SFB 823

Weak convergence of a pseudo maximum likelihood estimator for the extremal index

Betina Berghaus, Axel Bücher

Discussion Paper

Nr. 47/2016

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

BETINA BERGHAUS AND AXEL BÜCHER

Abstract. 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 a disjoint and a 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 is outperforming 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.

Key words. Clusters of extremes, extremal index, stationary time series, mixing coefficients, block maxima.

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 Leadbetter (1983); Leadbetter et al. (1983); O'Brien (1987); Hsing et al. (1988); Leadbetter and Rootzén (1988).

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 Hsing (1993); Smith and Weissman (1994); Weissman and Novak (1998); Ferro and Segers (2003); Süveges (2007); Robert (2009); Robert et al. (2009), among others. The present paper is on a slight modification of a blocks estimator 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.

Date: August 5, 2016.

In many papers on estimating the extremal index, either no asymptotic theory is given (such as in Süveges, 2007; Northrop, 2015), or the asymptotic theory is incomplete in the sense that theory is developed for a non-random threshold sequence, while in practice a random sequence must be used (as, e.g., in Weissman and Novak, 1998; Robert et al., 2009). 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, $X_1, X_2, \dots$ denotes a stationary sequence of real-valued random variables with stationary cumulative distribution function (cdf) 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

$$\lim_{n\to\infty} \mathbb{P}(M_{1:n} \le u_n) = e^{-\theta\tau}. \tag{1.1}$$

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

For simplicity, we assume that F is continuous and define a sequence of standard uniform random variables by $U_s = F(X_s)$. For $x \in (0,1)$, let $u_n = F^{\leftarrow}(x^{1/n})$, where $F^{\leftarrow}$ denotes the generalized inverse of the cdf F. Then $n\bar F(u_n) = n(1 - x^{1/n}) \to -\log(x)$ as $n \to \infty$ and therefore, by (1.1)

$$\mathbb{P}(N_{1:n}^n \le x) = \mathbb{P}(M_{1:n} \le u_n) \to e^{\theta \log x} = x^\theta, \tag{1.2}$$

where $N_{1:n} = F(M_{1:n}) = \max\{U_1, \dots, U_n\}$. In other words, $N_{1:n}^n$ asymptotically follows a beta-distribution with parameters $(\theta, 1)$, inspiring Northrop (2015) to estimate θ by the maximum likelihood estimator for the beta-distribution, based on a sample of estimated block maxima (see below). A slight modification of this estimator can be worked out by considering a further transformation to $Z_{1:n} = n(1 - N_{1:n})$. Equation (1.2) immediately implies that

$$\mathbb{P}(Z_{1:n} \ge x) = \mathbb{P}(N_{1:n}^n \le \{1 - x/n\}^n) \to \exp(-\theta x), \qquad n \to \infty, \tag{1.3}$$

that is, $Z_{1:n}$ asymptotically follows the exponential distribution with parameter θ.

Now suppose that we observe a stretch of length n from the time series $(X_s)_{s\ge1}$. 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 $i = 1, \dots, k_n$, let

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

denote the maximum over the X_s from the ith block. Also, let $N_{ni} = F(M_{ni}) = \max\{U_{(i-1)b_n+1}, \dots, U_{ib_n}\}$ and $Z_{ni} = b_n(1 - N_{ni})$. If b_n is sufficiently large, then, by (1.3), the (unobservable) random variables $Z_{n1}, \dots, Z_{nk}$ 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 suggests to estimate θ by the maximum-likelihood estimator for the Exponential(θ) distribution:

$$\tilde\theta_n = \Big(\frac{1}{k_n}\sum_{i=1}^{k_n} Z_{ni}\Big)^{-1}.$$

Note that $\tilde\theta_n$ should not be considered an estimator, as it is based on the unknown cdf F. Subsequently, we call $\tilde\theta_n$ 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

$$\hat N_{ni} = \hat F_n(M_{ni}) \quad \text{and} \quad \hat Z_{ni} = b_n(1-\hat N_{ni}),$$

where $\hat F_n(x) = n^{-1}\sum_{s=1}^n \mathbf{1}(X_s \le x)$ denotes the empirical cdf of $X_1,\dots,X_n$. We obtain the estimator

$$\hat\theta_n = \hat\theta_n^{\mathrm{dj}} = \Big(\frac{1}{k_n}\sum_{i=1}^{k_n} \hat Z_{ni}\Big)^{-1}, \tag{1.4}$$

which is, up to an error of order $o_{\mathbb{P}}(k_n^{-1/2})$, equal to the estimator $\{-\frac{1}{k_n}\sum_{i=1}^{k_n}\log(\hat N_{ni}^{b_n})\}^{-1}$ considered in Northrop (2015) (where no asymptotic theory is given). While deriving the asymptotic distribution of the oracle $\tilde\theta_n$ may appear tractable (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 $\hat\theta_n$ 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 $\hat\theta_n$. It will turn out that the impact of the rank transformation is non-negligible, resulting in different asymptotic variances of $\hat\theta_n$ and the corresponding oracle $\tilde\theta_n$. We also present asymptotic theory for a modification of $\hat\theta_n$ based on sliding block maxima. The asymptotic expansions derived in this paper also suggest an estimator for the asymptotic variance of $\hat\theta_n$, which is the second main contribution. A third contribution consists of a bias reduction method to improve the finite-sample approximation.

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. Consistency and asymptotic normality is then shown in Section 3. In Section 4, we propose a simple device to reduce the bias of the estimators. Estimators of the asymptotic variance are handled in Section 5. Examples are worked out in detail in Section 6, while finite-sample results and a case study are presented in Section 7 and 8, respectively. The complete proof of the main result for the disjoint blocks estimator is given is Section 9, and additional proofs are postponed to a supplementary material (Appendices A and B).

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

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

In time series extremes, one usually imposes assumptions on the decay of the mixing coefficients between sigma-fields generated by $\{X_i\,\mathbf{1}(X_s > F^{\leftarrow}(1-\varepsilon_n)) : s \le \ell\}$ and $\{X_s\,\mathbf{1}(X_s > F^{\leftarrow}(1-\varepsilon_n)) : s \ge \ell + k\}$, 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 $-\infty \le p < q \le \infty$ and $\varepsilon \in (0,1]$, let $\mathcal{B}^\varepsilon_{p:q}$ denote the sigma algebra generated by $U_s^\varepsilon := U_s \mathbf{1}(U_s > 1-\varepsilon)$ with $s \in \{p,\dots,q\}$ and define, for $\ell \ge 1$,

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

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 $N_{1:n} \le 1 - x/n$; the probability of that event converging to $e^{-\theta x}$ under the assumption of the existence of extremal index θ.

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

$$\begin{aligned}\alpha_{n,\ell}(x_1,\dots,x_m) = \sup\{|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| :&\\ A \in \mathcal{F}^{(x_1,\dots,x_m)}_{1:s,n}, B \in \mathcal{F}^{(x_1,\dots,x_m)}_{s+\ell:n,n}, 1 \le s \le n-\ell\}.&\end{aligned}$$

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

By Theorems 4.1 and 4.2 in Hsing et al. (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 $\{1,\dots,q_n\}$ to a probability distribution π on $\mathbb{N}$, that is,

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

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 et al., 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 $x_1 \ge$

$x_2 \geq 0, x_1 \neq 0$. 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

$$\lim_{n\to\infty} \mathbb{P}(N_n^{(x_1)}(B_n) = i, N_n^{(x_2)}(B_n) = j \mid N_n^{(x_1)}(B_n) > 0) = \pi_2^{(x_2/x_1)}(i,j) \qquad \forall\,(i,j) \in \mathcal{J},$$

where q_n is some $\Delta(u_n(x_1), u_n(x_2))$-separating sequence. In that case, the two-level point process $\boldsymbol{N}_n^{(x_1,x_2)} = (N_n^{(x_1)}, N_n^{(x_2)})$ 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

$$\mathrm{E}[|N_n^{(x_1)}(E) - N_n^{(x_2)}(E)|^{2+\delta}] \leq C'_\ell(x_1 - x_2) \qquad \forall\, \ell \geq x_1 \geq x_2 \geq 0.$$

(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

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

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

$$\ell_n = o(b_n^{2/(2+\delta)}), \quad k_n \alpha_{c_2}(\ell_n) = o(1);$$

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 $n \in \mathbb{N}$,

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

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

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

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

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

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

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

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 et al. (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

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}(N_{m:b_n} > 1 - \tfrac{x}{n} \mid U_1 \geq 1 - \tfrac{x}{n}) = 0 \tag{2.2}$$

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^2y^2 allows for additional flexibility, in that it allows for $O(n^2)$-non-negligible 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 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 estimator $\hat\theta_n^{\mathrm{dj}}$ defined in (1.4), as well as of a variant which is based on sliding blocks and which we will denote by $\hat\theta_n^{\mathrm{sl}}$. We begin by defining the latter estimator.

Divide the sample into $n - b_n + 1$ blocks of length b_n, i.e., 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 estimator for disjoint blocks, we will write $N_{nt}^{\mathrm{sl}} = F(M_{nt}^{\mathrm{sl}})$ and $Z_{nt}^{\mathrm{sl}} = b_n(1 - N_{nt}^{\mathrm{sl}})$ and define their empirical counterparts $\hat{N}_{nt}^{\mathrm{sl}} = \hat{F}_n(M_{nt}^{\mathrm{sl}})$ and $\hat{Z}_{nt}^{\mathrm{sl}} = b_n(1 - \hat{N}_{nt}^{\mathrm{sl}})$, where $\hat{F}_n$ is the empirical cdf of $X_1, \dots, X_n$. Just as for the disjoint blocks estimator, the (pseudo-)observations $\hat{Z}_{nt}^{\mathrm{sl}}$ are approximately exponentially distributed with mean θ^{-1}, which suggests to estimate θ by the reciprocal of their empirical mean, i.e.,

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

Note that no data has to be discarded if b_n is not a divisor of the sample size n. 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.

Theorem 3.1. *Suppose that Condition 2.1 and (2.2) is met. Then*

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

where

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

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 Z_{ni} as an iid 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 3.3 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 Appendix A in the supplement (for the sliding blocks estimator). Explicit calculations are possible for instance for a max-autoregressive process, see Section 6.1, or for the iid case.

Example 3.2. 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

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

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 et al. (2009), respectively. Moreover, 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 (9.9), we then obtain $\pi_2^{(\sigma)} = (1-\sigma)\,\mathbf{1}(i = 1, j = 0) + \sigma\,\mathbf{1}(i = 1, j = 1)$.

Remark 3.3 (Main idea for the proof). Define

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

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

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. Asympotic normality of $\hat{\theta}_n$ may be deduced from the delta method and weak convergence of $\sqrt{k}(\hat{T}_n - \theta^{-1})$. The roadmap to handle the latter is as follows: decompose

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

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 rowwise 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

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

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

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

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

$$\begin{aligned} \sqrt{k_n}(\hat{T}_n - T_n) = \frac{b_n}{\sqrt{k_n}}\sum_{i=1}^{k_n}(N_{ni} - \hat{N}_{ni}) &= \frac{b_n}{n\sqrt{k_n}}\sum_{i=1}^{k_n}\sum_{s=1}^{n}\{N_{ni} - \mathbf{1}(U_s \leq N_{ni})\} \\ &= \frac{1}{k_n^{3/2}}\sum_{i=1}^{k_n}\sum_{s=1}^{n}\{\mathbf{1}(U_s > 1 - Z_{ni}/b_n) - Z_{ni}/b_n\} \\ &= \frac{1}{k_n}\sum_{i=1}^{k_n} e_n(Z_{ni}) = \int_0^{\max_{i=1}^{k_n} Z_{ni}} e_n(x)\,\mathrm{d}\hat{H}_{k_n}(x). \end{aligned} \tag{3.6}$$

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; Rootzén, 2009; Robert, 2009; Drees and Rootzén, 2010), 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.

