WEAK CONVERGENCE OF A PSEUDO MAXIMUM LIKELIHOOD ESTIMATOR FOR THE EXTREMAL INDEX[1]

BY BETINA BERGHAUS AND AXEL BÜCHER

Ruhr-Universität Bochum

The extremes of a stationary time series typically occur in clusters. A primary measure for this phenomenon is the extremal index, representing the reciprocal of the expected cluster size. Both disjoint and sliding blocks estimator for the extremal index are analyzed in detail. In contrast to many competitors, the estimators only depend on the choice of one parameter sequence. We derive an asymptotic expansion, prove asymptotic normality and show consistency of an estimator for the asymptotic variance. Explicit calculations in certain models and a finite-sample Monte Carlo simulation study reveal that the sliding blocks estimator outperforms other blocks estimators, and that it is competitive to runs- and inter-exceedance estimators in various models. The methods are applied to a variety of financial time series.

1. Introduction. An adequate description of the extremal behavior of a time series is important in many applications, such as in hydrology, finance or actuarial science (see, e.g., Section 1.3 in the monograph [Beirlant et al. (2004)]). The extremal behavior can be characterized by the tail of the marginal law of the time series and by the serial dependence; that is, by the tendency that extremal observations tend to occur in clusters. A primary measure of extremal serial dependence is given by the extremal index $\theta \in [0, 1]$, which can be interpreted as being equal to the reciprocal of the mean cluster size. The underlying theory was worked out in Hsing, Hüsler and Leadbetter (1988), Leadbetter (1983), Leadbetter, Lindgren and Rootzén (1983), Leadbetter and Rootzén (1988), O'Brien (1987).

Estimating the extremal index based on a finite stretch from the time series has been extensively studied in the literature. Common approaches are based on the blocks method, the runs method and the inter-exceedance times method [see Beirlant et al. (2004), Section 10.3.4, for an overview]. The first two methods usually depend on two parameters to be chosen by the statistician: a threshold sequence and a cluster identification scheme parameter (such as a block length). In contrast, inter-exceedance type-estimators are attractive since they only depend on a threshold sequence. Some references are Ferro and Segers (2003), Hsing (1993),

Robert (2009), Robert, Segers and Ferro (2009), Smith and Weissman (1994), Süveges (2007), Süveges, Davison et al. (2010), Weissman and Novak (1998), among others. The present paper is on a blocks estimator (and a slightly modified version) due to Northrop (2015), which, remarkably, only depends on a cluster identification parameter. This makes the estimator practically appealing in comparison to other blocks methods.

In many papers on estimating the extremal index, either no asymptotic theory is given [such as in Northrop (2015), Süveges (2007)], or the asymptotic theory is incomplete in the sense that theory is developed for a nonrandom threshold sequence, while in practice a random sequence must be used [as, e.g., in Robert, Segers and Ferro (2009), Weissman and Novak (1998)]. As pointed out in the latter paper, "the mathematical treatment of such random threshold sequences requires complicated empirical process theory." In the present paper, the mathematical treatment is comprehensive, working out all the arguments needed from empirical process theory.

Let us proceed by motivating and defining the estimator: throughout, $X_1, X_2, \ldots$ denotes a stationary sequence of real-valued random variables with stationary cumulative distribution function (c.d.f.) F. The sequence is assumed to have an extremal index $\theta \in (0, 1]$: for any $\tau > 0$, there exists a sequence $u_n = u_n(\tau)$ such that $\lim_{n \to \infty} n \bar{F}(u_n) = \tau$ and such that

$$\lim_{n \to \infty} \mathbb{P}(M_{1:n} \leq u_n) = e^{-\theta \tau}.$$

Here, $\bar{F} = 1 - F$ and $M_{1:n} = \max(X_1, \ldots, X_n)$.

For simplicity, we assume that F is continuous (cf. Remark 3.6 below) and define a sequence of standard uniform random variables by $U_s = F(X_s)$. For $x \in (0, 1)$, let $u_n = F^{\leftarrow}(1 - x/n)$ and $u_n' = F^{\leftarrow}(e^{-x/n})$, where $F^{\leftarrow}$ denotes the generalized, left-continuous inverse of the c.d.f. F. Then $n \bar{F}(u_n) = x$ and $n \bar{F}(u_n') = n(1 - e^{-x/n}) \to x$ as $n \to \infty$, whence

$$(1.1) \qquad \mathbb{P}\big(n(1 - N_{1:n}) \geq x\big) = \mathbb{P}(M_{1:n} \leq u_n) \to e^{-\theta x},$$

$$(1.2) \qquad \mathbb{P}\big(-n \log(N_{1:n}) \geq x\big) = \mathbb{P}(M_{1:n} \leq u_n') \to e^{-\theta x},$$

where $N_{1:n} = F(M_{1:n}) = \max\{U_1, \ldots, U_n\}$. In other words, both $Y_{1:n} = -n \log(N_{1:n})$ and $Z_{1:n} = n(1 - N_{1:n})$ asymptotically follow an exponential distribution with parameter θ. The result concerning $Y_{1:n}$ inspired Northrop (2015) to estimate θ by the maximum likelihood estimator for the exponential distribution, based on a sample of estimated block maxima.

More precisely, suppose that we observe a stretch of length n from the time series $(X_s)_{s \geq 1}$. Divide the sample into k_n blocks of length b_n, and for simplicity assume that $n = b_n k_n$ (otherwise, the final block would consist of less than b_n observations and should be omitted). For $i = 1, \ldots, k_n$, let

$$M_{ni} = M_{((i-1)b_n+1):(ib_n)} = \max\{X_{(i-1)b_n+1}, \ldots, X_{ib_n}\}$$

denote the maximum over the X_s from the ith block. Also, let $N_{ni} = F(M_{ni}) = \max\{U_{(i-1)b_n+1}, \ldots, U_{ib_n}\}$ and $Y_{ni} = -b_n \log(N_{ni})$. If b_n is sufficiently large, then by (1.2), the (unobservable) random variables $Y_{n1}, \ldots, Y_{nk}$ form an approximate sample from the Exponential(θ)-distribution. Moreover, as common when working with block maxima of a time series, they may be considered as asymptotically independent, which prompted Northrop (2015) to estimate θ by the maximum-likelihood estimator for the Exponential(θ) distribution:

$$\tilde{\theta}_n^{\mathrm{N}} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} Y_{ni} \right)^{-1}.$$

Note that $\tilde{\theta}_n^{\mathrm{N}}$ should not be considered an estimator, as it is based on the unknown c.d.f. F. Subsequently, we call $\tilde{\theta}_n^{\mathrm{N}}$ an oracle for θ.

In practice, the U_s are not observable, whence they need to be replaced by their observable counterparts giving rise to the definitions

$$\hat{N}_{ni} = \hat{F}_n(M_{ni}) \quad \text{and} \quad \hat{Y}_{ni} = -b_n \log(\hat{N}_{ni}),$$

where $\hat{F}_n(x) = n^{-1} \sum_{s=1}^{n} \mathbf{1}(X_s \leq x)$ denotes the empirical c.d.f. of $X_1, \ldots, X_n$. We obtain, up to a bias correction discussed below, Northrop's estimator

$$(1.3) \qquad \hat{\theta}_n^{\mathrm{N}} = \hat{\theta}_n^{\mathrm{N,dj}} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \hat{Y}_{ni} \right)^{-1}.$$

In Northrop (2015), no asymptotic theory on $\hat{\theta}_n^{\mathrm{N}}$ is given. While deriving the asymptotic distribution of the oracle $\tilde{\theta}_n^{\mathrm{N}}$ may appear tractable [see also Robert (2009): essentially, a central limit theorem for rowwise dependent triangular arrays is to be shown, followed by an argument using the delta method], asymptotic theory on the estimator $\hat{\theta}_n^{\mathrm{N}}$ is substantially more difficult due to the additional serial dependence induced by the rank transformation (which on top of that operates between blocks instead of within blocks).

A central contribution of the present paper is the derivation of the asymptotic distribution of $\hat{\theta}_n^{\mathrm{N}}$. It will further turn out that the impact of the rank transformation is nonnegligible, resulting in different asymptotic variances of $\hat{\theta}_n^{\mathrm{N}}$ and the corresponding oracle $\tilde{\theta}_n^{\mathrm{N}}$. For that purpose, it will be convenient to consider the following (mathematically simpler) variant of Northrop's estimator:

$$(1.4) \qquad \hat{\theta}_n^{\mathrm{B}} = \hat{\theta}_n^{\mathrm{B,dj}} = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \hat{Z}_{ni} \right)^{-1}, \qquad \hat{Z}_{ni} = b_n(1 - \hat{N}_{ni}).$$

This estimator can either be motivated following the above lines, but using (1.1) rather than (1.2) as a starting point, or by consulting Robert (2009) and writing

$$(1.5) \qquad \frac{1}{\hat{\theta}_n^{\mathrm{B,dj}}} = \int_0^\infty \hat{p}_n^{(\tau)}(0) \, d\tau,$$

with $\hat{p}_n^{(\tau)}(0)$ denoting Robert's estimator for $e^{-\theta\tau}$ [page 275 in Robert (2009), with ">" replaced by "≥" in his definition of $\hat{N}_{r_n,j}^{(\tau)}$]. We will show below (Theorem 3.1) that $\hat{\theta}_n^{\mathrm{B}}$ and $\hat{\theta}_n^{\mathrm{N}}$ are in fact asymptotically equivalent. We also present asymptotic theory for modifications of $\hat{\theta}_n^{\mathrm{N}}$ and $\hat{\theta}_n^{\mathrm{B}}$ based on sliding block maxima, which is the second main contribution of the paper. Finally, the asymptotic expansions for $\hat{\theta}_n^{\mathrm{B}}$ suggest estimators for the asymptotic variance of $\hat{\theta}_n^{\mathrm{N}}$ and $\hat{\theta}_n^{\mathrm{B}}$ (and its sliding blocks variants); proving their consistency is the third main contribution.

The remaining parts of this paper are organized as follows: in Section 2, we present mathematical preliminaries needed to formulate and derive the asymptotic distributions of the estimators for θ. Asymptotic equivalence, consistency and asymptotic normality is then shown in Section 3. Estimators of the asymptotic variance are handled in Section 4. In Section 5, we propose a simple device to reduce the bias of the estimator and relate it to the ad hoc approach in Northrop (2015). Examples are worked out in detail in Section 6, while finite-sample results and a case study are presented in Sections 7 and 8, respectively. Sections 9 and 10 contain a sequence of auxiliary lemmas needed for the proof of the main results. Their proofs, as well as additional proofs are postponed to the Supplementary Material [Berghaus and Bücher (2018), Appendices A, B, C and D]. The Supplementary Material also contains additional simulation results (Appendix E).

2. Mathematical preliminaries. The serial dependence of the time series $(X_s)_s$ will be controlled via mixing coefficients. For two sigma-fields $\mathcal{F}_1, \mathcal{F}_2$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let

$$\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup_{A \in \mathcal{F}_1, B \in \mathcal{F}_2} \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right|.$$

In time series extremes, one usually imposes assumptions on the decay of the mixing coefficients between sigma-fields generated by $\{X_i \mathbf{1}(X_s > F^{\leftarrow}(1 - \varepsilon_n)) : s \leq \ell\}$ and $\{X_s \mathbf{1}(X_s > F^{\leftarrow}(1 - \varepsilon_n)) : s \geq \ell + k\}$, where $\varepsilon_n \to 0$ is some sequence reflecting the fact that only the dependence in the tail needs to be restricted [see, e.g., Rootzén (2009)]. For our purposes, we need slightly more to control even the dependence between the smallest of all block maxima [see also Condition 2.1(v) below]. More precisely, for $-\infty \leq p < q \leq \infty$ and $\varepsilon \in (0, 1]$, let $\mathcal{B}_{p:q}^\varepsilon$ denote the sigma algebra generated by $U_s^\varepsilon := U_s \mathbf{1}(U_s > 1 - \varepsilon)$ with $s \in \{p, \ldots, q\}$ and define, for $\ell \geq 1$,

$$\alpha_\varepsilon(\ell) = \sup_{k \in \mathbb{N}} \alpha\left(\mathcal{B}_{1:k}^\varepsilon, \mathcal{B}_{k+n:\infty}^\varepsilon \right).$$

Note that the coefficients are increasing in ε, whence they are bounded by the standard alpha-mixing coefficients of the sequence U_s, which can be retrieved for $\varepsilon = 1$. In Condition 2.1(iii) below, we will impose a condition on the decay of the mixing coefficients for small values of ε.

