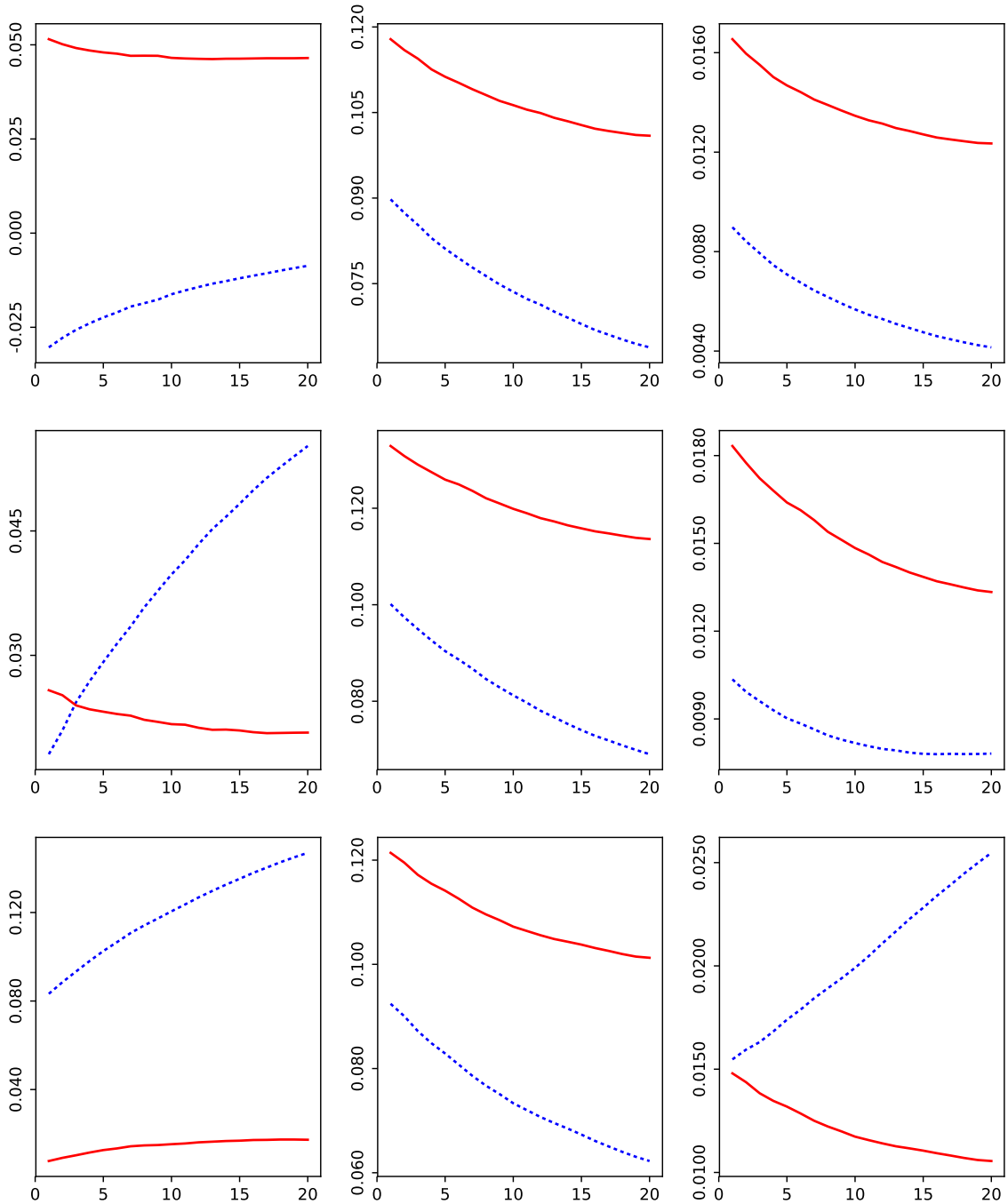

FIGURE 9. Bias (left column), standard error (center column), and mean squared error (right column) for the multilevel estimator $\widehat{\theta}(f)$ (dotted line) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (solid line) plotted against the choice of m, number of levels used in the estimators. Data are simulated from ARMAX models with $\theta = 0.25, 0.5, 0.75$ (top to bottom).

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 Figure 10.

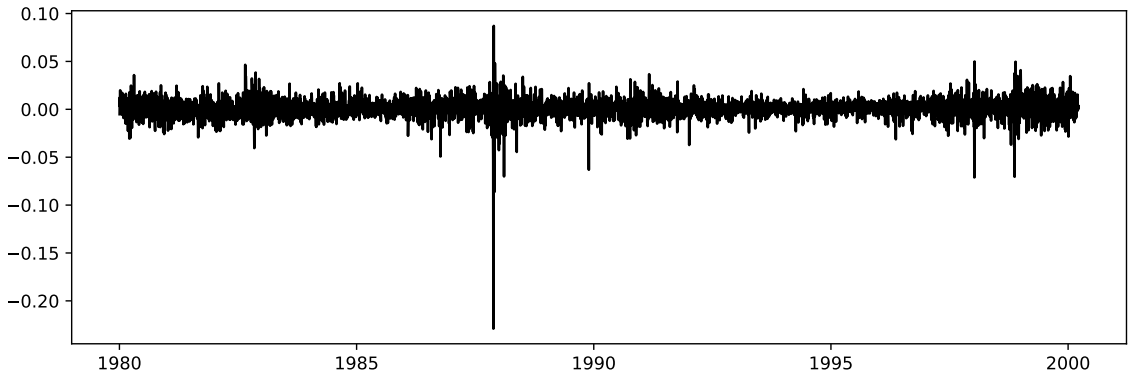

FIGURE 10. Daily Log Returns for S&P 500 from 1980 - 1999

There are $n = 5055$ returns in this data set, and the negative of their values form our sample. An initial estimate returned a value θ between 0.5 and 0.75, so we choose a relatively threshold of u_n^m. 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 (5.2) with $m' = 12$ levels, $\bar{u}_n^s$ being the $11 + 3(m' - s)$-th largest order statistic in the sample, $1 \leq s \leq m'$, and set $l = 15$, with $r_n^i = 10i$, $1 \leq i \leq l$.