Proof of Theorem 3.1 (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

$$D_n = \int_0^{\max Z_{ni}} e_n(x)\,\mathrm{d}\hat H_{k_n}(x), \quad D_{n,\ell} = \int_0^{\ell} e_n(x)\,\mathrm{d}\hat H_{k_n}(x), \quad D_\ell = \int_0^{\ell} e(x)\theta e^{-\theta x}\,\mathrm{d}x.$$

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. 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 $\ell \to \infty$. □

Proof of Theorem 3.1 (Sliding blocks). Let $\hat H_{k_n}^{\mathrm{sl}}$ denote the empirical distribution function of the Z_{nt}^{sl}, i.e., $\hat H_{k_n}^{\mathrm{sl}}(x) = \frac{1}{n-b_n+1}\sum_{t=1}^{n-b_n+1}\mathbf{1}(Z_{nt}^{\mathrm{sl}} \le x)$ and let

$$D_n^{\mathrm{sl}} = \int_0^{\max_t Z_{nt}^{\mathrm{sl}}} e_n(x)\,\mathrm{d}\hat H_{k_n}^{\mathrm{sl}}(x), \quad D_{n,\ell}^{\mathrm{sl}} = \int_0^{\ell} e_n(x)\,\mathrm{d}\hat H_{k_n}^{\mathrm{sl}}(x), \quad D_\ell^{\mathrm{sl}} = \int_0^{\ell} e(x)\theta e^{-\theta x}\,\mathrm{d}x.$$

With this notation the proof follows along the same lines as for the disjoint blocks, with Lemma 9.1, 9.2 and 9.3 replaced by Lemma A.1, A.2 and A.3, respectively. □

4. Bias reduction

Throughout this section, let $(\hat T_n, T_n) \in \{(\hat T_n^{\mathrm{dj}}, T_n^{\mathrm{dj}}), (\hat T_n^{\mathrm{sl}}, T_n^{\mathrm{sl}})\}$ denote any of the quantities defined in (3.1) or (3.2). A Taylor expansion allows to approximately decompose the bias of the estimator $\hat\theta_n = \hat T_n^{-1}$ into two parts:

$$\mu_n = \mathrm{E}[\hat T_n^{-1} - \theta] \approx -\theta^2\,\mathrm{E}[\hat T_n - \theta^{-1}] = -\theta^2\,\mathrm{E}[\hat T_n - T_n] - \theta^2\,\mathrm{E}[T_n - \theta^{-1}] =: \mu_{n1} + \mu_{n2}.$$

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 et al., 2009, Section 6). The first component μ_{n1} 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 4.1. *Additionally to the conditions of Theorem 3.1 suppose that Condition 2.1(iii) is met with $c_2 = 1$. Then*

$$\lim_{n\to\infty} k_n \operatorname{E}[\hat T_n - T_n] = -\frac{1}{\theta}.$$

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).

A proof can be found in Section B. As a consequence, we obtain that $\mu_{n1} = k_n^{-1}\theta + o(k_n^{-1})$. Plugging-in $\hat\theta_n$ as a consistent estimator for θ, we can estimate μ_{n1} by $\hat\mu_{n1} = k_n^{-1}\hat\theta_n$ and subtract it from $\hat\theta_n$ to obtain the bias-reduced estimator

$$\hat\theta_{n,bc} = \tfrac{k_n-1}{k_n}\hat\theta_n.$$

Note that if we are additionally willing to assume that $k_n \operatorname{E}[T_n - \theta^{-1}] = k_n \operatorname{E}[Z_{1:b_n} - \theta^{-1}] = o(1)$ as $n\to\infty$ (cf. Condition 2.1(vii)), we obtain that μ_{n1} is in fact the dominating bias-component. In common models, the assumption $k_n \operatorname{E}[T_n - \theta^{-1}] = o(1)$ is satisfied as soon as $k_n/b_n = o(1)$ (see Section 6). In comparison to the assumption $k_n/b_n^2 = o(1)$ imposed in Condition 2.1(iii), this requires larger block sizes. Similar assumptions have also been made for the bias-reduced estimators in Robert et al. (2009).

5. Variance estimation

For statistical inference on θ, estimators for the asymptotic variance formulas in Theorem 3.1 are needed. Unfortunately, the formulas itself 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. As explained below, this is particularly useful since a straightforward extension of our proposed estimator for σ_{dj}^2 to the sliding blocks estimator would require the choice of an additional tuning parameter.

By the central decomposition in (3.3) and the calculations in (3.6), we can write $\mathbb{T}_n^{\mathrm{dj}} = \sqrt{k_n}(\hat T_n^{\mathrm{dj}} - \theta^{-1})$ as

$$\begin{aligned}
\mathbb{T}_n^{\mathrm{dj}} &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} Z_{ni} - \theta^{-1} + \hat Z_{ni} - Z_{ni} \\
&= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \Big\{ Z_{ni} - \theta^{-1} + \tfrac{1}{k_n}\textstyle\sum_{j=1}^{k_n}\sum_{s\in I_j}\{\mathbf{1}(U_s > 1 - \tfrac{Z_{ni}}{b_n}) - \tfrac{Z_{ni}}{b_n}\}\Big\} \\
&= \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} B_{nj},
\end{aligned}$$

where

$$B_{nj} = Z_{nj} - \theta^{-1} + \textstyle\sum_{s\in I_j}\tfrac{1}{k_n}\sum_{i=1}^{k_n}\{\mathbf{1}(U_s > 1 - \tfrac{Z_{ni}}{b_n}) - \tfrac{Z_{ni}}{b_n}\},$$

and where $I_j = \{(j-1)k_n + 1, \dots, jk_n\}$ denotes the jth block of indices. The proof of Theorem 3.1 shows that $B_{n1}, \dots, B_{nk_n}$ are asymptotically independent and centered, and that their empirical mean multiplied by $\sqrt{k_n}$ converges to a centered normal distribution with variance σ_{dj}^2. Hence, their second empirical moment should be a consistent estimator for σ^2_{dj}. As the sample $B_{n1}, \dots, B_{nk_n}$ depends on unknown quantities, we must replace these objects by empirical counterparts, leading us to define

$$\hat B_{nj} = \hat Z_{nj} - \hat T_n + \sum_{s\in I_j} \frac{1}{k_n} \sum_{i=1}^{k_n} \{\mathbf{1}(\hat U_s > 1 - \tfrac{\hat Z_{ni}}{b_n}) - \tfrac{\hat Z_{ni}}{b_n}\}$$
$$= \hat Z_{nj} + \sum_{s\in I_j} \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbf{1}(\hat U_s > 1 - \tfrac{\hat Z_{ni}}{b_n}) - 2\cdot \hat T_n^{\mathrm{dj}},$$

where $\hat U_s = \hat F_n(X_s)$. The following proposition shows that

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

are in fact consistent estimators for σ^2_{dj} and σ^2_{sl}, 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.

Proposition 5.1 (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 $n\to\infty$,*

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

The proof is given in Section 9, while the finite sample performance is investigated in Section 7.

Following the above route to derive an estimator for the variance of the sliding blocks version is substantially more complicated. The corresponding decomposition of $\mathbb{T}_n^{\mathrm{sl}} = \sqrt{k_n}(\hat T_n^{\mathrm{sl}} - \theta^{-1})$ is

$$\mathbb{T}_n^{\mathrm{sl}} = \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n-b_n+1} \Big\{ \tfrac{n}{n-b_n+1} \tfrac{Z_{nt}^{\mathrm{sl}} - \theta^{-1}}{b_n} + \tfrac{1}{n-b_n+1} \sum_{s=1}^{n} \{\mathbf{1}(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\} \Big\}$$
$$= \frac{1}{\sqrt{k_n}} \sum_{s=1}^{n-b_n} \Big\{ \tfrac{Z_{ns}^{\mathrm{sl}} - \theta^{-1}}{b_n} + \tfrac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} \{\mathbf{1}(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\} \Big\} + o_{\mathbb{P}}(1)$$
$$= \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n-1} B_{nj}^{\mathrm{sl}} + o_{\mathbb{P}}(1),$$

where

$$B_{nj}^{\mathrm{sl}} = \tfrac{1}{b_n} \sum_{s\in I_j} (Z_{ns}^{\mathrm{sl}} - \theta^{-1}) + \sum_{s\in I_j} \tfrac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} \{\mathbf{1}(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\},$$

and where the $o_{\mathbb{P}}$ terms are due to omitting final blocks and due to $n/(n - b_n + 1) = 1 + o(1)$. Unlike for the disjoint blocks estimator, the B_{nj}^{sl} are not asymptotically independent. It can be seen from the proof, see in particular Lemma A.3, that a further 'blocking of blocks' is necessary to obtain asymptotically independent random variables. Precisely, let $k_n^* < k_n$ be an integer sequence converging to infinity, which formally should satisfy $k_n^* = o(k_n^{\delta/\{2(1+\delta)\}})$ with δ from Condition 2.1(ii). Form blocks of length k_n^* from the B_{nj}^{sl}, that is, let

$$A_{nl}^{\mathrm{sl}} = \textstyle\sum_{j=(l-1)k_n^*+1}^{lk_n^*} B_{nj}^{\mathrm{sl}}, \qquad l = 1, \dots, q_n = \lfloor k_n/(k_n^*) \rfloor.$$

Up to an incomplete final block (which we can absorb into the $o_{\mathbb{P}}(1)$), we can then write $\mathbb{T}_n^{\mathrm{sl}} = \frac{1}{\sqrt{k_n}} \sum_{l=1}^{q_n} A_{nl}^{\mathrm{sl}} + o_{\mathbb{P}}(1)$. Motivated by the proof, the A_{nl}^{sl} can now be regarded as asymptotically independent, which suggests to estimate

$$\tilde\sigma_{\mathrm{sl}}^2 = \frac{1}{k_n} \sum_{l=1}^{q_n} (\hat A_{nl}^{\mathrm{sl}})^2, \qquad \hat A_{nl}^{\mathrm{sl}} = \textstyle\sum_{j=(l-1)q_n+1}^{lq_n} \hat B_{nj}^{\mathrm{sl}},$$

where

$$\hat B_{nj}^{\mathrm{sl}} = \tfrac{1}{b_n} \textstyle\sum_{s \in I_j} (\hat Z_{ns}^{\mathrm{sl}} - \hat T_n^{\mathrm{sl}}) + \sum_{s \in I_j} \frac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} \{ \mathbf{1}(\hat U_s > 1 - \frac{\hat Z_{nt}^{\mathrm{sl}}}{b_n}) - \frac{\hat Z_{nt}^{\mathrm{sl}}}{b_n} \}.$$

In comparison to $\hat\sigma_{\mathrm{sl}}^2$, this estimator requires the choice of an additional tuning parameter k_n^*. We therefore do not pursue it any further in this paper.

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 3.1 are possible. These allow for a theoretical comparison with the blocks estimators from Robert (2009) and Robert et al. (2009). Moreover, we show that all assumptions imposed in Condition 2.1 are satisfied. In Section 6.2, we consider solutions of stochastic difference equations such as ARCH-processes. Complementing results from (Robert, 2009, Example 3.1) 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

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

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, i.e., $\pi(j) = \alpha^{j-1}(1-\alpha)$ for $j \ge 1$ (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

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

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

$$\pi_2^{(\sigma)}(j_1, 0) = \pi(j_1) - \sum_{j_2=1}^{j_1} \pi_2^{(\sigma)}(j_1, j_2) = (1-\alpha)\alpha^{j_1-1}\, \mathbf{1}(j_1 \leq z) + (\alpha^z - \sigma)\, \mathbf{1}(j_1 = 1 + z)$$

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

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

and that

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

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

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

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 et al., 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 et al. (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 et al. (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 5.1). It can be further be shown that, with $U_s = \exp(-1/X_s)$, we have

$$\mathrm{Var}\left\{\sum_{s=1}^n \mathbf{1}(U_s > 1 - y)\right\} \leq ny\{1 + 2\alpha/(1-\alpha)\}$$

for all $y \in (0,1)$, that is, Condition 2.1(iv) is met. Moreover, a simple calculation shows that $\mathbb{P}(\min_{i=1}^{2k_n} N'_{ni} \leq c) \leq 2k_n \mathbb{P}(N'_{n1} \leq c) = O(k_n c^{(1-\alpha)b_n/2}) = o(1)$, 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 $\delta > 0$,

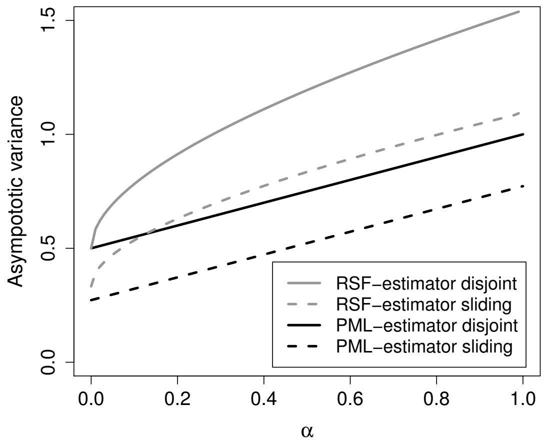

FIGURE 1. Asymptotic variances of $\sqrt{k_n}(\hat\theta_n/\theta - 1)$ within the ARMAX(α)-Model for the sliding and disjoint blocks estimators analyzed in this paper (PML) and in Robert et al. (2009) (RSF).

and that $\mathrm{E}[Z_{1:b_n}] - \theta^{-1} = O(b_n^{-1})$. The latter implies that Condition 2.1(vii) is satisfied if $k_n = o(b_n^2)$, i.e. if (2.1) holds. Finally, it can easily be seen that (2.2) is met.