The extremes of a time series may be conveniently described by the point process of normalized exceedances. The latter is defined, for a Borel set $A \subset E := (0, 1]$ and a number $x \in [0, \infty)$, by

$$N_n^{(x)}(A) = \sum_{s=1}^{n} \mathbf{1}(s/n \in A, U_s > 1 - x/n).$$

Note that $N_n^{(x)}(E) = 0$ iff $N_{1:n} \leq 1 - x/n$; the probability of that event converging to $e^{-\theta x}$ under the assumption of the existence of extremal index θ.

Fix $m \geq 1$ and $x_1 > \cdots > x_m > 0$. For $1 \leq p < q \leq n$, let $\mathcal{F}_{p:q,n}^{(x_1,\ldots,x_m)}$ denote the sigma-algebra generated by the events $\{U_i > 1 - x_j/n\}$ for $p \leq i \leq q$ and $1 \leq j \leq m$. For $1 \leq \ell \leq n$, define

$$\alpha_{n,\ell}(x_1, \ldots, x_m) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| :$$
$$A \in \mathcal{F}_{1:s,n}^{(x_1,\ldots,x_m)}, B \in \mathcal{F}_{s+\ell:n,n}^{(x_1,\ldots,x_m)}, 1 \leq s \leq n - \ell\}.$$

The condition $\Delta_n(\{u_n(x_j)\}_{1 \leq j \leq m})$ is said to hold if there exists a sequence $(\ell_n)_n$ with $\ell_n = o(n)$ such that $\alpha_{n,\ell_n}(x_1, \ldots, x_m) = o(1)$ as $n \to \infty$. A sequence $(q_n)_n$ with $q_n = o(n)$ is said to be $\Delta_n(\{u_n(x_j)\}_{1 \leq j \leq m})$-separating if there exists a sequence $(\ell_n)_n$ with $\ell_n = o(q_n)$ such that $nq_n^{-1}\alpha_{n,\ell_n}(x_1, \ldots, x_m) = o(1)$ as $n \to \infty$. If $\Delta_n(\{u_n(x_j)\}_{1 \leq j \leq m})$ is met, then such a sequence always exists, simply take $q_n = \lfloor \max\{n\alpha_{n,\ell_n}^{1/2}, (n\ell_n)^{1/2}\} \rfloor$.

By Theorems 4.1 and 4.2 in Hsing, Hüsler and Leadbetter (1988), if the extremal index exists and the $\Delta(u_n(x))$-condition is met ($m = 1$), then a necessary and sufficient condition for weak convergence of $N_n^{(x)}$ is convergence of the conditional distribution of $N_n^{(x)}(B_n)$ with $B_n = (0, q_n/n]$ given that there is at least one exceedance of $1 - x/n$ in $\{1, \ldots, q_n\}$ to a probability distribution π on $\mathbb{N}$, that is,

$$\lim_{n \to \infty} \mathbb{P}(N_n^{(x)}(B_n) = j | N_n^{(x)}(B_n) > 0) = \pi(j) \qquad \forall j \geq 1,$$

where q_n is some $\Delta(u_n(x))$-separating sequence. Moreover, in that case, the convergence in the last display holds for any $\Delta(u_n(x))$-separating sequence q_n. If the $\Delta(u_n(x))$-condition holds for any $x > 0$, then π does not depend on x [Hsing, Hüsler and Leadbetter (1988), Theorem 5.1].

A multivariate version of the latter results is stated in Perfekt (1994); see also the summary in Robert (2009), page 278, and the thesis Hsing (1984). Suppose that the extremal index exists and that the $\Delta(u_n(x_1), u_n(x_2))$-condition is met for any $x_1 \geq x_2 \geq 0, x_1 \neq 0$. Moreover, assume that there exists a family of probability measures $\{\pi_2^{(\sigma)} : \sigma \in [0, 1]\}$ on $\mathcal{J} = \{(i, j) : i \geq j \geq 0, i \geq 1\}$ such that, for all $(i, j) \in \mathcal{J}$,

$$\lim_{n \to \infty} \mathbb{P}(N_n^{(x_1)}(B_n) = i, N_n^{(x_2)}(B_n) = j | N_n^{(x_1)}(B_n) > 0) = \pi_2^{(x_2/x_1)}(i, j),$$

where q_n is some $\Delta(u_n(x_1), u_n(x_2))$-separating sequence. In that case, the two-level point process $\mathbf{N}_n^{(x_1, x_2)} = (N_n^{(x_1)}, N_n^{(x_2)})$ converges in distribution to a point process with characterizing Laplace transform explicitly stated in Robert (2009) on top of page 278. Note that

$$\pi_2^{(1)}(i, j) = \pi(i)\mathbf{1}(i = j), \qquad \pi_2^{(0)}(i, j) = \pi(i)\mathbf{1}(j = 0).$$

The following set of conditions will be imposed to establish asymptotic normality of the estimators.

CONDITION 2.1.   (i) *Extremal index and the point process of exceedances.* The extremal index $\theta \in (0, 1]$ exists and the above assumptions guaranteeing convergence of the one- and two-level point process of exceedances are satisfied.

(ii) *Moment assumption on the point process.* There exists $\delta > 0$ such that, for any $\ell > 0$, there exists a constant C_ℓ' such that

$$\mathrm{E}\big[\big|N_n^{(x_1)}(E) - N_n^{(x_2)}(E)\big|^{2+\delta}\big] \leq C_\ell'(x_1 - x_2) \qquad \forall \ell \geq x_1 \geq x_2 \geq 0, n \in \mathbb{N}.$$

(iii) *Asymptotic independence in the big-block/small-block heuristics.* There exists $c_2 \in (0, 1)$ and $C_2 > 0$ such that

$$\alpha_{c_2}(\ell) \leq C_2 \ell^{-\eta}$$

for some $\eta \geq 3(2 + \delta)/(\delta - \mu) > 3$ with $0 < \mu < \delta \wedge (1/2)$ and with $\delta > 0$ from Condition (ii). The block size $b_n \to \infty$ is chosen in such a way that

$$(2.1) \qquad\qquad k_n = o\big(b_n^2\big), \qquad n \to \infty,$$

and such that there exists a sequence $\ell_n \to \infty$ (to be thought of as the length of small blocks which are to be clipped-of at the end of each block of size b_n) satisfying $\ell_n = o(b_n^{2/(2+\delta)})$ and $k_n \alpha_{c_2}(\ell_n) = o(1)$; all convergences being for $n \to \infty$.

(iv) *Bound on the variance of the empirical process.* There exist some constants $c_1 \in (0, 1)$, $C_1 > 0$ such that, for all $y \in (0, c_1)$ and all $n \in \mathbb{N}$,

$$\mathrm{Var}\bigg\{\sum_{s=1}^{n} \mathbf{1}(U_s > 1 - y)\bigg\} \leq C_1\big(ny + n^2 y^2\big).$$

(v) *All standardized block maxima of size $b_n/2$ converge to* 1. For all $c \in (0, 1)$, we have

$$\lim_{n \to \infty} \mathbb{P}\Big(\min_{i=1}^{2k_n} N_{ni}' \leq c\Big) = 0,$$

where $N_{ni}' = \max\{U_s : s \in [(i - 1)b_n/2 + 1, \ldots, ib_n/2]\}$, for $i = 1, \ldots, 2k_n$, denote consecutive standardized block maxima of (approximate) size $b_n/2$.

(vi) *Existence of moments of maxima.* With $\delta > 0$ from Condition (ii), we have

$$\limsup_{n \to \infty} E\big[Z_{1:n}^{2+\delta}\big] < \infty.$$

(vii) *Bias.* As $n \to \infty$,

$$E[Z_{1:b_n}] = \theta^{-1} + o\big(k_n^{-1/2}\big).$$

Assumptions (i)–(iii) are suitable adaptations of Conditions (C1) and (C2) in Robert (2009); in fact, they can be seen to imply the latter. Among other things, these conditions are needed to apply his central result, Theorem 4.1, on the weak convergence of the tail empirical process on $[0, \infty)$. Note that the assumptions are satisfied for solutions of stochastic difference equations; see Example 3.1 in Robert (2009). The Assumption in (2.1) is a growth condition that is needed in the proof of Lemma 9.1. As argued in Robert, Segers and Ferro (2009), it is actually a weak requirement, as in many time series models it is a necessary condition for the bias condition in (vii) to be true (see Section 6 below). Finally, a positive extremal index can be guaranteed by assuming that

$$(2.2) \qquad \lim_{m \to \infty} \limsup_{n \to \infty} \mathbb{P}\Big(N_{m:b_n} > 1 - \frac{x}{n} \Big| U_1 \geq 1 - \frac{x}{n} \Big) = 0$$

for any $x > 0$; see Beirlant et al. (2004), formula (10.8). We will additionally need this assumption for the calculation of the asymptotic variance of the estimators.

In a slightly different form concerning only the tail, Assumption (iv) has also been made in Condition (C3) in Drees (2000) for proving weak convergence of the tail empirical process. In comparison to there, the extra factor $n^2 y^2$ allows for additional flexibility, in that it allows for $O(n^2)$-nonnegligible covariances, as long as their contribution is at most y^2. In Section 6, we show that the assumption holds for solutions of stochastic difference equations, such as the squared ARCH-model, and for max-autoregressive models.

Recall that $N_{ni}^{b_n}$ is approximately Beta$(\theta, 1)$-distributed. As a consequence, every standardized block maximum N_{ni} must converge to 1 as the sample size grows to infinity. Still, out of the sample of k_n block maxima, the smallest one could possibly be smaller than one, especially when the number of blocks is large. Assumption (v) prevents this from happening; note that a similar assumption has also been made in Bücher and Segers (2015), Condition 3.2. Imposing the assumption even for block maxima N'_{ni} of size $b_n/2$ guarantees that also the minimum over all big sub-block maxima (needed in the proof for the disjoint blocks estimator) and the minimum over all sliding block maxima of size b_n (needed in the proof for the sliding blocks estimator) converges to 1.

Assumption (vi) is needed to deduce uniform integrability of the sequence $Z_{1:b_n}^2$. It implies convergence of the variance of $Z_{1:b_n}$ to that of an exponential distribution with parameter θ. Finally, (vii) requires the approximation of the first moment of $Z_{1:b_n}$ by that of an exponential distribution to be sufficiently accurate.

3. Main results. In this section, we prove consistency and asymptotic normality of the disjoint blocks estimators $\hat{\theta}_n^{\mathrm{N,dj}}$ and $\hat{\theta}_n^{\mathrm{B,dj}}$ defined in (1.3) and (1.4), respectively, as well as of variants which are based on sliding blocks and which we will denote by $\hat{\theta}_n^{\mathrm{N,sl}}$ and $\hat{\theta}_n^{\mathrm{B,sl}}$, respectively. We begin by defining the latter estimators.