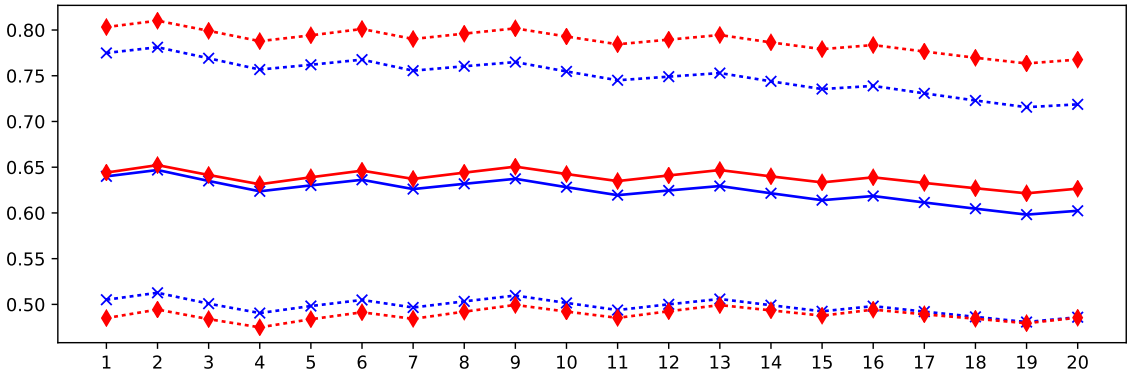

FIGURE 11. Values (solid lines) and point-wise 1 standard error intervals (dotted lines) for the multilevel estimator $\widehat{\theta}(f)$ ('x' marker) and the bias-corrected multilevel estimator $\widehat{\theta}^b(f)$ (diamond marker) plotted against the choice of m, number of levels used in the estimators. Standard errors are generated by bootstrap sampling of blocks of size 40. Data is drawn from the negative daily log returns for S&P 500.

The plots of the two estimators are shown above as a function of the number of levels m. The difference in the values of the two estimators, as well as their standard errors, are consistent with

those shown in Figure 8. Since in that case we have seen that the bias corrected estimator with the largest value of m is the "best" estimator, its value of 0.627 at $m = 20$ may be taken as the estimate of the extremal index for the losses.

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006): *Statistics of extremes: theory and applications*, John Wiley & Sons.

Berghaus, B. and A. Bücher (2017): *Weak convergence of a pseudo maximum likelihood estimator for the extremal index*, Preprint, arXiv: 1608.01903.

de Haan, L. and A. Ferreira (2006): *Extreme Value Theory: An Introduction*, New York: Springer.

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

Drees, H. and E. Kaufmann (1998): "Selecting the optimal sample fraction in univariate extreme value estimation," *Stochastic Processes and their Applications*, 75, 149–172.

Dupuis, D. (1998): "Exceedances over high thresholds: a guide to threshold selection," *Extremes*, 1, 251–261.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*, Berlin: Springer-Verlag.

Ferro, C. and J. Segers (2003): "Inference for clusters of extreme values," *Journal of the Royal Statistical Society. Series B (Methodological)*, 65, 545–556.

Hsing, T. (1991): "Estimating the parameters of rare events," *Stochastic Processes and their Applications*, 37, 117–139.

——— (1993): "Extremal index estimation for a weakly dependent stationary sequence," *The Annals of Statistics*, 21, 2043–2071.

Hsing, T., J. Hüsler, and M. Leadbetter (1988): "On the exceedance point process for a stationary sequence," *Probability Theory and Related Fields*, 78, 97–112.

Laurini, F. and J. A. Tawn (2003): "New estimators for the extremal index and other cluster characteristics," *Extremes*, 6(3), 189–211.

Leadbetter, M., G. Lindgren, and H. Rootzén (1983): *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer Verlag.

Nguyen, T. and G. Samorodnitsky (2012): "Tail Inference: where does the tail begin?" *Extremes*, 15, 437–461.

Northrop, P. J. (2015): "An efficient semiparametric maxima estimator of the extremal index" *Extremes*, 18(4), 585–603.

Resnick, S. (2007): *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, New York: Springer.

Resnick, S. and C. Stărică (1997): "Smoothing the Hill estimator," *Advances in Applied Probability*, 20, 271–293.

Robert, C. (2009): "Inference for the limiting cluster size distribution of extreme values," *The Annals of Statistics*, 37, 271–310.

Robert, C., J. Segers, and C. Ferro (2009): "A sliding block estimator for the extremal index," *Electronic Journal of Statistics*, 3, 993–1020.

Scotto, M., K. Turkman, and C. Anderson (2003): "Extremes of Some Sub-Sampled Time Series," *Journal of Time Series Analysis*, 24, 579–590.

Smith, R. and I. Weissman (1994): "Estimating the extremal index," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 515–528.

Süveges, M (2007): "Likelihood estimation of the extremal index," *Extremes*, 10(1-2), 41–55.

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

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853

E-mail address: ys598@cornell.edu

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

E-mail address: gs18@cornell.edu