6.2. **Stochastic Difference Equations.** Consider the stochastic difference equation

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

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

$$\mathrm{E}\log A_1 < 0, \quad \mathrm{E}[A_1^\kappa] = 1, \quad \mathrm{E}[A_1^\kappa \max(\log A_1, 0)] < \infty, \quad \mathrm{E}[B_1^\kappa] \in (0,\infty).$$

Under this assumption, there exists a unique stationary solution of (6.1) and the cdf 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 3.1.

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 Section B in the supplement.

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 three 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 (see Table 3.1 in de Haan et al., 1989).
- The **Markovian Copula-model** (Darsow et al., 1992):
$$X_s = F^{-}(U_s), \quad (U_s, U_{s-1}) \sim C_\vartheta, \qquad s \in \mathbb{Z}.$$
Here, F^{-} is the quantile function of some arbitrary continuous cdf 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, $\theta = \mathbb{P}(\max_{t\ge1} \prod_{s=1}^{t} A_s \le U)$, where $U, A_1, A_2, \dots$ are independent, U is standard uniform and A_s has cdf $H_\vartheta(s) = 1 - (1 + s^\vartheta)^{-(1+1/\vartheta)}$, $s \ge 0$, 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 cdf (the results are independent of this choice, as the estimators are rank-based). Algorithm 2 in Rémillard et al. (2012) allows 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 $n = 8,192 = 2^{13}$ and the block size b_n for the blocks estimators is chosen from the set $2^2, 2^3, \dots, 2^9$.

7.1. **Comparison with other estimators for the extremal index.**

We present results for five different estimators: the bias-reduced sliding blocks estimator from this paper, the bias-reduced sliding blocks estimator from Robert et al. (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 $\phi = 1.3$.

θ	PML-sliding	RSF-sliding	Intervals	ML-Süveges	Robert
0.25	0.91	1.35	0.53	**0.22**	1.77
0.50	1.58	2.24	0.99	**0.63**	2.07
0.75	2.03	2.34	1.17	**0.96**	2.31
1.00	1.78	0.12	0.88	**0.11**	2.22
0.422	3.18	4.85	**2.53**	3.19	4.00
0.460	3.53	5.45	**2.71**	1.92	4.26
0.727	**1.07**	1.46	1.08	1.44	1.19
0.997	**0.50**	1.33	5.34	2.19	0.65
0.95	1.26	1.99	9.53	4.19	**1.08**
0.80	0.79	**0.72**	5.75	2.33	1.16
0.60	1.71	2.75	0.60	**0.37**	2.33
0.40	3.14	5.23	**2.86**	3.68	4.74
0.20	**2.74**	5.28	2.81	14.03	4.59

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

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.

In Figures 2–4 we depict the mean-squared error $\mathrm{E}[(\hat\theta - \theta)^2]$ as a function of the block size parameter b_n, estimated on the basis of $N = 10,000$ simulation runs. For almost all models and estimators, the MSE-curves are U-shaped, representing the usual bias-variance tradeoff in extreme value theory. The minimal values of these curves are of particular interest, and are summarized in Table 1. We observe that the sliding blocks PML-estimator outperforms the other two blocks estimators in most scenarios. For the ARMAX-model, this is in agreement with the theoretical findings presented in Figure 1. In general however, there is no clear best estimator in terms of the MSE. For the ARMAX-model, the Süveges-estimator performs best, followed by the intervals estimator. For the ARCH-model (which may be regarded as the more relevant model, given its frequent use for the modeling of financial data) the picture is different: for small values of θ, the intervals estimator performs best, while for larger values the PML-estimator is the winner. For the Markovian copula model, each of the estimators under consideration performs best for one particular choice of the parameter. The sliding blocks PML-estimator is generally the most robust one, none of the reported MSE-values exceeding a value of $3.6 \cdot 10^{-3}$.

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

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

as estimators for the variances of $\sqrt{k_n}\hat\theta_n^{\mathrm{dj}}$ and $\sqrt{k_n}\hat\theta_n^{\mathrm{sl}}$, respectively. Results can be found in Figure 5, where we depict the curves

$$b_n \mapsto \mathrm{E}\Big[\Big(\frac{\hat\tau^2}{\mathrm{Var}(\sqrt{k_n}\hat\theta_n)} - 1\Big)^2\Big] = \mathrm{E}\Big[\Big(\frac{\hat\tau^2(b_n)}{\mathrm{Var}(\sqrt{k_n}\hat\theta_n(b_n))} - 1\Big)^2\Big]$$

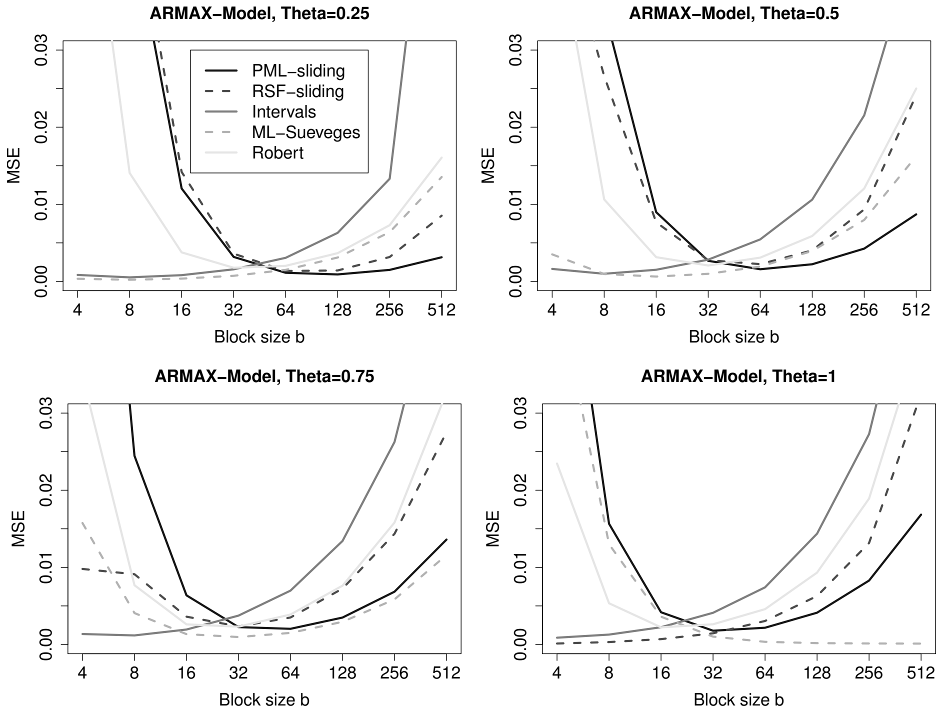

FIGURE 2. Mean squared error for the estimation of θ within the ARMAX-model for four values of $\theta \in \{0.25, 0.5, 0.75, 1\}$.

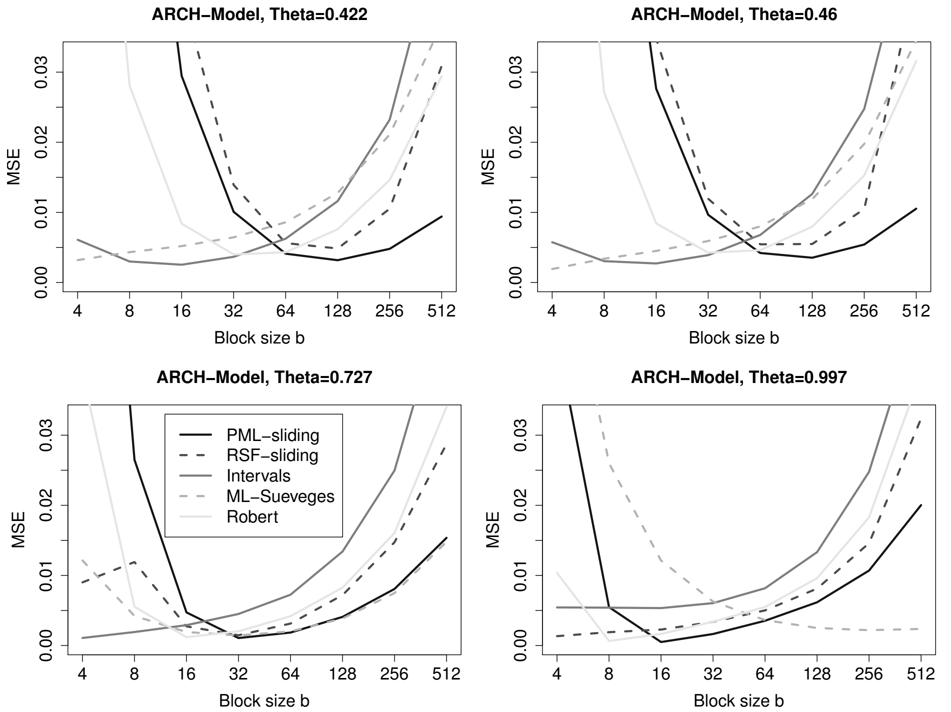

FIGURE 3. Mean squared error for the estimation of θ within the ARCH-model for four values of $\theta \in \{0.422, 0.460, 0.727, 0.997\}$.

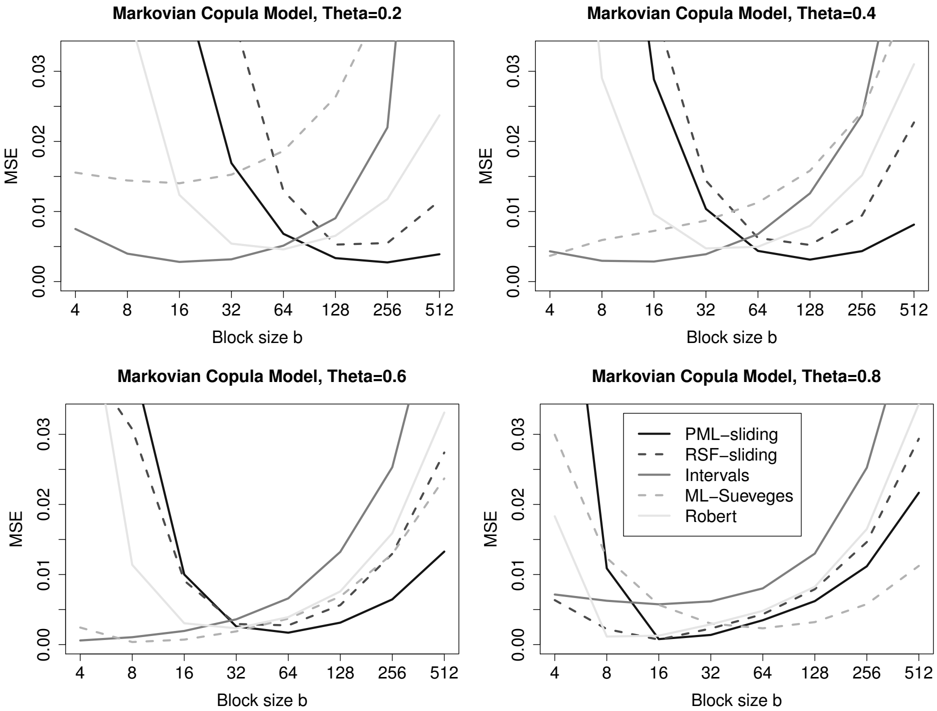

FIGURE 4. Mean squared error for the estimation of θ within the Markovian copula model for four values of $\theta \in \{0.2, 0.4, 0.6, 0.8\}$.