Divide the sample into $n - b_n + 1$ blocks of length b_n, that is, for $t = 1, \ldots, n - b_n + 1$, let

$$M_{nt}^{\mathrm{sl}} = M_{t:(t+b_n-1)} = \max\{X_t, \ldots, X_{t+b_n-1}\}.$$

Analogously to the notation used in the definition of the estimators for disjoint blocks, we will write $N_{nt}^{\mathrm{sl}} = F(M_{nt}^{\mathrm{sl}})$, $Z_{nt}^{\mathrm{sl}} = b_n(1 - N_{nt}^{\mathrm{sl}})$ and $Y_{nt}^{\mathrm{sl}} = -b_n \log(N_{nt}^{\mathrm{sl}})$ and define their empirical counterparts $\hat{N}_{nt}^{\mathrm{sl}} = \hat{F}_n(M_{nt}^{\mathrm{sl}})$, $\hat{Z}_{nt}^{\mathrm{sl}} = b_n(1 - \hat{N}_{nt}^{\mathrm{sl}})$ and $\hat{Y}_{nt}^{\mathrm{sl}} = -b_n \log(\hat{N}_{nt}^{\mathrm{sl}})$, where $\hat{F}_n$ is the empirical c.d.f. of $X_1, \ldots, X_n$. Just as for the disjoint blocks estimators, the (pseudo-)observations $\hat{Z}_{nt}^{\mathrm{sl}}$ and $\hat{Y}_{nt}^{\mathrm{sl}}$ are approximately exponentially distributed with mean θ^{-1}, which suggests to estimate θ by the reciprocal of their empirical mean:

$$\hat{\theta}_n^{\mathrm{B,sl}} = \left(\frac{1}{n - b_n + 1} \sum_{t=1}^{n-b_n+1} \hat{Z}_{nt}^{\mathrm{sl}} \right)^{-1}, \qquad \hat{\theta}_n^{\mathrm{N,sl}} = \left(\frac{1}{n - b_n + 1} \sum_{t=1}^{n-b_n+1} \hat{Y}_{nt}^{\mathrm{sl}} \right)^{-1}.$$

Up to a bias correction discussed below, $\hat{\theta}_n^{\mathrm{N,sl}}$ is the sliding blocks estimator proposed in Northrop (2015). Note that, for both estimators, no data has to be discarded if b_n is not a divisor of the sample size n.

The first central result is on first-order asymptotic equivalence between the proposed estimators, proven in Appendix C in the Supplementary Material [Berghaus and Bücher (2018)].

THEOREM 3.1. *Suppose that Condition* 2.1 *and* (2.2) *is met. Then, as* $n \to \infty$,

$$\sqrt{k_n}(\hat{\theta}_n^{\mathrm{B,dj}} - \hat{\theta}_n^{\mathrm{N,dj}}) = o_{\mathbb{P}}(1) \quad and \quad \sqrt{k_n}(\hat{\theta}_n^{\mathrm{B,sl}} - \hat{\theta}_n^{\mathrm{N,sl}}) = o_{\mathbb{P}}(1).$$

As a consequence of this theorem, we may concentrate on the mathematically simpler estimators $\hat{\theta}_n^{\mathrm{B,dj}}$ and $\hat{\theta}_n^{\mathrm{B,sl}}$ in the following asymptotic analysis. We will shortly write $\hat{\theta}_n^{\mathrm{dj}}$ and $\hat{\theta}_n^{\mathrm{sl}}$, respectively. Note that, while $\hat{\theta}_n^{\mathrm{sl}}$ is based on a substantially larger number of blocks than the disjoint blocks estimator, the blocks are heavily correlated. The following theorem is the central result of this paper and shows that both estimators are consistent and converge at the same rate to a normal distribution. The disjoint blocks estimator has a larger asymptotic variance than the sliding blocks estimator [see also Robert, Segers and Ferro (2009)].

THEOREM 3.2. *Suppose that Condition* 2.1 *and* (2.2) *is met. Then*

$$\sqrt{k_n}(\hat{\theta}_n^{\mathrm{dj}} - \theta) \rightsquigarrow \mathcal{N}(0, \theta^4 \sigma_{\mathrm{dj}}^2) \quad and \quad \sqrt{k_n}(\hat{\theta}_n^{\mathrm{sl}} - \theta) \rightsquigarrow \mathcal{N}(0, \theta^4 \sigma_{\mathrm{sl}}^2),$$

where

$$\sigma_{\mathrm{dj}}^2 = 4 \int_0^1 \frac{\mathrm{E}[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}]}{(1+\sigma)^3}\, d\sigma + 4\theta^{-1} \int_0^1 \frac{\mathrm{E}[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)]}{(1+\sigma)^3}\, d\sigma - \theta^{-2},$$

$$\sigma_{\mathrm{sl}}^2 = 4 \int_0^1 \frac{\mathrm{E}[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}]}{(1+\sigma)^3}\, d\sigma + 4\theta^{-1} \int_0^1 \frac{\mathrm{E}[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)]}{(1+\sigma)^3}\, d\sigma - \frac{4 - 4\log(2)}{\theta^2},$$

with $(\zeta_1^{(\sigma)}, \zeta_2^{(\sigma)}) \sim \pi_2^{(\sigma)}$. *In particular,* $\sigma_{\mathrm{dj}}^2 = \sigma_{\mathrm{sl}}^2 + \{3 - 4\log(2)\}/\theta^2 \approx \sigma_{\mathrm{sl}}^2 + 0.2274/\theta^2$.

It is interesting to note that the asymptotic variance of the disjoint blocks estimator is substantially more complicated than if one would naively treat the $\hat{Z}_{ni}$ (or the $\hat{Y}_{ni}$) as an i.i.d. sample from the exponential distribution with parameter θ [as is done in Northrop (2015); the variance would then simply be θ^2]. A heuristic explanation can be found in Remark 3.4 below. A formal proof is given at the end of this section, with several auxiliary lemmas postponed to Section 9 (for the disjoint blocks estimator) and to Section 10 (for the sliding blocks estimator). Explicit calculations are possible for instance for a max-autoregressive process; see Section 6.1, or for the i.i.d. case.

EXAMPLE 3.3. If the time series is serially independent, a simple calculation shows that $\pi(i) = \mathbf{1}(i = 1)$ and $\pi_2^{(\sigma)}(i, j) = (1 - \sigma)\mathbf{1}(i = 1, j = 0) + \sigma\mathbf{1}(i = 1, j = 1)$. This implies

$$\theta = 1, \mathrm{E}[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}] = \sigma, \qquad \mathrm{E}[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)] = 1 - \sigma$$

and, therefore, $\theta^4\sigma_{\mathrm{dj}}^2 = 1/2$ and $\theta^4\sigma_{\mathrm{sl}}^2 \approx 0.2726$. It is worthwhile to mention that these values are smaller than the variances of any of the disjoint and sliding blocks estimators considered in Robert, Segers and Ferro (2009), respectively. Moreover, note that asymptotic variance of the oracle $\tilde{\theta}_n^{\mathrm{dj}}$ is equal to $\theta^2 = 1$, which is twice as large as when the marginal c.d.f. is estimated. Finally, it can be seen that the same formulas are valid whenever $\theta = 1$: the fact that $\theta^{-1} \geq \sum_{i=1}^{\infty} i\pi(i)$ implies that $\pi(1) = 1$. By (A.9) in the Supplementary Material [Berghaus and Bücher (2018)], we then obtain $\pi_2^{(\sigma)} = (1 - \sigma)\mathbf{1}(i = 1, j = 0) + \sigma\mathbf{1}(i = 1, j = 1)$.

REMARK 3.4 (Main idea for the proof). Define $Z_{ni} = b_n(1 - N_{ni})$ and

$$(3.1) \qquad \hat{T}_n^{\mathrm{dj}} = \frac{1}{k_n}\sum_{i=1}^{k_n} \hat{Z}_{ni}, \qquad T_n^{\mathrm{dj}} = \frac{1}{k_n}\sum_{i=1}^{k_n} Z_{ni},$$

$$(3.2) \qquad \hat{T}_n^{\mathrm{sl}} = \frac{1}{n - b_n + 1}\sum_{t=1}^{n-b_n+1} \hat{Z}_{nt}^{\mathrm{sl}}, \qquad T_n^{\mathrm{sl}} = \frac{1}{n - b_n + 1}\sum_{t=1}^{n-b_n+1} Z_{nt}^{\mathrm{sl}}.$$

In the following, we only consider the disjoint blocks estimator, the argumentation for the sliding blocks estimator is similar. For the ease of notation, we will skip the upper index and just write $\hat{T}_n$ instead of $\hat{T}_n^{\mathrm{dj}}$, etc. Asymptotic normality of $\hat{\theta}_n$ may be deduced from the delta method and weak convergence of $\sqrt{k}(\hat{T}_n - \theta^{-1})$. The road map to handle the latter is as follows: decompose

$$(3.3) \qquad \sqrt{k_n}(\hat{T}_n - \theta^{-1}) = \sqrt{k_n}(\hat{T}_n - T_n) + \sqrt{k_n}(T_n - \theta^{-1}).$$

Using a big-block/small-block type argument, the asymptotics of the second summand on the right-hand side can be deduced from a central limit theorem for row-wise independent triangular arrays. Depending on the choice of the block sizes, an asymptotic bias term may appear, which we control by Condition 2.1(vii). The first summand is more involved, and also contributes to the limiting distribution: first, for $x \geq 0$, let

$$(3.4) \qquad e_n(x) = \frac{1}{\sqrt{k_n}} \sum_{s=1}^{n} \{\mathbf{1}(U_s > 1 - x/b_n) - x/b_n\}$$

denote the tail empirical process of $X_1, \ldots, X_n$ and let

$$(3.5) \qquad \hat{H}_{k_n}(x) = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbf{1}(Z_{ni} \leq x)$$

be the empirical distribution function of $Z_{n1}, \ldots, Z_{nk_n}$. Then

$$\sqrt{k_n}(\hat{T}_n - T_n) = \frac{b_n}{\sqrt{k_n}} \sum_{i=1}^{k_n} (N_{ni} - \hat{N}_{ni})$$

$$(3.6) \qquad = \frac{b_n}{n\sqrt{k_n}} \sum_{i=1}^{k_n} \sum_{s=1}^{n} \{N_{ni} - \mathbf{1}(U_s \leq N_{ni})\}$$

$$= \frac{1}{k_n^{3/2}} \sum_{i=1}^{k_n} \sum_{s=1}^{n} \{\mathbf{1}(U_s > 1 - Z_{ni}/b_n) - Z_{ni}/b_n\}$$

$$= \frac{1}{k_n} \sum_{i=1}^{k_n} e_n(Z_{ni}) = \int_0^{\max_{i=1}^{k_n} Z_{ni}} e_n(x) \, \mathrm{d}\hat{H}_{k_n}(x).$$

Since Z_{ni} is approximately exponentially distributed with parameter θ, one may expect that $\hat{H}_{k_n}(x)$ converges to $H(x) = 1 - \exp(-\theta x)$ in probability, for $n \to \infty$ and for any $x \geq 0$. Moreover, on an appropriate domain, $e_n \rightsquigarrow e$ for some Gaussian process e [Drees (2000, 2002), Drees and Rootzén (2010), Robert (2009), Rootzén (2009)], whence a candidate limit for the expression on the left-hand side of the previous display is given by $\int_0^\infty e(x)\theta e^{-\theta x} \, \mathrm{d}x$. The latter distribution is normal, and joint convergence of both terms on the right-hand side of (3.3) will finally allow for the derivation of the asymptotic distribution of $\hat{\theta}_n$. These heuristic arguments have to be made rigorous.

REMARK 3.5 (Disjoint blocks: Alternative proof).   As pointed out by a referee, the asymptotic distribution of the disjoint blocks estimator may alternatively be derived by completely relying on results in Robert (2009). The idea is as follows. First, recall (1.5), where $\hat{p}_n^{(\tau)}(0) = k_n^{-1} \sum_{j=1}^{k_n} \mathbf{1}(\hat{Z}_{ni} > \tau)$ for $\tau > 0$. Since $\hat{F}_n(x) < p$ if and only if $x < \hat{F}_n^{\leftarrow}(p)$, this expression coincides with the definition of $\hat{p}_n^{(\tau)}(0)$ used in Robert (2009), in the middle of page 275, up to a ">"-sign replaced by a "≥"-sign in his definition of $\hat{N}_{r_n,j}^{(\tau)}$. Hence, by Theorem 4.2 in that reference, assuming the latter replacement to be asymptotically negligible, we have $\sqrt{k_n}\{\hat{p}_n^{(\cdot)}(0) - p^{(\cdot)}(0)\} \rightsquigarrow \hat{e}_0(\cdot)$ in some appropriate metric space, where $\hat{e}_0$ is a Gaussian process. The continuous mapping theorem implies

$$\sqrt{k_n}\left\{\int_0^\cdot p_n^{(\tau)}(0)\,\mathrm{d}\tau - \int_0^\cdot p^{(\tau)}(0)\,\mathrm{d}\tau\right\} \rightsquigarrow \int_0^\cdot \hat{e}_0(\tau)\,\mathrm{d}\tau,$$

again on some appropriate metric space. Some tedious, but straightforward calculations show that the random variable $\lim_{t\to\infty} \int_0^t \hat{e}_0(\tau)\,\mathrm{d}\tau$ has the same law as the limit that we obtained with the approach stated in Remark 3.4. We do not give any further details on this approach as it is limited to the case of disjoint blocks.

REMARK 3.6 (On continuity of F).   In the Introduction, we assumed for simplicity that F is continuous. Some thoughts reveal that the main limit relations motivating the estimators, that is, (1.1) and (1.2), continue to hold under the weaker assumption that

$$\lim_{x \to x_F} \frac{1 - F(x-)}{1 - F(x)} = 1,$$

where x_F denotes the right endpoint of the support of F. By Theorem 1.7.13 in Leadbetter, Lindgren and Rootzén (1983), this condition is also necessary for the extremal index to exist. However, the proofs of our theoretical results do not easily generalize to this weaker assumption, the reason being that we heavily rely on the asymptotic equivalence of e_n in (3.4) and $\bar{e}_n$ on page 281 in Robert (2009) (to apply his Theorem 4.1 on weak convergence of $\bar{e}_n$) and on centredness of e_n on $[0, \varepsilon b_n]$ (to show negligibility of certain terms in Lemmas 9.1 and 10.1). A further discussion is beyond the scope of this paper.

PROOF OF THEOREM 3.2 (DISJOINT BLOCKS).   Write $\hat{T}_n = \hat{T}_n^{\mathrm{dj}}$ and $T_n = T_n^{\mathrm{dj}}$. Recall the definitions of e_n and $\hat{H}_{k_n}$ in (3.4) and (3.5), respectively. For $\ell \in \mathbb{N}$, let

$$D_n = \int_0^{\hat{m}} e_n(x)\,\mathrm{d}\hat{H}_{k_n}(x), \qquad D_{n,\ell} = \int_0^\ell e_n(x)\,\mathrm{d}\hat{H}_{k_n}(x),$$

$$D_\ell = \int_0^\ell e(x)\theta e^{-\theta x}\,\mathrm{d}x,$$

where $\hat{m} = \max Z_{ni}$. Also, let $G_n = \sqrt{k_n}(T_n - \mathrm{E}\, T_n)$ and let G be defined as in Lemma 9.3. Suppose we have shown that:

  (i) For all $\delta > 0$: $\lim_{\ell \to \infty} \limsup_{n \to \infty} \mathbb{P}(|D_{n,\ell} - D_n| > \delta) = 0$.
  (ii) For all $\ell \in \mathbb{N}$: $D_{n,\ell} + G_n \rightsquigarrow D_\ell + G$ as $n \to \infty$.
  (iii) $D_\ell + G \rightsquigarrow D + G \sim \mathcal{N}(0, \sigma_{\mathrm{dj}}^2)$ as $\ell \to \infty$.

It then follows from (3.6) and Wichura's theorem [Billingsley (1979), Theorem 25.5] that

$$\sqrt{n}(\hat{T}_n - \mathrm{E}\, T_n) = D_n + G_n \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{dj}}^2), \qquad n \to \infty.$$

By Condition 2.1(vii), we obtain that $\sqrt{k_n}(\hat{T}_n - \theta^{-1}) \rightsquigarrow \mathcal{N}(0, \sigma_{\mathrm{dj}}^2)$. The theorem then follows from the delta-method.

The assertion in (i) is proved in Lemma 9.1. The assertion in (ii) is proved in Lemma 9.5 (it is a consequence of the continuous mapping theorem and Lemmas 9.2 and 9.4), The assertion in (iii) follows from the fact that $D_\ell + G$ is normally distributed with variance σ_ℓ^2 as specified in Lemma 9.5, and the fact that by Lemma 9.6 $\sigma_\ell^2 \to \sigma_{\mathrm{dj}}^2$ for $\ell \to \infty$. $\square$

PROOF OF THEOREM 3.2 (SLIDING BLOCKS). Let $\hat{H}_{k_n}^{\mathrm{sl}}$ denote the empirical distribution function of the Z_{nt}^{sl}, $\hat{H}_{k_n}^{\mathrm{sl}}(x) = \frac{1}{n - b_n + 1} \sum_{t=1}^{n - b_n + 1} \mathbf{1}(Z_{nt}^{\mathrm{sl}} \le x)$, and let

$$D_n^{\mathrm{sl}} = \int_0^{\hat{m}^{\mathrm{sl}}} e_n(x) \, \mathrm{d}\hat{H}_{k_n}^{\mathrm{sl}}(x), \qquad D_{n,\ell}^{\mathrm{sl}} = \int_0^\ell e_n(x) \, \mathrm{d}\hat{H}_{k_n}^{\mathrm{sl}}(x),$$

$$D_\ell^{\mathrm{sl}} = \int_0^\ell e(x) \theta e^{-\theta x} \, \mathrm{d}x,$$

where $\hat{m}^{\mathrm{sl}} = \max_t Z_{nt}^{\mathrm{sl}}$. With this notation, the proof follows along the same lines as for the disjoint blocks, with Lemmas 9.1, 9.2 and 9.3 replaced by Lemmas 10.1, 10.2 and 10.3, respectively. $\square$

4. Variance estimation.

For statistical inference on θ, estimators for the asymptotic variance formulas in Theorem 3.2 are needed. Unfortunately, the formulas themselves are too complicated to base such estimators on a simple plug-in principle. Rather than that, we rely on an asymptotic expansion of the disjoint blocks estimator resulting from a careful inspection of the proofs. Note that, since $\sigma_{\mathrm{dj}}^2 = \sigma_{\mathrm{sl}}^2 - \{3 - 4\log(2)\}/\theta^2$, an estimator for the variance of the disjoint blocks estimator can immediately be transferred into one for the sliding blocks estimator. This is particularly useful since a straightforward extension of our proposed estimator for σ_{dj}^2 to the sliding blocks estimator is not possible and would require the choice of an additional tuning parameter.

The proof of Theorem 3.2, in particular the central decomposition in (3.3) and the calculations in (3.6), allows us to write $\mathbb{T}_n^{\mathrm{dj}} = \sqrt{k_n}(\hat{T}_n^{\mathrm{dj}} - \theta^{-1})$ as

$$\frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} (Z_{nj} - \theta^{-1}) + \int_0^\infty e_n(x) \, dH(x) + o_\mathbb{P}(1) = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} B_{nj} + o_\mathbb{P}(1),$$

where

$$B_{nj} = Z_{nj} - \theta^{-1} + \int_0^\infty \sum_{s \in I_j} \left\{ \mathbf{1}\left(U_s > 1 - \frac{x}{b_n} \right) - \frac{x}{b_n} \right\} dH(x)$$

and where $I_j = \{(j-1)k_n + 1, \ldots, jk_n\}$ denotes the jth block of indices. The proof of Theorem 3.2 shows that $B_{n1}, \ldots, B_{nk_n}$ are asymptotically independent (big block/small block heuristics) and centred, and that their empirical mean multiplied by $\sqrt{k_n}$ converges to a centred normal distribution with variance σ_{dj}^2. Hence, their second empirical moment should be a consistent estimator for σ_{dj}^2. As the sample $B_{n1}, \ldots, B_{nk_n}$ depends on unknown quantities, we must replace these objects by empirical counterparts, leading us to define

$$\hat{B}_{nj} = \hat{Z}_{nj} - \hat{T}_n + \sum_{s \in I_j} \frac{1}{k_n} \sum_{i=1}^{k_n} \left\{ \mathbf{1}\left(\hat{U}_s > 1 - \frac{\hat{Z}_{ni}}{b_n} \right) - \frac{\hat{Z}_{ni}}{b_n} \right\}$$

$$= \hat{Z}_{nj} + \sum_{s \in I_j} \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbf{1}\left(\hat{U}_s > 1 - \frac{\hat{Z}_{ni}}{b_n} \right) - 2 \cdot \hat{T}_n^{\mathrm{dj}},$$

where $\hat{U}_s = \hat{F}_n(X_s)$. The following proposition shows that

$$\hat{\sigma}_{\mathrm{dj}}^2 = \frac{1}{k_n} \sum_{j=1}^{k_n} \hat{B}_{nj}^2, \qquad \hat{\sigma}_{\mathrm{sl}}^2 = \hat{\sigma}_{\mathrm{dj}}^2 - \{3 - 4\log(2)\}(\hat{\theta}_n^{\mathrm{sl}})^{-2}$$

are in fact consistent estimators for σ_{dj}^2 and σ_{sl}^2, respectively, provided that moments of order slightly larger than 4 exist. To simplify the proofs, we assume beta-mixing of the times series, since it allows for stronger coupling results than alpha-mixing. We also impose a further growth condition on the block size, which allows for a further simplification within the proof [which is given in in the Supplementary Material, Berghaus and Bücher (2018)].

PROPOSITION 4.1 (Consistency of variance estimators). *Additionally to the assumptions imposed in Condition 2.1 suppose that $b_n = o(k_n^2)$ for $n \to \infty$ (hence, $b_n^{1/2} \ll k_n \ll b_n^2$), that Condition 2.1(iii) is met with the alpha-mixing coefficient $\alpha_{c_2}(\ell)$ replaced by the beta-mixing coefficient $\beta_1(\ell)$ (see the proof for a precise definition) and that Condition 2.1(ii) and (vi) are met with $\delta > 2$. Then, as $n \to \infty$,*

$$\hat{\sigma}_{\mathrm{dj}}^2 \xrightarrow{P} \sigma_{\mathrm{dj}}^2 \quad and \quad \hat{\sigma}_{\mathrm{sl}}^2 \xrightarrow{P} \sigma_{\mathrm{sl}}^2.$$

5. Bias reduction. While the previous sections were concerned with the $O(1/\sqrt{k_n})$-asymptotics, we will now have a heuristic look at the $O(1/k_n)$-asymptotics, in particular in terms of expectations. As a result, we will obtain a bias reduction scheme. Let $(\hat{T}_n, T_n, \sigma^2) \in \{(\hat{T}_n^{\mathrm{dj}}, T_n^{\mathrm{dj}}, \sigma_{\mathrm{dj}}^2), (\hat{T}_n^{\mathrm{sl}}, T_n^{\mathrm{sl}}, \sigma_{\mathrm{sl}}^2)\}$ denote any of the quantities defined in (3.1), (3.2) or Theorem 3.2. A Taylor expansion allows us to write

$$\hat{T}_n^{-1} - \theta = -\theta^2(\hat{T}_n - T_n) - \theta^2(T_n - \theta^{-1}) + \theta^3(\hat{T}_n - \theta^{-1})^2 + O_{\mathbb{P}}(k_n^{-3/2})$$

$$\equiv a_{n1} + a_{n2} + a_{n3} + O_{\mathbb{P}}(k_n^{-3/2}).$$

Let $\mu_{nj} = \mathrm{E}[a_{nj}]$. The second component μ_{n2} is inherent to the time series $(X_s)_{s \in \mathbb{N}}$ itself. In many examples, it can be seen to be of the order $O(b_n^{-1})$; see, for instance, Section 6 or similar calculations made in Robert, Segers and Ferro (2009), Section 6. Since $\sqrt{k_n}(\hat{T}_n - \theta^{-1}) \rightsquigarrow \mathcal{N}(0, \sigma^2)$, it seems plausible that the third component μ_{n3} satisfies $\mu_{n3} = k_n^{-1}\theta^3\sigma^2 + o(k_n^{-1})$, though we will not give a precise proof. Finally, consider the first component μ_{n1}, which is essentially due to the use of the empirical distribution function in the definition of the estimator. The following lemma gives a first-order asymptotic expansion, which turns out to be the same for the disjoint and sliding blocks estimator.