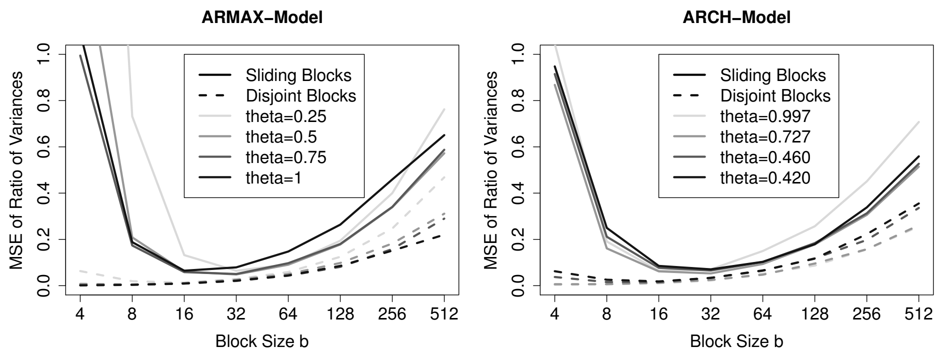

FIGURE 5. Mean squared error $\mathrm{E}[(\hat\tau^2/\operatorname{Var}(\hat\theta_n)-1)^2]$ within the ARMAX-model (left) and the squared ARCH-model (right).

estimated on the basis of 10,000 simulation runs. Here, $(\hat\tau^2, \hat\theta_n) \in \{(\hat\tau^2_{\mathrm{dj}}, \hat\theta_n^{\mathrm{dj}}), (\hat\tau^2_{\mathrm{sl}}, \hat\theta_n^{\mathrm{sl}})\}$ and $\operatorname{Var}(\sqrt{k_n}\hat\theta_n(b_n))$ 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 behavior as for the estimation of θ depicted in Figures 2–4: the curves are U-shaped and possess a minimum at some intermediate values of b_n. Due to the fact that estimator $\hat\tau^2_{\mathrm{sl}}$ is based on an additional

		ARMAX-model				Squared ARCH-model			
	b_n/θ	0.25	0.5	0.75	1	0.422	0.46	0.727	0.997
disjoint	16	0	0	0.13	0.50	0	0	0.27	**0.95**
	32	0.03	0.63	0.85	**0.91**	0.07	0.10	**0.94**	0.88
	64	0.79	**0.93**	**0.94**	**0.95**	0.76	0.79	**0.92**	0.87
	128	**0.94**	**0.94**	**0.94**	**0.93**	**0.93**	**0.93**	**0.90**	0.87
	256	**0.94**	**0.92**	**0.92**	**0.91**	**0.93**	**0.92**	0.87	0.86
	512	**0.91**	0.89	0.88	0.85	**0.90**	**0.90**	0.85	0.82
sliding	16	0	0	0.02	0.18	0	0	0.09	**0.92**
	32	0	0.46	0.76	0.85	0.02	0.04	**0.92**	0.81
	64	0.69	**0.90**	**0.93**	**0.92**	0.67	0.70	**0.90**	0.79
	128	**0.92**	**0.93**	**0.92**	0.88	**0.90**	**0.91**	0.87	0.78
	256	**0.92**	**0.90**	0.87	0.83	**0.91**	**0.90**	0.83	0.75
	512	0.87	0.84	0.81	0.73	0.86	0.86	0.79	0.70

TABLE 2. Empirical coverage probabilities of 95%-confidence bands. Values above 90% are in boldface.

estimation step (which is potentially biased, if b_n is small), the approximation works better for the disjoint blocks estimator.

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

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

for θ, where $(\hat{\tau}^2, \hat{\theta}_n) \in \{(\hat{\tau}^2_{\mathrm{dj}}, \hat{\theta}_n^{\mathrm{dj}}), (\hat{\tau}^2_{\mathrm{sl}}, \hat{\theta}_n^{\mathrm{sl}})\}$ and 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 Table 2, with coverage probabilities above 0.9 in boldface. It can be seen that the probabilities strongly depend on the block size b_n, with at least one reasonable choice for every model, usually close to the MSE-minimal choice in Figures 2 and 3. The larger width of the confidence sets for the disjoint blocks estimator (not presented here; it is 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 ($n = 6,780$ observations for each index). Clusters of large negative returns can be financially damaging and are hence of interest for risk management.

In Figure 6, 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, alongside with a 95%-confidence band based on the variance estimator from Section 5 and the normal approximation. Interestingly, the curves appear to be quite smooth. 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 unique estimate for the extremal index. The dotted lines in Figure 6 correspond to the following ad-hoc solution (which is

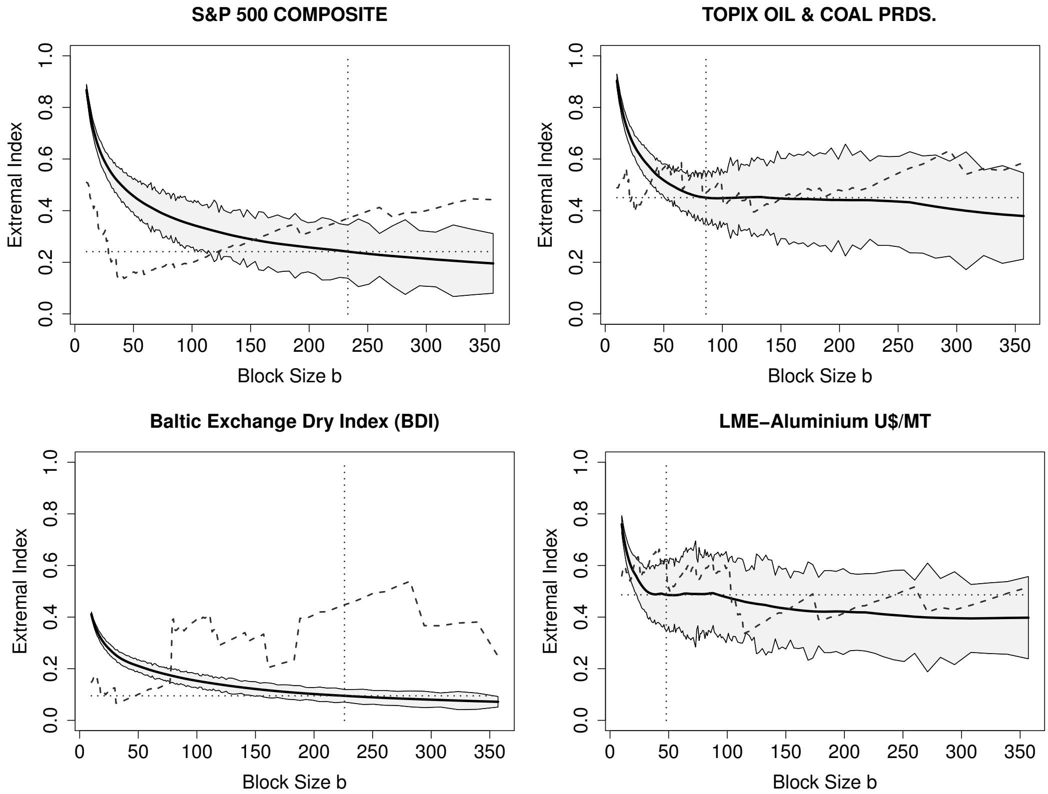

FIGURE 6. Extremal index estimates for four financial time series as a function of the block size. The solid line is the bias–reduced sliding blocks estimate, the shaded region is the pointwise 95%-confidence band. The dashed line is the intervals estimator. The dotted lines correspond to the plateau-search algorithm described in the main text.

essentially searching for a plateau in the plots): first, calculate absolute differences $d(b) = |\hat\theta(b+1) - \hat\theta(b)|, b = 10, \dots, 356$. Let D denote the empirical mean of $d(10), \dots, d(356)$. The chosen block length $\hat b$ is the minimal block length such that the sum over five successive values, $d(\hat b) + \cdots + d(\hat b + 4)$, is smaller than $D/2$. It can be seen that this choice approximately catches the plateaus visible in the 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 3, we state the resulting estimates of the extremal index and the width of the corresponding confidence intervals. Interestingly, the extremal index lies around 0.3 for most of the equity indexes (S&P 500 Composite, MSCI World, etc.), while it is 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

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

TABLE 3. 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.

production. In particular, this index is free of speculation which might explain why the extremal index is much smaller than for the other time series under consideration.

9. PROOFS

Throughout the proofs, C and C' denote generic constants whose values may change from line to line. The notation $o, o_{\mathbb{P}}, O, O_{\mathbb{P}}$ always refers to $n \to \infty$, if not mentioned otherwise.

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

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

Proof. For some $\varepsilon \in (0, c_1 \wedge c_2)$, let $A_n = A_n(\varepsilon)$ denote the event $\{\min_{i=1}^{k_n} N_{ni} > 1 - \varepsilon/2\} = \{\max_{i=1}^{k_n} Z_{ni} < \varepsilon b_n/2\}$. By Condition 2.1(v), we have $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. We may write

$$D_n - D_{n,\ell} = R_{n,\ell}\, \mathbf{1}_{A_n} + o_{\mathbb{P}}(1)$$

as $n \to \infty$, where, with $I_j = \{(j-1)b_n + 1, \dots, jb_n\}$ for $j = 1, \dots, k_n$ (and $I_j = \varnothing$ else),

$$R_{n,\ell} = k_n^{-3/2} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} \sum_{s \in I_j} f(U_s, Z_{ni}) g_{n,\ell}(Z_{ni})$$

and

$$f(U_s, Z_{ni}) = \mathbf{1}(U_s > 1 - \tfrac{Z_{ni}}{b_n}) - \tfrac{Z_{ni}}{b_n}, \qquad g_{n,\ell}(Z_{ni}) = \mathbf{1}(b_n\varepsilon/2 > Z_{ni} \geq \ell).$$

Now, decompose $R_{n,\ell} = R_{n,\ell,0} + R_{n,\ell,1} + R_{n,\ell,-1} + R_{n,\ell,2}$ according to whether the second sum over j is such that $j = i, j = i+1, j = i-1$ or $|j-i| \geq 2$, respectively. It suffices to show that $R_{n,\ell,0}\,\mathbf{1}_{A_n} = o_{\mathbb{P}}(1)$ and $R_{n,\ell,\pm 1}\,\mathbf{1}_{A_n} = o_{\mathbb{P}}(1)$ as $n \to \infty$, and that

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{P}(|R_{n,\ell,2}\,\mathbf{1}_{A_n}| > \delta) = 0 \tag{9.1}$$

for all $\delta > 0$.

First, since $R_{n,\ell,0} = k_n^{-3/2} \sum_{i=1}^{k_n} Z_{ni} \cdot g_{n,\ell}(Z_{ni})$, we have $\mathrm{E}\,|R_{n,\ell,0}| \leq k_n^{-1/2}\,\mathrm{E}\,|Z_{ni}| = o(1)$ as $n \to \infty$ by Condition 2.1(vi).

Second, we can write $R_{n,\ell,1} = \bar{R}_{n,\ell,1} - R_{n,\ell,0} = \bar{R}_{n,\ell,1} - o_{\mathbb{P}}(1)$, where

$$\bar{R}_{n,\ell,1} = k_n^{-3/2} \sum_{i=1}^{k_n-1} \sum_{s \in I_{i+1}} \mathbf{1}(U_s > 1 - \tfrac{Z_{ni}}{b_n}) g_{n,\ell}(Z_{ni})$$

whence it suffices to show that $\bar{R}_{n,\ell,1}\,\mathbf{1}_{A_n} = o_{\mathbb{P}}(1)$. For that purpose, define

$$U_s^{\varepsilon} = U_s\,\mathbf{1}(U_s > 1-\varepsilon), \qquad Z_{ni}^{\varepsilon/2} = b_n(1 - N_{ni}^{\varepsilon/2}) = b_n(1 - \max_{s \in I_i} U_s^{\varepsilon/2}). \tag{9.2}$$

Note that $Z_{ni}^{\varepsilon/2}$ is $\mathcal{B}^{\varepsilon/2}_{\{(i-1)b_n+1\}:ib_n}$ measurable, whence the mixing coefficients become available. On the event A_n, we have $\bar{R}_{n,\ell,1} = \bar{R}^{\varepsilon}_{n,\ell,1}$, where $\bar{R}^{\varepsilon}_{n,\ell,1}$ is defined exactly as $\bar{R}_{n,\ell,1}$, but with U_s and Z_{ni} replaced by U_s^{ε} and $Z_{ni}^{\varepsilon/2}$, respectively. By stationarity, we obtain

$$\mathrm{E}\,|\bar{R}^{\varepsilon}_{n,\ell,1}| = (k_n - 1)k_n^{-3/2} \sum_{s=1}^{b_n} \mathrm{E}\Big[\mathbf{1}\big(U^{\varepsilon}_{b_n+s} > 1 - \tfrac{Z_{n1}^{\varepsilon/2}}{b_n}\big) g_{n,\ell}(Z_{n1}^{\varepsilon/2})\Big].$$

Recall Theorem 3 in Bradley (1983) (coupling for strongly mixing random variables): if X and Y are two random variables in some Borel space S and $\mathbb{R}$, respectively, if U is uniform on $[0,1]$ and independent of (X,Y) and if $q > 0$ and $\gamma > 0$ are such that $q \leq \|Y\|_\gamma = (\mathrm{E}\,|Y|^\gamma)^{1/\gamma}$, then there exists measurable function f such that $Y^* = f(X,Y,U)$ has the same distribution as Y, is independent of X and satisfies

$$\mathbb{P}(|Y - Y^*| \geq q) \leq 18(\|Y\|_\gamma/q)^{\gamma/(2\gamma+1)} \alpha(\sigma(X), \sigma(Y))^{2\gamma/(2\gamma+1)}. \tag{9.3}$$

Apply this theorem with $X = U^{\varepsilon}_{b_n+s}$, $Y = Z_{n1}^{\varepsilon/2}$, $\gamma = 2+\delta$ and $q = q_n = \|Z_{n1}^{\varepsilon/2}\|_{2+\delta}$ to obtain that

$$\mathrm{E}\,|\bar{R}^{\varepsilon}_{n,\ell,1}| \leq k_n^{-1/2} \sum_{s=1}^{b_n} \Big\{ \mathrm{E}[\mathbf{1}(U^{\varepsilon}_{b_n+s} > 1 - \tfrac{Z_{n1}^{\varepsilon/2*} + q_n}{b_n})] + 18 \cdot \alpha(\sigma(U^{\varepsilon}_{b_n+s}), \sigma(Z_{n1}^{\varepsilon/2}))^{\frac{4+2\delta}{5+2\delta}} \Big\}$$

where $Z_{n1}^{\varepsilon/2*}$ is independent of X and has the same distribution as $Z_{n1}^{\varepsilon/2}$. Note that $\alpha(\sigma(U^{\varepsilon}_{b_n+s}), \sigma(Z_{n1}^{\varepsilon/2})) \leq \alpha_{c_2}(s)$. Since $U_s^{\varepsilon} \leq U_s$, it follows that

$$\mathrm{E}\,|\bar{R}^{\varepsilon}_{n,\ell,1}| \leq k_n^{-1/2} \Big\{ \mathrm{E}[Z_{n1}^{\varepsilon/2*}] + q_n + 18 \times \sum_{s=1}^{b_n} \alpha_{c_2}(s)^{\frac{4+2\delta}{5+2\delta}} \Big\},$$

which converges to 0 by Conditions 2.1(iii) and (vi). To conclude, $R_{n,\ell,1}\,\mathbf{1}_{A_n} = o_{\mathbb{P}}(1)$.