LEMMA 5.1. *Additionally to the conditions of Theorem 3.2 suppose that Condition 2.1(iii) is met with $c_2 = 1$. Then*

$$\lim_{n \to \infty} k_n \mathrm{E}[\hat{T}_n - T_n] = -\frac{1}{\theta},$$

where $(\hat{T}_n, T_n) \in \{(\hat{T}_n^{\mathrm{dj}}, T_n^{\mathrm{dj}}), (\hat{T}_n^{\mathrm{sl}}, T_n^{\mathrm{sl}})\}$ as defined in (3.1) and (3.2).

The proof is given in Appendix D. As a consequence, we have $\mu_{n1} = k_n^{-1}\theta + o(k_n^{-1})$. Now, plugging-in $\hat{\theta}_n$ and $\hat{\sigma}_n^2$ as a consistent estimator for θ and σ^2, we can estimate μ_{n1} and μ_{n3} by $\hat{\mu}_{n1} = k_n^{-1}\hat{\theta}_n$ and $\hat{\mu}_{n3} = k_n^{-1}\hat{\theta}_n^3\hat{\sigma}_n^2$, respectively. Subtracting these expression from $\hat{\theta}_n$, we obtain the bias-reduced estimator

$$\hat{\theta}_{n,bc} = \hat{\theta}_n - k_n^{-1}\hat{\theta}_n - k_n^{-1}\hat{\theta}_n^3\hat{\sigma}_n^2.$$

The $O(1/\sqrt{k_n})$-asymptotics will not be affected, but $\hat{\theta}_{n,bc}$ shows a better finite-sample performance and is therefore used in Section 7.

Note that if we are additionally willing to assume that $k_n\mu_{n2} = k_n \mathrm{E}[T_n - \theta^{-1}] = k_n \mathrm{E}[Z_{1:b_n} - \theta^{-1}] = o(1)$ as $n \to \infty$ [cf. Condition 2.1(vii)], we obtain that μ_{n1} and μ_{n3} are in fact the dominating bias-components. In common models, the assumption $k_n \mathrm{E}[T_n - \theta^{-1}] = o(1)$ is satisfied as soon as $k_n/b_n = o(1)$ (see Section 6). In comparison to the assumption $k_n/b_n^2 = o(1)$, in Condition 2.1(iii) larger block sizes are required. Similar assumptions have also been made for the bias reductions in Robert, Segers and Ferro (2009).

Finally, note that the bias reduction based on $\hat{\mu}_{n1}$ can actually be alternatively motivated by the fact that $\hat{\theta}_n^{\text{dj}} - k_n^{-1}\hat{\theta}_n^{\text{dj}}$ is equal to $(k_n^{-1}\sum_{i=1}^{k_n}\tilde{Z}_{ni})^{-1}$, where $\tilde{Z}_{ni} = b_n(1 - \hat{F}_{n,-i}(M_{ni}))$ with $\hat{F}_{n,-i}$ being the empirical c.d.f. of $(X_s)_{s\notin I_i}$. The idea of using $\hat{F}_{n,-i}$ rather than $\hat{F}_n$ has been used in Northrop (2015) as a bias reduction scheme.

6. Examples. Two examples are worked out in this section. For the max-autoregressive processes, considered in Section 6.1, explicit calculations for the asymptotic variance formulas in Theorem 3.2 are possible. These allow for a theoretical comparison with the blocks estimators from Robert (2009) and Robert, Segers and Ferro (2009). All assumptions imposed in Condition 2.1 are shown to hold. In Section 6.2, we consider solutions of stochastic difference equations such as ARCH-processes. Complementing results from Robert (2009) we show that Condition 2.1(iv) is satisfied.

6.1. *Max-autoregressive processes.* Consider the max-autoregressive process of order one, ARMAX(1) in short, defined by the recursion

$$X_s = \max\{\alpha X_{s-1}, (1-\alpha)Z_s\}, \qquad s \in \mathbb{Z},$$

where $\alpha \in [0, 1)$ and where $(Z_s)_s$ denotes an i.i.d. sequence of standard Fréchet random variables. A stationary solution of this recursion is given by

$$X_s = \max_{j\geq 0}(1-\alpha)\alpha^j Z_{s-j},$$

which shows that the stationary distribution is standard Fréchet as well. The sequence has extremal index $\theta = 1 - \alpha$ and its cluster size distribution is geometric, that is, $\pi(j) = \alpha^{j-1}(1-\alpha)$ for $j \geq 1$ [see, e.g., Chapter 10 in Beirlant et al. (2004)]. Moreover, it follows from Proposition 5.3.7 in Hsing (1984) and some simple calculations that

$$\begin{aligned}
\pi_2^{(\sigma)}(j_1, j_2) &= \alpha^{j_2-1}\{(\sigma - \alpha^{j_1-j_2+1})\mathbf{1}(\alpha^{j_1-j_2+1} < \sigma \leq \alpha^{j_1-j_2}) \\
&\quad + (\alpha^{j_1-j_2} - \alpha\sigma)\mathbf{1}(\alpha^{j_1-j_2} < \sigma \leq \alpha^{j_1-j_2-1})\} \\
&= \alpha^{j_2-1}\{(\sigma - \alpha^{z+1})\mathbf{1}(j_1 = j_2 + z) \\
&\quad + (\alpha^{z+1} - \alpha\sigma)\mathbf{1}(j_1 = j_2 + z + 1)\}
\end{aligned}$$

for $j_1 \geq j_2 > 0$, where $z = \lfloor \log\sigma/\log\alpha \rfloor \in \mathbb{N}_0$. The formula in Proposition 5.3.7 in Hsing (1984) is wrong for $j_2 = 0$, but can be corrected to

$$\pi_2^{(\sigma)}(j_1, 0) = (1-\alpha)\alpha^{j_1-1}\mathbf{1}(j_1 \leq z) + (\alpha^z - \sigma)\mathbf{1}(j_1 = 1 + z)$$

for $j_1 \geq 1$. Based on these formulas, some straightforward calculations yield

$$\mathrm{E}[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}] = \frac{\alpha^{z+1} + \sigma\{1 + z(1-\alpha)\}}{(1-\alpha)^2}$$

and

$$E[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)] = \frac{1-\alpha^{z+1}}{1-\alpha} - \sigma(z+1).$$

Note that, for $\alpha \to 0$, we obtain $E[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}] \to \sigma$ and $E[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)] \to 1-\sigma$, which corresponds to the i.i.d. scenario. The latter two displays imply

$$E[\zeta_1^{(\sigma)}\zeta_2^{(\sigma)}] + \theta^{-1}E[\zeta_1^{(\sigma)}\mathbf{1}(\zeta_2^{(\sigma)}=0)] = \frac{1+\alpha\sigma}{(1-\alpha)^2}$$

and hence

$$\sigma_{\mathrm{dj}}^2 = \frac{1+\alpha}{2(1-\alpha)^2}, \qquad \sigma_{\mathrm{sl}}^2 = \frac{8\log 2 - 5 + \alpha}{2(1-\alpha)^2}.$$

Since $\theta = 1 - \alpha$, the asymptotic variances of $\sqrt{k_n}(\hat{\theta}_n/\theta - 1)$ simply reduce to the affine linear functions $(1+\alpha)/2$ and $(8\log 2 - 5 + \alpha)/2$ for the disjoint and the sliding blocks estimator, respectively. These functions can be compared with the asymptotic variance formulas in Robert, Segers and Ferro (2009), Formula 5.1, and in Robert (2009), page 285, variance of $\hat{\theta}_{1,n}^{(\tau)}$. Note that the variance of $\hat{\theta}_{1,n}^{(\tau)}$ in Robert (2009) is exactly the same as the one of the disjoint blocks estimator in Robert, Segers and Ferro (2009). The asymptotic variance formulas depend on an additional parameter $\tau > 0$ to be chosen by the statistician. Assuming we would have access to the optimal value (which can be calculated numerically, but must be estimated in practice), we obtain the variance curves depicted in Figure 1. We observe that, for the ARMAX-model, the PML-estimators analyzed in this paper have a smaller asymptotic variance than the (theoretically optimal) estimators in Robert, Segers and Ferro (2009) and Robert (2009).

Regarding the additional assumptions in Condition 2.1, some tedious calculations show that Condition 2.1(ii) is satisfied for $\delta = 1$. $(X_s)_{s \in \mathbb{Z}}$ can further be shown to be a geometrically ergodic Markov chain; see Formula (3.5) in Bradley (2005). As a consequence of Theorem 3.7 in that reference, $(X_s)_{s \in \mathbb{Z}}$ is geometrically β-mixing, whence Condition 2.1(iii) is satisfied (and also the condition on beta-mixing imposed in Proposition 4.1). It can be further be shown that, with $U_s = \exp(-1/X_s)$, we have $\mathrm{Var}\{\sum_{s=1}^n \mathbf{1}(U_s > 1 - y)\} \le ny\{1 + 2\alpha/(1-\alpha)\}$ for all $y \in (0,1)$, that is, Condition 2.1(iv) is met. Moreover, a simple calculation shows that $\mathbb{P}(\min_{i=1}^{2k_n} N'_{ni} \le c) \le 2k_n \mathbb{P}(N'_{n1} \le c) = O(k_n c^{(1-\alpha)b_n/2}) = o(1)$, provided that $k_n = o(b_n^2)$. Hence, Condition 2.1(v). Based on an explicit calculation of the distribution of Z_{n1}, it can also be seen that Condition 2.1(vi) is satisfied for any $\delta > 0$, and that $E[Z_{1:b_n}] - \theta^{-1} = O(b_n^{-1})$. The latter implies that Condition 2.1(vii) is satisfied if $k_n = o(b_n^2)$; see (2.1). It can easily be seen that (2.2) is met.
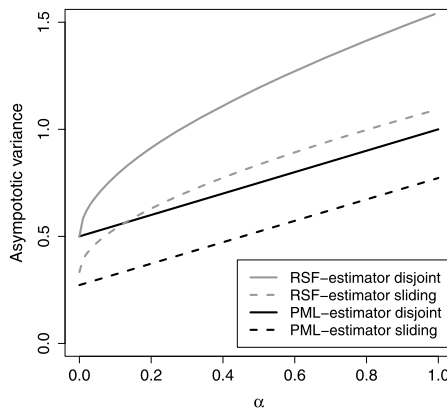
FIG. 1.   *Asymptotic variances of $\sqrt{k_n}(\hat{\theta}_n/\theta - 1)$ within the ARMAX(α)-Model for the sliding and disjoint blocks estimators analyzed in this paper (PML) and in Robert, Segers and Ferro (2009) (RSF).*

6.2. *Stochastic difference equations.*   Consider the equation

$$(6.1) \qquad\qquad X_s = A_s X_{s-1} + B_s, \qquad s \in \mathbb{N},$$

where $(A_s, B_s)_s$ are i.i.d. $[0, \infty)^2$-valued random vectors. If $A_s = \alpha_1 Z_s^2$ and $B_s = \alpha_0 Z_s^2$ for some $\alpha_0, \alpha_1 > 0$ and some i.i.d. real-valued sequence $(Z_s)_s$, the above equation defines the popular (squared) ARCH(1)-time series model. For simplicity, we assume that the distribution of (A_1, B_1) is absolutely continuous.

The existence of a stationary solution of (6.1) as well as the tail behavior of the stationary distribution F of X_s has been studied in Kesten (1973), Theorem 5. More precisely, consider the condition:

(S) There exists some $\kappa > 0$ such that

$$\mathrm{E}\log A_1 < 0, \qquad \mathrm{E}\big[A_1^\kappa\big] = 1,$$

$$\mathrm{E}\big[A_1^\kappa \max(\log A_1, 0)\big] < \infty, \qquad \mathrm{E}\big[B_1^\kappa\big] \in (0, \infty).$$

Under this assumption, there exists a unique stationary solution of (6.1) and the c.d.f. F of X_s satisfies $1 - F(x) \sim cx^{-\kappa}$ as $x \to \infty$ for some constant $c > 0$. Moreover, F is continuous [Vervaat (1979), Theorem 3.2] and, in particular, in the max-domain of attraction of $G_{1/\kappa}$, the generalized extreme value distribution with extreme-value index $1/\kappa$.

Explicit calculations for the (two-level) cluster size distribution have been carried out in Perfekt (1994), Example 4.2. Unfortunately, the formulas are complicated and do not allow for simple expressions of the asymptotic variances in Theorem 3.2.

Slight adaptations of Assumptions (i)–(iii) of Condition 2.1 have been checked in Robert (2009), Example 3.1. We complement those results by showing that also

(iv) is satisfied. The result is inspired by Section 4 in Drees (2000) and is in fact a modification of Lemma 4.1 in that paper to the present needs. Its proof is given in Appendix D in the Supplementary Material [Berghaus and Bücher (2018)].

LEMMA 6.1. *Suppose that Condition* (S) *is met and let* $(X_s)_s$ *denote a stationary solution of* (6.1). *Then Condition* 2.1(iv) *is met.*

7. Finite-sample performance.   A simulation study is performed to illustrate the finite-sample performance of the proposed estimators and methods. Results are presented for four time series models:

- The *ARMAX-model* from Section 6.1:

$$X_s = \max\{\alpha X_{s-1}, (1-\alpha)Z_s\}, \qquad s \in \mathbb{Z},$$

  where $\alpha \in [0, 1)$ and where $(Z_s)_s$ is an i.i.d. sequence of standard Fréchet random variables. We consider $\alpha = 0, 0.25, 0.5, 0.75$ resulting in $\theta = 1, 0.75, 0.5, 0.25$.

- The *squared ARCH-model* from Section 6.2:

$$X_s = (2 \times 10^{-5} + \lambda X_{s-1})Z_s^2, \qquad s \in \mathbb{Z},$$

  where $\lambda \in (0, 1)$ and where $(Z_s)_s$ denotes an i.i.d. sequence of standard normal random variables. We consider $\lambda = 0.1, 0.5, 0.9, 0.99$ which implies $\theta = 0.997, 0.727, 0.460, 0.422$, respectively [Table 3.1 in de Haan et al. (1989)].

- The *ARCH-model*:

$$X_s = \left(2 \times 10^{-5} + \lambda X_{s-1}^2\right)^{1/2} Z_s, \qquad s \in \mathbb{Z},$$

  where $\lambda \in (0, 1)$ and where $(Z_s)_s$ denotes an i.i.d. sequence of standard normal random variables. We consider $\lambda = 0.1, 0.5, 0.7, 0.99$ which implies $\theta = 0.999, 0.835, 0.721, 0.571$, respectively [Table 3.2 in de Haan et al. (1989)].

- The *Markovian Copula-model* [Darsow, Nguyen and Olsen (1992)]:

$$X_s = F^{\leftarrow}(U_s), \qquad (U_s, U_{s-1}) \sim C_\vartheta, s \in \mathbb{Z}.$$

Here, $F^{\leftarrow}$ is the left-continuous quantile function of some arbitrary continuous c.d.f. F, $(U_s)_s$ is a stationary Markovian time series of order 1 and C_ϑ denotes the Survival Clayton Copula with parameter $\vartheta > 0$. For this model, $\theta = \mathbb{P}(\max_{t \geq 1} \prod_{s=1}^{t} A_s \leq U)$, where $U, A_1, A_2, \ldots$ are independent, U is standard uniform and A_s has c.d.f. $H_\vartheta(s) = 1 - (1 + s^\vartheta)^{-(1+1/\vartheta)}$, $s \geq 0$; see Perfekt (1994) or Beirlant et al. (2004), Section 10.4.2. We consider choices $\vartheta = 0.23, 0.41, 0.68, 1.06, 1.90$ such that (approximately) $\theta = 0.2, 0.4, 0.6, 0.8, 0.95$ and fix F as the standard uniform c.d.f. (the results are independent of this choice, as the estimators are rank-based). Algorithm 2 in Rémillard, Papageorgiou and Soustra (2012) allows us to simulate from this model.

Additional simulation results for the AR-model and the doubly stochastic process from Smith and Weissman (1994) turned out to be quite similar to the ARMAX-model and are not presented for the sake of brevity. In all scenarios under consideration, the sample size is fixed to $n = 8192 = 2^{13}$ and the block size b_n for the blocks estimators is chosen from the set $2^2, 2^3, \ldots, 2^9$.

7.1. *Comparison with other estimators for the extremal index.*   We present results for six different estimators: the bias-reduced sliding blocks estimator $\hat{\theta}_n^{\mathrm{B}}$, the sliding blocks estimator from Northrop (2015) (i.e., $\hat{\theta}_n^{\mathrm{N}}$, but with $\hat{F}_n$ replaced by $\hat{F}_{n,-i}$ in the ith block), the bias-reduced sliding blocks estimator from Robert, Segers and Ferro (2009) (with a data-driven choice of the threshold as outlined in Section 7.1 of that paper), the integrated version of the blocks estimator from Robert (2009), the intervals estimator from Ferro and Segers (2003) and the ML-estimator from Süveges (2007). Results for other versions of these estimators (e.g., the disjoint blocks versions or the versions based on a fixed threshold) are not presented as their performance was dominated by the above versions in almost all scenarios under consideration. The parameters σ and ϕ for the Robert-estimator [last display on page 276 of Robert (2009)] are chosen as $\sigma = 0.7$ and $\phi = 1.3$. The intervals estimator and the Süveges-estimator require the choice of a threshold u, which we choose as the $1 - 1/b_n$ empirical quantile of the observed data. All estimators are constrained to the interval $[0, 1]$, except for Table 1 where we also report results for the unconstrained versions.

In Figure 2 (ARCH), as well as in Appendix E of the Supplement Material Berghaus and Bücher (2018) (ARMAX, squared ARCH and Markovian Copula), we depict the mean-squared error $\mathrm{E}[(\hat{\theta} - \theta)^2]$ as a function of the block size parameter b, estimated on the basis of $N = 10{,}000$ simulation runs. For most models and estimators, the MSE-curves are U-shaped, representing the usual bias-variance tradeoff in extreme value theory (an exception being the Süveges-estimator within the ARCH-model for $\theta = 0.571$, a possible reason being its high bias due to fact that his central assumption $D^{(2)}$ is not satisfied in this model). Explicit pictures of the squared bias and variance can be found in Appendix E of the Supplementary Material. For the blocks estimators considered in this paper, the bias is decreasing in b (the asymptotics for the exponential distribution kick in), while the variance is increasing (the convergence rate of the estimators being $1/\sqrt{k_n}$). In terms of the bias, $\hat{\theta}_n^{\mathrm{N}}$ is clearly superior to $\hat{\theta}_n^{\mathrm{B}}$ for small block sizes.

The minimal values of the curves in Figure 2 are of particular interest, and are summarized in Table 1. We observe that the sliding blocks estimators $\hat{\theta}_n^{\mathrm{B}}$ and $\hat{\theta}_n^{\mathrm{N}}$ outperform the other two blocks estimators in most scenarios. For the ARMAX-model, this is in agreement with the theoretical findings presented in Figure 1. Comparing $\hat{\theta}_n^{\mathrm{B}}$ and $\hat{\theta}_n^{\mathrm{N}}$, we see that $\hat{\theta}_n^{\mathrm{N}}$ seems to be preferable in most scenarios. In general, there is no clear best estimator in terms of the MSE: $\hat{\theta}_n^{\mathrm{N}}$ wins six times, the Süveges-estimator six times, $\hat{\theta}_n^{\mathrm{B}}$ four times and the intervals estimator is best in one scenario.

TABLE 1
Minimal mean squared error multiplied with 10^3 for the ARMAX-model (top 4 rows), the squared ARCH-model (upper middle 4 rows), the ARCH-model (lower middle 4 rows) and the Markovian copula model (bottom 5 rows). The estimator with the (row-wise) smallest MSE is in boldface. Values in brackets refer to the unconstrained estimator

θ	$\hat{\theta}_n^{B,sl}$	$\hat{\theta}_n^{N,sl}$	RSF-sliding	Intervals	ML-Süveges	Robert
0.25	0.91	0.51	1.35	0.53	**0.22**	1.77
0.50	1.58	0.78	2.24	0.99	**0.63**	2.07
0.75	2.03	**0.67**	2.34	1.17	0.96	2.31
1.00	**0.00 (1.78)**	0.05 (0.11)	0.10 (0.12)	0.88	0.11	2.22
0.422	3.18	2.86	4.85	**2.53**	3.19	4.00
0.460	3.53	2.98	5.45	2.71	**1.92**	4.26
0.727	1.07	**0.46**	1.46	1.08	1.44	1.19
0.997	**0.01 (0.50)**	1.56	1.31 (1.33)	5.34	2.19	0.65
0.571	4.82	**4.81**	7.65	6.02	20.94	5.58
0.721	3.32	2.63	4.22	3.70	**0.28**	3.65
0.835	1.89	1.02	1.74	1.83	**0.31**	2.09
0.999	**0.00 (0.98)**	0.16 (0.17)	0.73 (0.76)	1.01	1.15	1.13
0.20	0.63	**0.52**	1.72	0.63	15.14	1.56
0.40	0.99	**0.68**	1.61	0.79	3.80	1.29
0.60	1.65	0.92	1.72	4.77	**0.43**	1.65
0.80	0.97	**0.18**	0.72	13.00	2.53	0.63
0.95	**0.82 (0.94)**	4.60	2.87	12.05 (12.50)	4.32	1.65 (2.06)

7.2. *Estimation of the asymptotic variance and coverage of confidence bands.* We consider the ARMAX-, squared ARCH- and ARCH-model as described above. We are interested in the performance of

$$\hat{\tau}_{dj}^2 = (\hat{\theta}_n^{dj})^4 \hat{\sigma}_{dj}^2 \quad \text{and} \quad \hat{\tau}_{sl}^2 = (\hat{\theta}_n^{sl})^4 \hat{\sigma}_{sl}^2$$

as estimators for the variances of $\sqrt{k_n}\hat{\theta}_n^{x,dj}$ and $\sqrt{k_n}\hat{\theta}_n^{x,sl}$, respectively, where $x \in \{B, N\}$. Results can be found in Figure 3 (as well as in Figures 16 and 17 of the Supplementary Material), where we depict the curves

$$b_n \mapsto E\left[\left(\frac{\hat{\tau}^2(b_n)}{\text{Var}(\sqrt{k_n}\hat{\theta}_n(b_n))} - 1\right)^2\right], \qquad b_n \mapsto E\left[\frac{\hat{\tau}^2(b_n)}{\text{Var}(\sqrt{k_n}\hat{\theta}_n(b_n))} - 1\right],$$