The sum $R_{n,\ell,-1}$ can be treated analogously so that it remains to show (9.1). Decompose $R_{n,\ell,2} = \bar{S}_{n,\ell,1} + \bar{S}_{n,\ell,2}$ where

$$\bar{S}_{n,\ell,1} = k_n^{-3/2} \sum_{i=3}^{k_n} \sum_{j=1}^{i-2} \sum_{s \in I_j} f(U_s, Z_{ni}) g_{n,\ell}(Z_{ni})$$

and where $\bar{S}_{n,\ell,2}$ is defined analogously with the second sum ranging from $i+2$ to k_n. We will only treat $\bar{S}_{n,\ell,1}$ in the following, as $\bar{S}_{n,\ell,2}$ can be treated analogously. Recall (9.2) and note that, on the event A_n, we have $f(U_s, Z_{ni}) g_{n,\ell}(Z_{ni}) = f(U_s^\varepsilon, Z_{ni}^{\varepsilon/2}) g_{n,\ell}(Z_{ni}^{\varepsilon/2})$. Therefore, again on the event A_n,

$$\bar{S}_{n,\ell,1} = k_n^{-3/2} \sum_{i=3}^{k_n} \sum_{j=1}^{i-2} \sum_{s \in I_j} f(U_s^\varepsilon, Z_{ni}^{\varepsilon/2}) g_{n,\ell}(Z_{ni}^{\varepsilon/2}) = \frac{1}{k_n} \sum_{i=3}^{k_n} e_{1:i-2}(Z_{ni}^{\varepsilon/2}) g_{n,\ell}(Z_{ni}^{\varepsilon/2}) =: \bar{S}_{n,\ell,1}^\varepsilon,$$

where, for $p, q \in \{1, \dots, k_n\}$, $p < q$, and $x \geq 0$,

$$e_{n,p:q}(x) = \frac{1}{\sqrt{k_n}} \sum_{i=p}^{q} \sum_{s \in I_i} \{\mathbf{1}(U_s^\varepsilon > 1 - x/b_n) - x/b_n\}.$$

We will show that (9.1) is met with $R_{n,\ell,2}\,\mathbf{1}_{A_n}$ replaced by $\bar{S}_{n,\ell,1}^\varepsilon$, and for that purpose we consider the first central moment of $\bar{S}_{n,\ell,1}^\varepsilon$.

Note that $|e_{1:j}(x)\,\mathbf{1}(x \geq \ell)| \leq j b_n/\sqrt{k_n}$ and that, for all $x, y \geq 0$ with $y - q \leq x \leq y + q$ for some $q > 0$, we have

$$|e_{1:j}(x)| \leq |e_{1:j}(y+q)| \vee |e_{1:j}((y-q) \vee 0)| + 2q\sqrt{k_n},$$

as can be shown by a case-by-case study and monotonicity arguments. The previous two inequalities, together with (9.3) with $X = (U_1^\varepsilon, \dots, U_{(i-1)b_n}^\varepsilon)$, $Y = Z_{ni}^{\varepsilon/2}$, $\gamma = 2 + \delta$ and $q = q_n = \|Z_{n1}^{\varepsilon/2}\|_{2+\delta}/\sqrt{k_n}$, imply that

$$\begin{aligned} \mathrm{E}[|\bar{S}_{n,\ell,1}^\varepsilon|] \leq \frac{1}{k_n} \sum_{i=3}^{k_n} \mathrm{E}\Big[\big\{ |e_{1:i-2}(Z_{ni}^{\varepsilon/2*} + q_n)| + |e_{1:i-2}((Z_{ni}^{\varepsilon/2*} - q_n) \vee 0)| + 2\|Z_{n1}^{\varepsilon/2}\|_{2+\delta} \big\} \\ \times \mathbf{1}\big(\tfrac{b_n\varepsilon}{2} + q_n > Z_{ni}^{\varepsilon/2*} \geq \ell - q_n\big) \Big] + \frac{1}{k_n} 18 \big(\sqrt{k_n}\big)^{\frac{2+\delta}{5+2\delta}} \sum_{i=3}^{k_n} \frac{i b_n}{\sqrt{k_n}} \alpha_\varepsilon(b_n)^{\frac{4+2\delta}{5+2\delta}}, \end{aligned}$$

where $Z_{ni}^{\varepsilon/2*}$ is independent of $(U_1^\varepsilon, \dots, U_{(i-1)b_n}^\varepsilon)$ and has the same distribution as $Z_{ni}^{\varepsilon/2}$. The second sum on the right-hand side is of the order (note that $\eta > 3$)

$$O\big(b_n k_n^{1/2 + \frac{2+\delta}{10+4\delta}} \alpha_{c_2}(b_n)^{\frac{4+2\delta}{5+2\delta}}\big) = O\big(k_n^{\frac{7+3\delta}{10+4\delta}} b_n^{1 - \eta\frac{4+2\delta}{5+2\delta}}\big) = O\big((k_n/b_n^2)^{\frac{7+3\delta}{10+4\delta}} b_n^{-\frac{\delta}{5+2\delta}}\big)$$

which converges to 0 by Condition (2.1).

Since $\|Z_{n1}^{\varepsilon/2}\|_{2+\delta} \mathbb{P}(Z_{n1}^{\varepsilon/2} \geq \ell - q_n)$ converges to 0 for $n \to \infty$ followed by $\ell \to \infty$, it remains to consider the sums over $\mathrm{E}\big[|e_{1:i-2}(Z_{ni}^{\varepsilon/2*} \pm q_n)\,\mathbf{1}(\frac{b_n\varepsilon}{2} + q_n > Z_{ni}^{\varepsilon/2*} \geq \ell - q_n)\big]$. We only treat the sum involving the plus-sign. After conditioning on $Z_{ni}^{\varepsilon/2*}$ we are left with bounding $\mathrm{E}\,|e_{1:i-2}(z)|$ for $z \in [\ell, \varepsilon b_n]$ (note that $\frac{b_n\varepsilon}{2} + q_n > Z_{ni}^{\varepsilon/2*}$ implies that $Z_{ni}^{\varepsilon/2*} + q_n \leq b_n\varepsilon/2 + 2q_n \leq b_n\varepsilon$ for sufficiently large n). Decompose $e_{1:i-2} = e_{1:i-2}^{\text{even}} + e_{1:i-2}^{\text{odd}}$

where $e^{\text{even}}_{1:i-2}$ and $e^{\text{odd}}_{1:i-2}$ denote the sum over the even and the odd blocks, respectively. It suffices to treat both sums separately, and we give the details for the sum over the even blocks. Let

$$V_j = V_j(z) = \sum_{s \in I_{2j}} \{\mathbf{1}(U_s^\varepsilon > 1 - z/b_n) - z/b_n\},$$

such that $e^{\text{even}}_{1:i-2}(z) = k_n^{-1/2} \sum_{j=1}^{\lfloor i/2 \rfloor - 1} V_j$. Note that $\alpha(\sigma(V_j), \sigma(V_{j+1})) \le \alpha_{c_2}(b_n)$. Repeatedly applying the coupling construction from (9.3) above (with $\gamma = 2$, $V_1^* = V_1$ and, in the jth step, $X = (V_1^*, \dots, V_j^*)$ and $Y = V_{j+1}$), together with Theorem 5.1 in Bradley (2005), we can inductively construct an iid sequence $(V_j^*)_{j \ge 1}$ such that V_j^* has the same distribution as V_j for any j and such that

$$\mathbb{P}(|V_j - V_j^*| \ge q_n') \le 18 \cdot k_n^{1/5} \alpha_{c_2}(b_n)^{4/5},$$

where $q_n' = \|V_j\|_2 / \sqrt{k_n}$. Note that, since $z \le \varepsilon b_n$, we have $\|V_j\|_2 \le C\sqrt{z + z^2}$ by Condition 2.1(iv). Now

$$\mathrm{E}\,|e^{\text{even}}_{1:i-2}(z)| \le k_n^{-1/2}\, \mathrm{E} \Big| \textstyle\sum_{j=1}^{\lfloor i/2 \rfloor - 1} V_j^* \Big| + i k_n^{-1/2}\, \mathrm{E}\,|V_j - V_j^*|.$$

Since V_j^* is a centered iid sequence, we have the bound

$$\mathrm{E} \Big| \textstyle\sum_{j=1}^{\lfloor i/2 \rfloor - 1} V_j^* \Big| \le \Big\{ \mathrm{Var}\Big(\textstyle\sum_{j=1}^{\lfloor i/2 \rfloor - 1} V_j^* \Big) \Big\}^{1/2} \le i^{1/2}\, \|V_j\|_2.$$

By the Cauchy-Schwarz-inequality, we further have

$$\mathrm{E}\,|V_j - V_j^*| \le q_n' + \mathrm{E}\,|V_j - V_j^*|\, \mathbf{1}(|V_j - V_j^*| \ge q_n') \le q_n' + 2\|V_j\|_2 \sqrt{18}\, k_n^{1/10} \alpha_{c_2}(b_n)^{2/5}.$$

As a consequence,

$$\mathrm{E}\,|e^{\text{even}}_{1:i-2}(z)| \le \{\sqrt{i/k_n} + i k_n^{-1} + 9 \cdot i k_n^{-2/5} \alpha_{c_2}(b_n)^{2/5}\} \|V_j\|_2$$

for any $z \in [0, \varepsilon b_n]$, where $\|V_j\|_2 \le C\sqrt{z + z^2} \le C(1+z)$ by Condition 2.1(iv). A similar bound for the sum over the odd blocks finally implies that

$$\begin{aligned} &\mathrm{E}\left[|e_{1:i-2}(Z_{ni}^{\varepsilon/2*} + q_n)\, \mathbf{1}(\tfrac{b_n \varepsilon}{2} + q_n > Z_{ni}^{\varepsilon/2*} \ge \ell - q_n)\right] \\ &\quad \le C\{\sqrt{i/k_n} + i k_n^{-1} + 9 \cdot i k_n^{-2/5} \alpha_{c_2}(b_n)^{2/5}\}\, \mathrm{E}\left[(1 + Z_{n1}^{\varepsilon/2*} + q_n)\, \mathbf{1}(Z_{n1}^{\varepsilon/2*} \ge \ell - q_n)\right] \end{aligned}$$

after conditioning on $Z_{ni}^{\varepsilon/2*}$. Note that the limes superior for $n \to \infty$ of the moment on the right-hand side can be made arbitrary small by increasing ℓ. To finalize the treatment of $\mathrm{E}[|\bar{S}^\varepsilon_{n,\ell,1}|]$ we are hence left with bounding the expression

$$\frac{1}{k_n} \sum_{i=3}^{k_n} \{\sqrt{i/k_n} + i k_n^{-1} + 9 \cdot i k_n^{-2/5} \alpha_{c_2}(b_n)^{2/5}\} \le C + C' \cdot k_n^{3/5} \alpha_{c_2}(b_n)^{2/5}.$$

Since $\alpha_{c_2}(b_n)^{2/5} = O(b_n^{-2\eta/5}) = O(b_n^{-6/5})$, we obtain that $k_n^{3/5} \alpha_{c_2}(b_n)^{2/5} = O((k_n/b_n^2)^{3/5})$, which converges to zero under the assumption that $k_n/b_n^2 = o(1)$. □