$(\hat{\tau}^2, \hat{\theta}_n) \in \{(\hat{\tau}_{dj}^2, \hat{\theta}_n^{B\,dj}), (\hat{\tau}_{sl}^2, \hat{\theta}_n^{B,sl}), (\hat{\tau}_{dj}^2, \hat{\theta}_n^{N,dj}), (\hat{\tau}_{sl}^2, \hat{\theta}_n^{N,sl})\}$, estimated on the basis of 10,000 simulation runs. Here, $\text{Var}(\sqrt{k_n}\hat{\theta}_n(b_n))$ is approximated by the empirical variance of $\sqrt{k_n}\hat{\theta}_n(b_n)$ over additional 10,000 simulations. Qualitatively, we observe a similar behaviour as for the estimation of θ depicted in Figure 2: the curves are U-shaped and possess a minimum at some intermediate values of b_n. Due to the fact that estimator $\hat{\tau}_{sl}^2$ is based on an additional estimation step (which
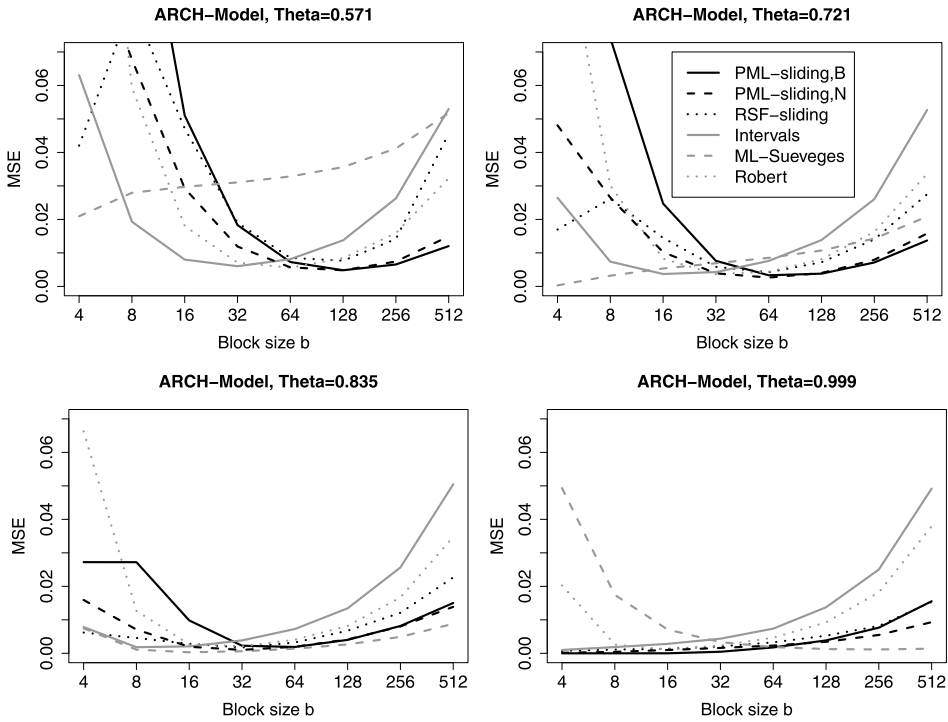
FIG. 2. *Mean squared error for the estimation of θ within the ARCH-model for four values of* θ ∈ {0.571, 0.721, 0.835, 0.999}.

is potentially biased, if b_n is small), the approximation works better for the disjoint blocks estimator. Also, the approximation is far better for $\hat{\theta}_n^B$ than for $\hat{\theta}_n^N$ (in particular for the bias), which may be explained by the fact that $\hat{\tau}_{sl}^2$ is based on an explicit expansion for $\hat{\theta}_n^B$. In particular, the fact that the bias of $\hat{\theta}_n^N$ is eventually increasing for larger block sizes may be explained by the $1/\sqrt{k_n}$-approximation of $\hat{\theta}_n^N$ by $\hat{\theta}_n^B$ (Theorem 3.1).

We are also interested in the coverage probabilities of the confidence sets

$$\text{CI}_{1-\alpha} = \left[\hat{\theta}_n - k_n^{-1/2}\hat{\tau}u_{1-\alpha/2}, \hat{\theta}_n + k_n^{-1/2}\hat{\tau}u_{1-\alpha/2}\right]$$

for θ, where $u_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution. Empirical coverage probabilities for $1 - \alpha = 0.95$ based on $N = 10,000$ simulation runs are presented in Tables 2 ($\hat{\theta}_n^B$-versions) and 3 ($\hat{\theta}_n^N$-versions), with coverage probabilities above 0.9 in boldface. Since the variance approximation is worse for $\hat{\theta}_n^N$, the coverage probabilities are worse as well. Moreover, it can be seen that the probabilities strongly depend on the block size b_n, with, for $\hat{\theta}_n^B$, at least one reasonable choice for every model, usually close to the MSE-minimal choice in Figure 2 (and Figure 5 in the Supplementary Material). The larger width of the confidence sets for the disjoint blocks estimator (not presented here; it is
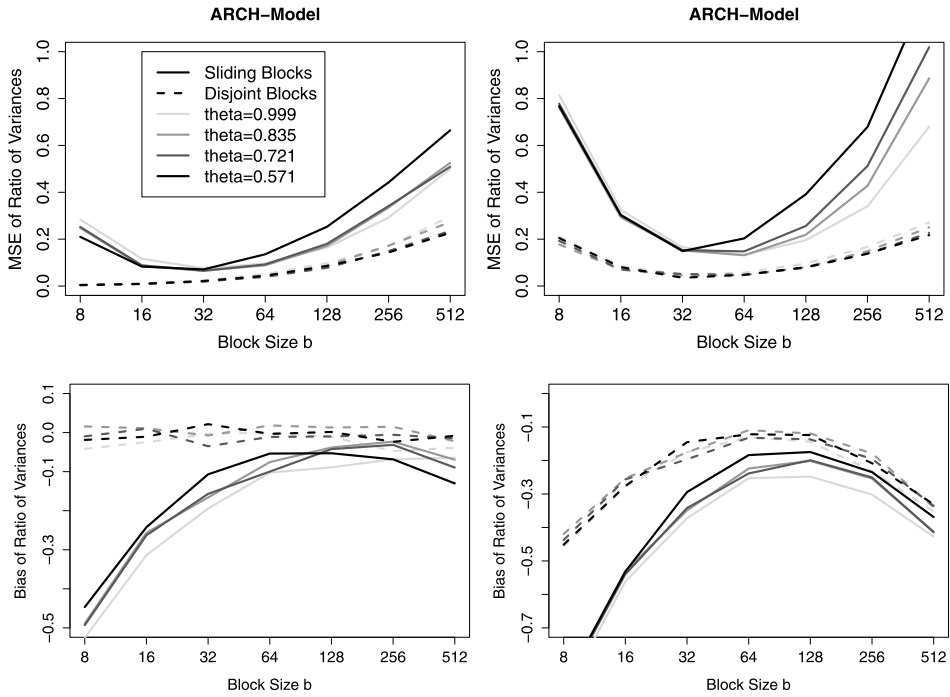
FIG. 3. *Mean squared error* $\mathrm{E}[(\hat{\tau}^2/\mathrm{Var}(\hat{\theta}_n) - 1)^2]$ *and bias* $\mathrm{E}[\hat{\tau}^2/\mathrm{Var}(\hat{\theta}_n)] - 1$ *within the ARCH-model for the unconstrained estimators* $\hat{\theta}_n^{\mathrm{B}}$ (*left*) *and* $\hat{\theta}_n^{\mathrm{N}}$ (*right*).

due to the larger variance) results in a slightly better performance compared to the sliding blocks estimator.

8. Case study. The use of the PML-estimators and the corresponding confidence sets is illustrated on negative daily log returns of a variety of financial market indices and prices including equity (e.g., S&P 500 Composite, MSCI World), commodities (e.g., TOPIX Oil & Coal, Gold Bullion LBM, Raw Sugar) and U.S. treasury bonds between 04 January 1990 and 30 December 2015 ($n = 6780$ observations for each index). Clusters of large negative returns can be financially damaging and are hence of interest for risk management.

In Figure 4, we depict estimates of the extremal index for four typical time series as a function of the block length parameter, ranging from $b = 10$ to $b = 357$. The solid curves correspond to the bias corrected sliding blocks estimator $\hat{\theta}_n^{\mathrm{B,sl}}$, alongside with a 95%-confidence band based on the variance estimator from Section 4 and the normal approximation. Interestingly, the curves appear to be quite smooth, which is a typical and nice property of the sliding blocks estimator. For comparison, the (far rougher) dashed lines correspond to the intervals estimator from Ferro and Segers (2003). As highlighted by many other authors, there is no simple optimal solution for the choice of the *best* block length parameter and a

TABLE 2
Empirical coverage probabilities of 95%-confidence bands of the constrained estimators $\hat{\theta}_n^B$. Values above 90% are in boldface

	b_n/θ	ARMAX-model				ARCH-model			
		0.25	0.5	0.75	1	0.571	0.721	0.835	0.999
Disjoint	16	0	0	0.13	**1.0**	0.00	0.00	0.04	**1.00**
	32	0.03	0.63	0.85	**0.99**	0.01	0.42	0.87	**0.97**
	64	0.80	**0.93**	**0.95**	**0.98**	0.68	**0.91**	**0.94**	**0.93**
	128	**0.94**	**0.94**	**0.94**	**0.95**	**0.93**	**0.94**	**0.92**	**0.91**
	256	**0.93**	**0.92**	**0.91**	**0.92**	**0.93**	**0.92**	**0.90**	0.89
	512	**0.91**	**0.90**	0.88	0.87	**0.90**	0.88	0.86	0.84
Sliding	16	0	0	0.02	**1.00**	0.00	0.00	0.00	**1.00**
	32	0.01	0.46	0.75	**1.00**	0.00	0.20	0.76	**0.95**
	64	0.71	**0.90**	**0.93**	**0.96**	0.53	0.86	**0.92**	0.89
	128	**0.92**	**0.93**	**0.92**	**0.92**	0.89	**0.92**	0.88	0.85
	256	**0.91**	0.89	0.87	0.86	**0.90**	0.88	0.84	0.81
	512	0.88	0.85	0.81	0.76	0.85	0.81	0.77	0.73

unique estimate for the extremal index. The dotted lines in Figure 4 correspond to case-by-case visual choices, trying to capture plateaus in the respective plots.

For the ease of comparison, this procedure has been repeated for all 20 time series under consideration (despite the fact that the entire curves provide a more detailed picture of the extremal dependence). In Table 4, we state the resulting

TABLE 3
Empirical coverage probabilities of 95%-confidence bands of the constrained Northrop estimators $\hat{\theta}_n^N$. Values above 90% are in boldface

|	b_n/θ	ARMAX-model				ARCH-model			
		0.25	0.5	0.75	1	0.422	0.46	0.727	0.997
Disjoint	16	0.00	0.47	0.84	**0.95**	0.00	0.01	0.62	0.82
	32	0.50	0.87	**0.92**	**0.96**	0.07	0.64	**0.91**	0.88
	64	0.88	**0.93**	**0.93**	**0.97**	0.69	**0.90**	**0.92**	**0.92**
	128	**0.91**	**0.92**	**0.92**	**0.96**	0.89	**0.92**	**0.92**	**0.94**
	256	**0.90**	**0.90**	**0.91**	**0.96**	**0.90**	**0.91**	**0.94**	**0.94**
	512	0.87	0.87	**0.91**	**0.94**	0.86	0.89	**0.92**	**0.93**
Sliding	16	0.00	0.22	0.70	**0.91**	0.00	0.00	0.32	0.62
	32	0.31	0.80	0.88	**0.94**	0.01	0.42	0.87	0.77
	64	0.82	0.89	**0.90**	**0.94**	0.54	0.84	0.88	0.85
	128	0.87	0.88	0.88	**0.92**	0.83	0.88	0.86	0.84
	256	0.85	0.84	0.81	0.86	0.83	0.83	0.80	0.76
	512	0.75	0.72	0.69	0.76	0.73	0.69	0.67	0.62
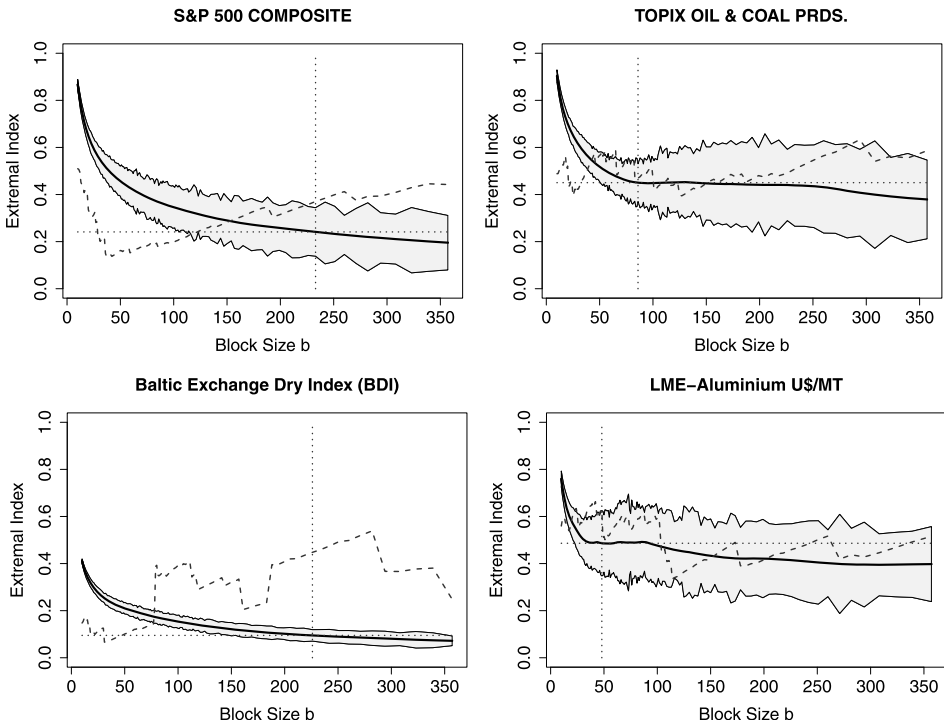
FIG. 4. *Extremal index estimates for four financial time series as a function of the block size. The solid line is the bias-reduced sliding blocks estimate; the shaded region is the pointwise 95%-confidence band. The dashed line is the intervals estimator. The dotted lines correspond to the selected block length based on visual inspection of the graphs.*

estimates of the extremal index and the width of the corresponding confidence intervals. Interestingly, the estimates of the extremal index lie around 0.3 for most of the equity indexes (S&P 500 Composite, MSCI World, etc.), while they are around 0.45 for many of the commodity prices (Coffee, Cotton, Aluminium). The smallest value of 0.12 is attained for the Baltic Exchange Dry Index, an index measuring the price of moving the major raw materials by sea and usually regarded as an efficient economic indicator of future economic growth and production.