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

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

Proof. Recall that $H(x) = 1 - \exp(-\theta x)$. We have to show that

$$\int_0^\ell e_n(x)\,\mathrm{d}(\hat H_{k_n} - H)(x) = o_\mathbb{P}(1), \qquad n \to \infty,$$

which follows from Lemma C.8 in Berghaus and Bücher (2016), provided we can show that

$$\sup_{x\in[0,\ell]} |\hat H_{k_n}(x) - H(x)| = o_\mathbb{P}(1), \qquad n \to \infty.$$

The last display in turn follows from pointwise convergence (in probability) of $\hat H_{k_n}$ to H by a standard Gilvenko-Cantelli-type argument. For the pointwise convergence, note that $\mathrm{E}[\hat H_{k_n}(x)] = H_{k_n}(x) := \mathbb{P}(Z_{n1} \le x) \to H(x)$ by (1.3). By similar arguments as in the proof of Proposition 3.1 in Robert et al. (2009) (but under slightly different assumptions) it can be shown that

$$\lim_{n\to\infty} k_n \operatorname{Var}\{\hat H_{k_n}(x)\} = e^{-\theta x}(1 - e^{-\theta x}).$$

This implies pointwise convergence in probability and hence the Lemma. □

Lemma 9.3 (Joint convergence of fidis)**.** *Under Condition 2.1, for any $x_1, \dots, x_m \in [0,\infty)$,*

$$\Big(e_n(x_1), \dots, e_n(x_m), G_n\Big)' \rightsquigarrow \Big(e(x_1), \dots, e(x_m), G\Big)',$$

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

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

Here, $r(0,0) = h(0) = 0$ and, for $x \ge y \ge 0$ with $x \ne 0$,

$$r(x,y) = \theta x \sum_{i=1}^\infty \sum_{j=0}^i ij\pi_2^{(y/x)}(i,j), \qquad h(x) = \int_0^x \sum_{i=1}^\infty i p_2^{(x,y)}(i,0)\,\mathrm{d}y - x/\theta,$$

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

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

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

Proof. Note that weak convergence of the first m components of the vector follows from Theorem 4.1 in Robert (2009). Regarding joint convergence with the $(m+1)$st component, we only consider the case $m = 1$ and set $x_1 = x$; the general case can be treated analogously.

Recall the definition of ℓ_n in Condition 2.1(iii). Decompose blocks $I_i = I_i^+ \cup I_i^-$, where

$$I_i^+ = \{(i-1)b_n + 1, \dots, ib_n - \ell_n\}, \qquad I_i^- = \{ib_n - \ell_n + 1, \dots, ib_n\}.$$

and let

$$e_n^+(x) = k_n^{-1/2} \sum_{i=1}^{k_n} \sum_{s \in I_i^+} \{\mathbf{1}(U_s > 1 - x/b_n) - x/b_n\}$$

$$G_n^+ = k_n^{-1/2} \sum_{i=1}^{k_n} Z_{ni}^+ - \mathrm{E}[Z_{ni}^+], \qquad Z_{ni}^+ = b_n(1 - \max_{s \in I_i^+} U_s).$$

As a consequence of Lemma 6.6 in Robert (2009), $e_n^-(x) = e_n(x) - e_n^+(x) = o_{\mathbb{P}}(1)$. Let us show the same for G_n. Denote $G_n^- = G_n - G_n^+$ and $Z_{ni}^- = Z_{ni} - Z_{ni}^+$. For $\varepsilon \in (0, c_1 \wedge c_2)$, let $A_n^+ = \{\min_{i=1}^{k_n} N_{ni}^+ > 1 - \varepsilon\}$ and note that $\mathbb{P}(A_n^+) \to 1$ by Condition 2.1(v). It then suffices to show that $G_n^- \mathbf{1}_{A_n^+} = o_{\mathbb{P}}(1)$. We can write $G_n^- \mathbf{1}_{A_n^+} = \tilde G_n^- \mathbf{1}_{A_n^+} = \tilde G_n^- + o_{\mathbb{P}}(1)$, where

$$\tilde G_n^- = k_n^{-1/2} \sum_{i=1}^{k_n} \{Z_{ni}^- - \mathrm{E}[Z_{ni}^-]\} \, \mathbf{1}(N_{ni}^+ > 1 - \varepsilon)$$

Now, $N_{ni}^+ > 1 - \varepsilon$ implies that $Z_{ni}^- = Z_{ni}^{\varepsilon -}$, where the latter variable is defined in terms of the U_i^ε instead of the U_i. Hence, $\tilde G_n^- = k_n^{-1/2} \sum_{i=1}^{k_n} S_{ni}^\varepsilon$, where

$$S_{ni}^\varepsilon = \{Z_{ni}^{\varepsilon -} - \mathrm{E}[Z_{ni}^-]\} \, \mathbf{1}(N_{ni}^{\varepsilon +} > 1 - \varepsilon)$$

is $\mathcal{B}^\varepsilon_{\{(i-1)b_n+1\}:(ib_n)}$-measurable. As a consequence, by stationarity

$$\begin{aligned} \mathrm{Var}(\tilde G_n^-) &= \mathrm{Var}(S_{n1}^\varepsilon) + \frac{2}{k_n} \sum_{i=1}^{k_n} (k_n - i) \, \mathrm{Cov}(S_{n1}^\varepsilon, S_{n,1+i}^\varepsilon) \\ &\le 3 \, \mathrm{Var}(S_{n1}^\varepsilon) + \frac{2}{k_n} \sum_{i=2}^{k_n} (k_n - i) \, \mathrm{Cov}(S_{n1}^\varepsilon, S_{n,1+i}^\varepsilon) \end{aligned} \tag{9.4}$$

Let us first show that $\mathrm{Var}(S_{n1}^\varepsilon) = o(1)$ as $n \to \infty$, which would follow, if we show that, for any $p \in (2, 2 + \delta)$, $|Z_{n1}^{\varepsilon -}| \le |Z_{n1}^-| \to 0$ in L_p (the inequality follows by studying the cases $N_{ni}^+ > 1 - \varepsilon$ and $\le 1 - \varepsilon$). Since $\ell_n = o(b_n)$ we have, for any $y > 0$,

$$\begin{aligned} \mathbb{P}(Z_{n1}^- \ne 0) &= \mathbb{P}\Big(\max_{s \in I_1} U_s > \max_{s \in I_1^+} U_s \Big) \\ &\le \mathbb{P}\Big(\max_{s=1}^{b_n - \ell_n} U_s \le 1 - y/b_n \Big) + \mathbb{P}\Big(\max_{s=1}^{\ell_n} U_s > 1 - y/b_n \Big) \\ &\le \mathbb{P}\Big(Z_{1:b_n - \ell_n} \ge y(b_n - \ell_n)/b_n \Big) + \ell_n y / b_n \\ &\to \exp(-\theta y), \end{aligned} \tag{9.5}$$

which can be made arbitrary small by increasing y. Hence, $Z_{n1}^- = o_{\mathbb{P}}(1)$. Since $\mathrm{E}\,|Z_{n1}^-|^p \le C \,\mathrm{E}\,|Z_{1:b_n - \ell_n}|^p < \infty$ for any $p \in (2, 2+\delta)$ by Condition 2.1(vi), we can conclude that $Z_{n1}^- \to 0$ in L_p.

It remains to treat the sum over the covariances on the right-hand side of (9.4). By Lemma 3.11 in Dehling and Philipp (2002) (which is a slightly more general version of Lemma 6.3 in Robert, 2009), for any $p \in (2, 2 + \delta)$,

$$|\, \mathrm{Cov}(S_{n1}^\varepsilon, S_{n,1+i}^\varepsilon)| \le 10 (\mathrm{E}\,|S_{n1}^\varepsilon|^p)^{2/p} \alpha_{c_2}((i-1)b_n)^{1-2/p}$$

(note that S_{ni}^{ε} is $\mathcal{B}^{\varepsilon}_{(ib_n-b_n+1):(ib_n)}$-measurable). Now, for $i \ge 2$, $\alpha_{c_2}((i-1)b_n) \le \alpha_{c_2}(i-1) \le C(i-1)^{-\eta}$ by monotonicity of $\alpha_{c_2}(\ell)$. The sum over the covariances in (9.4) can thus be bounded by a multiple of

$$(\mathrm{E}\,|S_{n1}^{\varepsilon}|^p)^{2/p} \sum_{i=2}^{k_n} \alpha_{c_2}((i-1)b_n)^{1-2/p} \le (\mathrm{E}\,|S_{n1}^{\varepsilon}|^p)^{2/p} \sum_{i=1}^{\infty} i^{-\eta(1-2/p)}.$$

The series converges and the moment converges to 0 by arguments as given above.

Now, since $(e_n^-(x), G_n^-) = o_{\mathbb{P}}(1)$ and $\mathbb{P}(A_n^+) \to 1$, it suffices to show that $(e_n^+(x), G_n^+)\,\mathbf{1}_{A_n^+}$ converges weakly to the claimed normal distribution. This in turn follows from the Cramér-Wold device, provided we show that for any $\lambda_1, \lambda_2 \in \mathbb{R}$

$$(\lambda_1 e_n^+(x) + \lambda_2 G_n^+)\,\mathbf{1}_{A_n^+} \rightsquigarrow \lambda_1 e(x) + \lambda_2 G.$$

The left-hand side can be rewritten as $(k_n^{-1/2} \sum_{i=1}^{k_n} \tilde f_{i,n})\,\mathbf{1}_{A_n^+} = k_n^{-1/2} \sum_{i=1}^{k_n} \tilde f_{i,n} + o_{\mathbb{P}}(1)$, where $\tilde f_{i,n} = f_{i,n}\,\mathbf{1}(Z_{ni}^+ < \varepsilon b_n)$ and

$$f_{i,n} = \lambda_1 \sum\nolimits_{s \in I_i^+} \{\mathbf{1}(U_s > 1 - x/b_n) - x/b_n\} + \lambda_2 (Z_{ni}^+ - \mathrm{E}[Z_{ni}^+]).$$

Note that $\tilde f_{i,n}$ is $\mathcal{B}^{\varepsilon}_{\{(i-1)b_n+1\}:\{ib_n-\ell_n\}}$-measurable. A standard argument based on characteristic functions (see, e.g., the proof of Lemma 6.7 in Robert, 2009) shows that the weak limit of $k_n^{-1/2} \sum_{i=1}^{k_n} \tilde f_{i,n}$ is the same as if the $(\tilde f_{i,n})_{i=1,\dots,k_n}$ were considered as iid. Now,

$$\frac{\sum_{i=1}^{k_n} \mathrm{E}[|\tilde f_{i,n}|^p]}{\big(\sum_{i=1}^{k_n} \mathrm{E}[|\tilde f_{i,n}|^2]\big)^{p/2}} = k_n^{1-p/2} \frac{\mathrm{E}[|\tilde f_{i,n}|^p]}{\Big(\mathrm{E}[|\tilde f_{i,n}|^2]\Big)^{p/2}}.$$

By Minkowski's inequality, for any $p \in (2, 2+\delta)$, $\sup_n \mathrm{E}[|\tilde f_{1,n}|^p] < \infty$ by Condition 2.1(vi) and (ii). As a consequence, provided $\lim_{n\to\infty} \mathrm{E}[\tilde f_{1,n}^2]$ exists, Ljapunov's condition is satisfied (Billingsley, 1979, Theorem 27.3) and $k_n^{-1/2} \sum_{i=1}^{k_n} \tilde f_{i,n}$ converges to a normal distribution with variance equal to $\lim_{n\to\infty} \mathrm{E}[\tilde f_{1,n}^2]$.

The latter limit is equal to $\lim_{n\to\infty} \mathrm{E}[f_{1,n}^2]$, whence it remains to be shown that

$$\lim_{n\to\infty} \mathrm{E}[f_{1,n}^2] = \lambda_1^2 r(x,x) + 2\lambda_1\lambda_2 h(x) + \lambda_2^2/\theta^2,$$

which in turn follows, observing the expressions for the limiting covariances $r(x,x)$ in Theorem 4.1 in Robert (2009), from

$$\lim_{n\to\infty} \mathrm{Cov}\left\{ \sum\nolimits_{s\in I_1^+} \mathbf{1}(U_s > 1 - x/b_n), b_n(1 - \max\nolimits_{s \in I_1^+} U_s) \right\} = h(x),$$
$$\lim_{n\to\infty} \mathrm{Var}\left\{ b_n(1 - \max\nolimits_{s\in I_1^+} U_s) \right\} = \theta^{-2}.$$

Repeating arguments from above, we may replace the set I_1^+ by I_1 in the preceding display, whence it is in fact sufficient to show that

$$\lim_{n\to\infty} \mathrm{Cov}(N_n^{(x)}(E), Z_{1:n}) = h(x), \qquad \lim_{n\to\infty} \mathrm{Var}(Z_{1:n}) = \theta^{-2}.$$

By an application of Theorem 2.20 in van der Vaart (1998), the second assertion follows directly from $Z_{1:n} \rightsquigarrow \exp(\theta)$ and Condition 2.1(vi). For the first convergence, abbreviate

$N_n^{(x)} = N_n^{(x)}(E)$ and note that

$$\mathbb{P}(N_n^{(x)} = i, Z_{1:n} > y) = \mathbb{P}(N_n^{(y)} = 0, N_n^{(x)} = i) \to \begin{cases} p_2^{(x,y)}(i,0) & x \geq y \geq 0 \\ 0 & y > x \geq 0, \end{cases}$$

see Perfekt (1994); Robert (2009), that is, $(N_n^{(x)}, Z_{1:n})$ converges jointly. By uniform integrability, we may deduce that

$$\mathrm{E}[N_n^{(x)} Z_{1:n}] = \sum_{i=1}^\infty i \int_0^\infty \mathbb{P}(Z_{1:n} > y, N_n^{(x)} = i) \,\mathrm{d}y \to \sum_{i=1}^\infty i \int_0^x p_2^{(x,y)}(i,0) \,\mathrm{d}y.$$

The lemma finally follows from $\mathrm{E}[Z_{1:n}] \to \theta^{-1}$ and $\mathrm{E}[N_n^{(x)}] \to x$. □

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

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

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.

Proof. This follows directly from Theorem 4.1 in Robert (2009). □

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

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

Proof. As a consequence of Lemma 9.2, Lemma 9.4 and the continuous mapping theorem, we have

$$D_{n,\ell} + G_n = \theta \int_0^\ell e_n(x)\, e^{-\theta x} \,\mathrm{d}x + G_n + o_{\mathbb{P}}(1) \rightsquigarrow \theta \int_0^\ell e(x)\, e^{-\theta x} \,\mathrm{d}x + G.$$

The right-hand side is normally distributed with variance σ_ℓ^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 3.1, respectively.

Proof. Since

$$\lim_{\ell \to \infty} \sigma_\ell^2 = \sigma_\infty^2 = \theta^2 \int_0^\infty \int_0^\infty r(x,y) e^{-\theta(x+y)} \,\mathrm{d}x \,\mathrm{d}y + 2\theta \int_0^\infty h(x) e^{-\theta x} \,\mathrm{d}x + \theta^{-2}, \quad (9.6)$$

we only have to show, that $\sigma_\infty^2 = \sigma_{\mathrm{dj}}^2$. First of all, note that, for $x > y$,

$$r(x,y) = \theta x \sum_{i=1}^\infty \sum_{j=0}^i ij \pi_2^{(y/x)}(i,j) = \theta x \, \mathrm{E}[\zeta_1^{(y/x)} \zeta_2^{(y/x)}],$$

where $(\zeta_1^{(y/x)}, \zeta_2^{(y/x)}) \sim \pi_2^{(y/x)}$. Using this representation and substituting $\sigma = \frac{y}{x}$ we obtain

$$\theta^2 \int_0^\infty \int_0^\infty r(x,y) e^{-\theta(x+y)} \,\mathrm{d}x\,\mathrm{d}y = 2\theta^2 \int_0^\infty \int_0^x \theta x\,\mathrm{E}[\zeta_1^{(y/x)} \zeta_2^{(y/x)}] e^{-\theta(x+y)} \,\mathrm{d}x\,\mathrm{d}y$$
$$= 2\theta^3 \int_0^1 \mathrm{E}[\zeta_1^{(\sigma)} \zeta_2^{(\sigma)}] \int_0^\infty x^2 e^{-\theta(1+\sigma)x} \,\mathrm{d}x \,\mathrm{d}\sigma = 4 \int_0^1 \frac{\mathrm{E}[\zeta_1^{(\sigma)} \zeta_2^{(\sigma)}]}{(1+\sigma)^3} \,\mathrm{d}\sigma,$$

which is exactly the first summand in σ_{dj}^2.

Consider the second integral in σ_∞^2. By the definition of $p_2^{(x,y)}$ in Lemma 9.3 we have

$$\sum_{i=1}^\infty i p_2^{(x,y)}(i,0) = \mathrm{E}\Big[\sum_{j=1}^\eta \zeta_{j1}^{(y/x)} \mathbf{1}\Big(\sum_{j=1}^\eta \zeta_{j2}^{(y/x)} = 0\Big)\Big],$$

where $\eta \sim \mathrm{Poisson}(\theta x)$ is independent of iid random vectors $(\zeta_{i1}^{(y/x)}, \zeta_{i2}^{(y/x)}) \sim \pi_2^{(y/x)}, i \in \mathbb{N}$. With the identity $\mathbb{P}(\zeta_{12}^{(\sigma)} = 0) = 1 - \sigma$, which we will show later, the latter expectation can further be rewritten as

$$\sum_{k=1}^\infty \mathrm{E}\Big[\sum_{j=1}^k \zeta_{j1}^{(y/x)} \mathbf{1}\Big(\sum_{j=1}^k \zeta_{j2}^{(y/x)} = 0\Big)\Big]\mathbb{P}(\eta = k)$$
$$= \sum_{k=1}^\infty k\,\mathrm{E}\big[\zeta_{11}^{(y/x)} \mathbf{1}(\zeta_{12}^{(y/x)} = 0)\big]\mathbb{P}(\zeta_2^{(y/x)} = 0)^{k-1}\mathbb{P}(\eta = k)$$
$$= \sum_{k=1}^\infty k\,\mathrm{E}\big[\zeta_{11}^{(y/x)} \mathbf{1}(\zeta_{12}^{(y/x)} = 0)\big](1 - y/x)^{k-1}\frac{(\theta x)^k}{k!} e^{-\theta x}$$
$$= \mathrm{E}\big[\zeta_{11}^{(y/x)} \mathbf{1}(\zeta_{12}^{(y/x)} = 0)\big]\theta x e^{-\theta y}. \tag{9.7}$$

Hence, substituting $\sigma = y/x$,

$$h(x) = \theta x^2 \int_0^1 \mathrm{E}\big[\zeta_1^{(\sigma)} \mathbf{1}(\zeta_2^{(\sigma)} = 0)\big] e^{-\theta\sigma x} \,\mathrm{d}\sigma - \frac{x}{\theta} \tag{9.8}$$

and therefore

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

which corresponds to the remaining summands in σ_{dj}^2.

It remains to be shown that

$$\mathbb{P}(\zeta_{12}^{(\sigma)} = 0) = 1 - \sigma. \tag{9.9}$$

By the definition of $\pi_2^{(\sigma)}$ in Section 2, we have

$$\mathbb{P}(\zeta_2^{(\sigma)} = 0) = 1 - \mathbb{P}(\zeta_2^{(\sigma)} > 0) = 1 - \lim_{n\to\infty} \mathbb{P}(N_n^{(\sigma x)}(B_n) > 0 | N_n^{(x)}(B_n) > 0)$$
$$= 1 - \lim_{n\to\infty} \mathbb{P}(N_{1:q_n} > 1 - \tfrac{\sigma x}{n} | N_{1:q_n} > 1 - \tfrac{x}{n})$$
$$= 1 - \lim_{n\to\infty} \mathbb{P}\Big(\frac{N_{1:q_n} - (1 - \frac{x}{n})}{\frac{x}{n}} > 1 - \sigma \big| N_{1:q_n} > 1 - \tfrac{x}{n}\Big).$$

Finally, by (2.2), we can use identity (10.21) in Beirlant et al. (2004), which is an implication of Theorem 3.1 in Segers (2005), to deduce that, as $n \to \infty$,

$$\mathbb{P}\Big(\tfrac{N_{1:q_n}-(1-\frac{x}{n})}{\frac{x}{n}} > 1-\sigma \Big| N_{1:q_n} > 1-\tfrac{x}{n}\Big) = \mathbb{P}\Big(\tfrac{U_1-(1-\frac{x}{n})}{\frac{x}{n}} > 1-\sigma \Big| U_1 > 1-\tfrac{x}{n}\Big) + o(1),$$

which converges to σ as asserted. □

Proof of Proposition 5.1. Let

$$\beta_\varepsilon(\ell) = \sup_{k\in\mathbb{N}} \beta(\mathcal{B}^\varepsilon_{1:k}, \mathcal{B}^\varepsilon_{k+\ell:\infty}) = \sup_{k\in\mathbb{N}} \frac{1}{2} \sup \sum_{i\in I}\sum_{j\in J} |\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|,$$

where the last supremum is over all finite partitions $(A_i)_{i\in I} \subset \mathcal{B}^\varepsilon_{1:k}$ and $(B_j)_{j\in J} \subset \mathcal{B}^\varepsilon_{k+\ell:\infty}$ of Ω. Decompose

$$\hat\sigma^2_{\mathrm{dj}} = \frac{1}{k_n}\sum_{j=1}^{k_n} \hat B^2_{nj} = A_{n1} + 2A_{n2} + A_{n3},$$

where

$$A_{n1} = \frac{1}{k_n}\sum_{j=1}^{k_n} B^2_{nj}, \quad A_{n2} = \frac{1}{k_n}\sum_{j=1}^{k_n} (\hat B_{nj} - B_{nj})B_{nj}, \quad A_{n3} = \frac{1}{k_n}\sum_{j=1}^{k_n} (\hat B_{nj} - B_{nj})^2.$$

By the Cauchy-Schwarz inequality, it suffices to show that $A_{n3} = o_\mathbb{P}(1)$ and that $A_{n1} = \sigma^2_{\mathrm{dj}} + o_\mathbb{P}(1)$.

Let us first show that $A_{n3} = o_\mathbb{P}(1)$. Note that $U_s > 1 - Z_{nj}/b_n$ iff $\hat U_s > 1 - \hat Z_{nj}/b_n$, almost surely. As a consequence, by a similar calculation as in (3.6), we can write

$$\begin{aligned}\hat B_{nj} - B_{nj} &= \hat Z_{nj} - Z_{nj} + \frac{1}{\theta} - \hat T_n + \frac{1}{k_n}\sum_{i=1}^{k_n}(Z_{ni} - \hat Z_{ni}) \\ &= \frac{e_n(Z_{nj})}{\sqrt{k_n}} + \frac{1}{\theta} - \hat T_n - \frac{1}{\sqrt{k_n}}\sqrt{k_n}(\hat T_n - T_n) = \frac{e_n(Z_{nj})}{\sqrt{k_n}} + O_\mathbb{P}(k_n^{-1/2})\end{aligned}$$

almost surely, where the $O_\mathbb{P}$-term is uniformly in $j = 1, \dots, n$. We may further write

$$e_n(Z_{nj}) = -\sqrt{n/k_n}\cdot \mathbb{F}_n(1 - Z_{nj}/b_n),$$

where $\mathbb{F}_n(u) = n^{-1/2}\sum_{s=1}^n \{\mathbf{1}(U_s \le u) - u\}$ denotes the usual empirical process. By weak convergence of that process (a consequence of the assumption on beta-mixing) we can conclude that $\max_{j=1}^n |e_n(Z_{nj})| = O_\mathbb{P}(b_n^{1/2})$. Hence,

$$\begin{aligned}A_{n3} &= \frac{1}{k_n^2}\sum_{j=1}^{k_n}\{e_n(Z_{nj}) + O_\mathbb{P}(1)\}^2 = \Big\{\frac{1}{k_n^2}\sum_{j=1}^{k_n} e_n^2(Z_{nj})\Big\} + O_\mathbb{P}(b_n^{1/2}k_n^{-1}) + O_\mathbb{P}(k_n^{-1}) \\ &\le \frac{1}{k_n}\max_{j=1}^n |e_n(Z_{nj})| \int_0^\infty |e_n(z)|\,\mathrm{d}\hat H_{k_n}(z) + o_\mathbb{P}(1).\end{aligned}$$

Repeating arguments from the proof of Theorem 3.1 (Wichura's theorem), it can be seen that the dominating term on the right-hand side of this display is of the order $O_\mathbb{P}(\sqrt{b_n}/k_n)$, which converges to 0 by assumption.