9. Auxiliary lemmas for proving Theorem 3.2 (disjoint blocks).

LEMMA 9.1 (Approximation by an integral with bounded support). *Under Condition* 2.1, *for all $\delta > 0$,*

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} \mathbb{P}\big(|D_{n,\ell} - D_n| > \delta\big) = 0.$$

TABLE 4
Sliding Blocks Estimates of the extremal index and width of corresponding confidence intervals for negative daily log returns of 20 *financial market indices and prices*

Index/Prices	Extremal index	Width of C-interval
Raw Sugar Cents/lb	0.54	0.17
Coffee-Brazilian Cents/lb	0.49	0.13
LME-Aluminium U$/MT	0.49	0.14
Palladium U$/Troy Ounce	0.46	0.11
TOPIX OIL & COAL PRDS.	0.45	0.08
US T-Bill 10 YEAR	0.44	0.12
Cotton Cents/lb	0.42	0.12
S&P GSCI Precious Metal	0.42	0.12
MSCI WORLD EX US	0.36	0.11
Crude Oil-Brent Cur. Month	0.35	0.10
Gold Bullion LBM	0.33	0.10
RUSSELL 2000	0.31	0.09
S&P GSCI Commodity Total Return	0.30	0.09
S&P 500 COMPOSITE	0.29	0.10
LMEX Index	0.27	0.10
G12-DS Banks	0.26	0.09
G7-DS Banks	0.26	0.10
EU-DS Banks	0.26	0.08
S&P500 BANKS	0.22	0.08
Baltic Exchange Dry Index (BDI)	0.12	0.02

LEMMA 9.2 (Approximation by a Lebesgue integral). *Suppose that Condition* 2.1 *is met. Then, as* $n \to \infty$,

$$D_{n,\ell} = D'_{n,\ell} + o_{\mathbb{P}}(1) \qquad \text{where } D'_{n,\ell} = \int_0^\ell e_n(x)\theta e^{-\theta x}\,\mathrm{d}x.$$

LEMMA 9.3 (Joint convergence of fidis). *Under Condition* 2.1, *for any* $x_1, \ldots, x_m \in [0, \infty)$, *as* $n \to \infty$,

$$\big(e_n(x_1), \ldots, e_n(x_m), G_n\big)' \rightsquigarrow \big(e(x_1), \ldots, e(x_m), G\big)',$$

the random vector on the right-hand side being $\mathcal{N}_{m+1}(\mathbf{0}, \boldsymbol{\Sigma}^{\mathrm{dj}}(x_1, \ldots, x_m))$-*distributed with*

$$\boldsymbol{\Sigma}^{\mathrm{dj}}(x_1, \ldots, x_m) = \begin{pmatrix} r(x_1, x_1) & \ldots & r(x_1, x_m) & h(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \ldots & r(x_m, x_m) & h(x_m) \\ h(x_1) & \ldots & h(x_m) & \theta^{-2} \end{pmatrix}.$$

Here, $r(0,0) = h(0) = 0$ *and, for* $x \geq y \geq 0$ *with* $x \neq 0$,

$$r(x, y) = \theta x \sum_{i=1}^{\infty} \sum_{j=0}^{i} ij\pi_2^{(y/x)}(i, j), \qquad h(x) = \int_0^x \sum_{i=1}^{\infty} ip_2^{(x,y)}(i, 0)\,\mathrm{d}y - x/\theta,$$

where, for $i \geq j \geq 0, i \geq 1$,

$$p_2^{(x,y)}(i, j) = \mathbb{P}\{\mathbf{N}_E^{(x,y)} = (i, j)\}, \qquad \mathbf{N}_E^{(x,y)} = \sum_{i=1}^{\eta} (\zeta_{i1}^{(y/x)}, \zeta_{i2}^{(y/x)})$$

with $\eta \sim \mathrm{Poisson}(\theta x)$ *independent of i.i.d. random vectors* $(\zeta_{i1}^{(y/x)}, \zeta_{i2}^{(y/x)}) \sim \pi_2^{(y/x)}, i \in \mathbb{N}$.

LEMMA 9.4.   *Under Condition* 2.1, *as* $n \to \infty$,

$$\{(e_n(x), G_n)'\}_{x \in [0,\infty)} \rightsquigarrow \{(e(x), G)'\}_{x \in [0,\infty)}$$

in $D([0, \infty)) \times \mathbb{R}$, *where* $(e, G)'$ *is a centered Gaussian process with continuous sample paths and covariance functional as specified in Lemma* 9.3. *Here,* $D([0, \infty))$ *is equipped with the metric* $d(f, g) = \int_0^{\infty} e^{-t}[d_t(f, g) \wedge 1]\,\mathrm{d}t$ *where* d_t *denotes the* J_1*-metric applied to the restrictions of* f *and* g *to* $[0, t]$.

LEMMA 9.5.   *Under Condition* 2.1, *for any* $\ell \in \mathbb{N}$,

$$D_{n,\ell} + G_n \rightsquigarrow \mathcal{N}(0, \sigma_\ell^2),$$

as $n \to \infty$, *where*

$$\sigma_\ell^2 = \theta^2 \int_0^\ell \int_0^\ell r(x, y)e^{-\theta(x+y)}\,\mathrm{d}x\,\mathrm{d}y + 2\theta \int_0^\ell h(x)e^{-\theta x}\,\mathrm{d}x + \theta^{-2}.$$

LEMMA 9.6.   *Under Condition* 2.1, *as* $\ell \to \infty$,

$$\sigma_\ell^2 \to \sigma_{\mathrm{dj}}^2,$$

where σ_ℓ^2 *and* σ_{dj}^2 *are defined in Lemma* 9.5 *and Theorem* 3.2, *respectively.*

10. Auxiliary lemmas for proving Theorem 3.2 (sliding blocks).

LEMMA 10.1 (Approximation by an integral with bounded support—sliding blocks).   *Under Condition* 2.1, *for all* $\delta > 0$,

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} \mathbb{P}(|D_{n,\ell}^{\mathrm{sl}} - D_n^{\mathrm{sl}}| > \delta) = 0.$$

LEMMA 10.2 (Approximation by a Lebesgue integral—sliding blocks).   *Suppose Condition* 2.1 *is met. Then, as* $n \to \infty$,

$$D_{n,\ell}^{\mathrm{sl}} = D_{n,\ell}^{\prime\mathrm{sl}} + o_{\mathbb{P}}(1) \qquad \text{where } D_{n,\ell}^{\prime\mathrm{sl}} = \int_0^\ell e_n(x)\theta e^{-\theta x}\,\mathrm{d}x.$$

LEMMA 10.3 (Joint convergence of fidis—sliding blocks).   *Let*

$$G_n^{sl} = \sqrt{k_n}\big(T_n^{sl} - \mathrm{E}\, T_n^{sl}\big), \qquad T_n^{sl} = \frac{1}{n - b_n + 1} \sum_{t=1}^{n-b_n+1} Z_{nt}^{sl}.$$

Under Condition 2.1, *for any* $x_1, \ldots, x_m \in [0, \infty)$, *as* $n \to \infty$,

$$\big(e_n(x_1), \ldots, e_n(x_m), G_n^{sl}\big)' \rightsquigarrow \big(e(x_1), \ldots, e(x_m), G^{sl}\big)',$$

the random vector on the right-hand side being $\mathcal{N}_{m+1}(\mathbf{0}, \mathbf{\Sigma}^{sl}(x_1, \ldots, x_m))$-*distributed with*

$$\mathbf{\Sigma}^{sl}(x_1, \ldots, x_m) = \begin{pmatrix} r(x_1, x_1) & \ldots & r(x_1, x_m) & h(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \ldots & r(x_m, x_m) & h(x_m) \\ h(x_1) & \ldots & h(x_m) & \dfrac{2(\log(4) - 1)}{\theta^2} \end{pmatrix},$$

where r and h are defined in Lemma 9.3.

Acknowledgements.

SUPPLEMENTARY MATERIAL

Supplement to: "Weak convergence of a pseudo maximum likelihood estimator for the extremal index" (DOI: 10.1214/17-AOS1621SUPP; .pdf). The supplement contains missing proofs for the results in this paper and additional simulation results.

REFERENCES

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. Wiley, Chichester. MR2108013

BERGHAUS, B. and BÜCHER, A. (2018). Supplement to "Weak convergence of a pseudo maximum likelihood estimator for the extremal index." DOI:10.1214/17-AOS1621SUPP.

BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York. MR0534323

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042

BÜCHER, A. and SEGERS, J. (2015). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. Available at arXiv:1511.07613.

DARSOW, W. F., NGUYEN, B. and OLSEN, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36** 600–642.

DE HAAN, L., RESNICK, S. I., ROOTZÉN, H. and DE VRIES, C. G. (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Process. Appl.* **32** 213–224.

DREES, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301.

DREES, H. (2002). Tail empirical processes under mixing conditions. In *Empirical Process Techniques for Dependent Data* 325–342. Birkhäuser, Boston, MA.

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886

FERRO, C. A. T. and SEGERS, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 545–556.

HSING, T. (1984). Point processes associated with extreme value theory. Ph.D. thesis, Univ. North Carolina at Chapel Hill, ProQuest LLC, Ann Arbor, MI.

HSING, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* **21** 2043–2071.

HSING, T., HÜSLER, J. and LEADBETTER, M. R. (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78** 97–112.

KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248. MR0440724

LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

LEADBETTER, M. R. and ROOTZÉN, H. (1988). Extremal theory for stochastic processes. *Ann. Probab.* **16** 431–478. MR0929071

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

O'BRIEN, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.* **15** 281–291. MR0877604

PERFEKT, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab.* **4** 529–548. MR1272738

RÉMILLARD, B., PAPAGEORGIOU, N. and SOUSTRA, F. (2012). Copula-based semiparametric models for multivariate time series. *J. Multivariate Anal.* **110** 30–42. MR2927508

ROBERT, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.* **37** 271–310.

ROBERT, C. Y., SEGERS, J. and FERRO, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020.

ROOTZÉN, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.* **119** 468–490.

SMITH, R. L. and WEISSMAN, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56** 515–528. MR1278224

SÜVEGES, M. (2007). Likelihood estimation of the extremal index. *Extremes* **10** 41–55. MR2407640

SÜVEGES, M., DAVISON, A. C. et al. (2010). Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.* **4** 203–221.

VERVAAT, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab.* **11** 750–783. MR0544194

WEISSMAN, I. and NOVAK, S. Y. (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inference* **66** 281–288.

FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTR. 150
4470 BOCHUM
GERMANY
E-MAIL: betina.berghaus@rub.de
            axel.buecher@rub.de