It remains to be shown that $A_{n1} = \sigma^2_{\mathrm{dj}} + o_{\mathbb{P}}(1)$. For that purpose, write $A_{n1} = C_{n1} + 2C_{n2} + C_{n3}$, where

$$C_{n1} = \frac{1}{k_n}\sum_{j=1}^{k_n}(Z_{nj} - \theta^{-1})^2$$

$$C_{n2} = \frac{1}{k_n}\sum_{j=1}^{k_n}(Z_{nj} - \theta^{-1})\Big\{ \textstyle\sum_{s\in I_j} \frac{1}{k_n}\sum_{i=1}^{k_n} \mathbf{1}(U_s > 1 - \frac{Z_{ni}}{b_n}) - \frac{Z_{ni}}{b_n}\Big\}$$

$$C_{n3} = \frac{1}{k_n}\sum_{j=1}^{k_n}\Big\{ \textstyle\sum_{s\in I_j} \frac{1}{k_n}\sum_{i=1}^{k_n} \mathbf{1}(U_s > 1 - \frac{Z_{ni}}{b_n}) - \frac{Z_{ni}}{b_n}\Big\}^2$$

From the proof of Lemma 9.6 we know that $\sigma^2_{\mathrm{dj}} = \sigma^2_\infty$, where σ^2_∞ is defined in (9.6). Therefore, it suffices to show that

$$C_{n1} \xrightarrow{p} \theta^{-2}, \quad C_{n2} \xrightarrow{p} \theta\int_0^\infty h(x)e^{-\theta x}\,\mathrm{d}x, \quad C_{n3} \xrightarrow{p} \theta^2\int_0^\infty\int_0^\infty r(x,y)e^{-\theta(x+y)}\,\mathrm{d}x\,\mathrm{d}y.$$

The first convergence can be shown by considering expectations and variances: first, $\mathrm{E}[C_{n1}] = \mathrm{E}[(Z_{n1} - \theta^{-1})^2] \to \theta^{-2}$ by Condition 2.1(vi) and weak convergence of Z_{n1}. Second,

$$\mathrm{Var}(C_{n1}) = \frac{1}{k_n}\mathrm{Var}\left\{(Z_{n1} - \theta^{-1})^2\right\} + \frac{1}{k_n}\sum_{\ell=1}^{k_n}\frac{k_n - \ell}{k_n}\mathrm{Cov}\{(Z_{n1} - \theta^{-1})^2, (Z_{n,1+\ell} - \theta^{-1})^2\}$$

which is of the order $O(k_n^{-1})$ by a standard inequality for covariances of strongly mixing time series and by finiteness of moments of Z_{nj} of order larger than 4.

Consider C_{n2}. For integer $\ell \ge 1$, let

$$C_{n2}(\ell) = \frac{1}{k_n^2}\sum_{\substack{j,i\in\{1,\dots,k_n\}\\ |j-i|\ge 2}}\Big\{(Z_{nj} - \theta^{-1})\textstyle\sum_{s\in I_j} f(U_s, Z_{ni})\Big\}\mathbf{1}(Z_{ni} \le \ell),$$

where $f(u,z) = \mathbf{1}(u > 1 - z/b_n) - z/b_n$. Using similar arguments as in the proof of Lemma 9.1 it can be shown that, for any $\delta > 0$, $\limsup_{n\to\infty}\mathbb{P}(|C_{n2}(\ell) - C_{n2}| > \delta)$ converges to 0 for $\ell \to \infty$. Therefore, by Wichura's theorem (Billingsley, 1979, Theorem 25.5), it is sufficient to show that

$$C_{n2}(\ell) \to C_2(\ell) = \theta\int_0^\ell h(x)e^{-\theta x}\,\mathrm{d}x, \qquad n \to \infty,$$

holds for any $\ell \in \mathbb{N}$. For that purpose, we will show that $\mathrm{E}[C_{n2}(\ell)] \to C_2(\ell)$ and that $\mathrm{Var}(C_{n2}(\ell)) \to 0$ as $n \to \infty$.

Recall Berbee's coupling Lemma (Berbee, 1979): if X and Y are two random variables in some Borel spaces S_1 and S_2, respectively, then there exists a random variable U independent of (X,Y) and a measurable function f such that $Y^* = f(X,Y,U)$ has the same distribution as Y, is independent of X and satisfies $\mathbb{P}(Y \ne Y^*) = \beta(\sigma(X), \sigma(Y))$. Apply this lemma with $X = (U_s)_{s\in I_j}$ and $Y = Z_{ni}$ (with $|i-j| \ge 2$) to construct a random variable $Z_{ni}^* \sim H_{k_n}$ (H_{k_n} denoting the cdf of Z_{n1}) independent of $(U_s)_{s\in I_j}$ satisfying $\mathbb{P}(Z_{ni} \neq Z_{ni}^*) \le \beta(b_n)$. Write

$$\begin{aligned} &\mathrm{E}\left[(Z_{nj} - \tfrac{1}{\theta}) \textstyle\sum_{s\in I_j} f(U_s, Z_{ni})\, \mathbf{1}(Z_{ni} \le \ell)\right] \qquad (9.10)\\ &= \mathrm{E}\left[(Z_{nj} - \tfrac{1}{\theta}) \textstyle\sum_{s\in I_j} f(U_s, Z_{ni}^*)\, \mathbf{1}(Z_{ni}^* \le \ell)\right] \\ &\quad + \mathrm{E}\Big[(Z_{nj} - \tfrac{1}{\theta}) \textstyle\sum_{s\in I_j} \big\{ f(U_s, Z_{ni})\, \mathbf{1}(Z_{ni} \le \ell) - f(U_s, Z_{ni}^*)\, \mathbf{1}(Z_{ni}^* \le \ell)\big\}\, \mathbf{1}(Z_{ni} \neq Z_{ni}^*)\Big] \end{aligned}$$

By Hölder's and Minkowski's inequality, the second expectation on the right-hand side of this display can be bounded in absolute value by

$$\|Z_{nj} - \tfrac{1}{\theta}\|_3 \textstyle\sum_{s\in I_j} \big\{ \|f(U_s, Z_{ni})\, \mathbf{1}(Z_{ni} \le \ell)\|_3 + \|f(U_s, Z_{ni}^*)\, \mathbf{1}(Z_{ni}^* \le \ell)\|_3 \big\} \beta(b_n)^{1/3}.$$

This bound converges to 0, since $|f(U_s, Z_{ni})| \le 1$ and since the assumptions imply that $\limsup_{n\to\infty} \|Z_{n1} - \frac{1}{\theta}\|_3 \le C$ and that $b_n \beta(b_n)^{1/3} = o(1)$.

As a consequence, rewriting the first summand on the right-hand side of (9.10), we obtain that

$$\mathrm{E}[C_{n2}(\ell)] = \mathrm{E}[h_n(Z_{n1}^*)\, \mathbf{1}(Z_{n1}^* \le \ell)] + o(1),$$

where $h_n(x) = \mathrm{E}\left[(Z_{n1} - \theta^{-1}) \sum_{s\in I_1} f(U_s, x)\right]$. By Condition 2.1(ii) and (vi) $h_n(Z_{n1}^*)$ is uniformly integrable. Hence, to obtain that $\mathrm{E}[C_{n2}(\ell)] \to C_2(\ell)$ we only have to show that $h_n(Z_{n1}^*)\, \mathbf{1}(Z_{n1}^* \le \ell) \rightsquigarrow h(Z)\, \mathbf{1}(Z \le \ell)$ with Z being exponentially distributed with parameter θ. This in turn follows from the extended continuous mapping theorem, since $Z_{n1}^* \rightsquigarrow Z$ and $h_n(x_n)\, \mathbf{1}(x_n \le \ell) \to h(x)\, \mathbf{1}(x \le \ell)$ for any sequence $x_n \to x \neq \ell$. To see the latter, note that, for $x < \ell$ and n large enough, Minkowski's inequality and Condition 2.1(ii) and (vi) imply that

$$|h_n(x_n) - h_n(x)| = \big|\, \mathrm{E}\left[(Z_{n1} - \theta^{-1})\{N_{b_n}^{(x_n)}(E) - N_{b_n}^{(x)}(E)\}\right]\big| \le C \times |x_n - x|^{1/(2+\delta)}.$$

Consider the variance of $C_{n2}(\ell)$. By the Cauchy-Schwarz inequality, up to negligible terms, it can be written as

$$\begin{aligned} k_n^{-4} \sum_{(i,i',j,j')\in J} \mathrm{Cov}\Big(&(Z_{nj} - \theta^{-1}) \textstyle\sum_{s\in I_j} f(U_s, Z_{ni})\, \mathbf{1}(Z_{ni} \le \ell), \\ &(Z_{nj'} - \theta^{-1}) \textstyle\sum_{s'\in I_{j'}} f(U_{s'}, Z_{ni'})\, \mathbf{1}(Z_{ni'} \le \ell)\Big) \qquad (9.11) \end{aligned}$$

where J denote the set of all $(i, i', j, j') \in \{1, \dots, k_n\}^4$ such that any two of the indexes are at distance larger than 2. We have to show that all covariances in this sum converge to 0, uniformly in the indexes.

First, consider the case where either $i \vee j < i' \wedge j'$ or $i' \vee j' < i \wedge j$. Recall Lemma 3.11 in Dehling and Philipp (2002): for real-valued random variables X, Y and real numbers $r, s, t > 1$ such that $1/r + 1/s + 1/t = 1$, we have

$$\big|\, \mathrm{E}[XY] - \mathrm{E}[X]\, \mathrm{E}[Y]\big| \le 10 \|X\|_r \|Y\|_s \alpha(\sigma(X), \sigma(Y))^{1/t}. \qquad (9.12)$$

Therefore, for some $\varepsilon \in (0, \delta)$, the covariances inside the sum in (9.11) are bounded by

$$\|(Z_{nj} - \theta^{-1}) \textstyle\sum_{s\in I_j} f(U_s, Z_{ni})\, \mathbf{1}(Z_{ni} \le \ell)\|_{2+\varepsilon}^2 \{\alpha_1(b_n)\}^{\varepsilon/(2+\varepsilon)},$$

which can be seen to be $o(1)$ by Minkowski's inequality and the Cauchy-Schwarz inequality.

The other cases are slightly more difficult. Consider the case $i < j' < j < i'$. Apply Berbee's coupling Lemma with $X = (U_s)_{s \in I_{j'} \cup I_j \cup I_{i'}}$ and $Y = (U_s)_{s \in U_i}$. Then the mixed moment inside the covariance can be written as

$$\begin{aligned}
&\mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z_{ni}) \mathbf{1}(Z_{ni} \le \ell)(Z_{nj'} - \theta^{-1}) \sum_{s' \in I_{j'}} f(U_{s'}, Z_{ni'}) \mathbf{1}(Z_{ni'} \le \ell)\Big] \\
&\qquad = \mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z^*_{ni}) \mathbf{1}(Z^*_{ni} \le \ell) \\
&\qquad\qquad\qquad \times (Z_{nj'} - \theta^{-1}) \textstyle\sum_{s' \in I_{j'}} f(U_{s'}, Z_{ni'}) \mathbf{1}(Z_{ni'} \le \ell)\Big] + o(1),
\end{aligned}$$

where the remainder term has been handled by Hölder's and Minkowski's inequality just as in (9.10). A second application of Berbee's coupling Lemma (with $X = ((U^*_s)_{s \in I_i}, (U_s)_{s \in I_{j'} \cup I_j})$ and $Y = (U_s)_{s \in I_{i'}}$) allows to rewrite the dominating term in the last display as

$$\begin{aligned}
&\mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z^*_{ni}) \mathbf{1}(Z^*_{ni} \le \ell) \\
&\qquad\qquad \times (Z_{nj'} - \theta^{-1}) \textstyle\sum_{s' \in I_{j'}} f(U_{s'}, Z^*_{ni'}) \mathbf{1}(Z^*_{ni'} \le \ell)\Big] + o(1) \\
&= \mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z^*_{ni}) \mathbf{1}(Z^*_{ni} \le \ell)\Big] \\
&\qquad\qquad \times \mathrm{E}[(Z_{nj'} - \theta^{-1}) \textstyle\sum_{s' \in I_{j'}} f(U_{s'}, Z^*_{ni'}) \mathbf{1}(Z^*_{ni'} \le \ell)\Big] + o(1),
\end{aligned}$$

where the latter equality follows from (9.12). Since

$$\begin{aligned}
&\mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z^*_{ni}) \mathbf{1}(Z^*_{ni} \le \ell)\Big] \\
&\qquad\qquad = \mathrm{E}\Big[(Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z_{ni}) \mathbf{1}(Z_{ni} \le \ell)\Big] + o(1)
\end{aligned}$$

we finally obtain that

$$\begin{aligned}
&\mathrm{Cov}\Big((Z_{nj} - \theta^{-1}) \textstyle\sum_{s \in I_j} f(U_s, Z^*_{ni}) \mathbf{1}(Z^*_{ni} \le \ell), \\
&\qquad\qquad (Z_{nj'} - \theta^{-1}) \textstyle\sum_{s' \in I_{j'}} f(U_{s'}, Z^*_{ni'}) \mathbf{1}(Z^*_{ni'} \le \ell)\Big) = o(1)
\end{aligned}$$

All other cases can be treated similarly by a successive application of Berbee's coupling Lemma. Also, C_{n3} can be treated similarly. □

Acknowledgments

The authors would like to thank Johan Segers and Gregor Weiß for providing us with R-implementations of several estimators for the extremal index and with the financial market data, respectively. This research has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation, which is gratefully acknowledged. Parts of this paper were written when A. Bücher was a visiting professor at TU Dortmund University.

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*, Volume 112 of *Mathematical Centre Tracts*. Amsterdam: Mathematisch Centrum.

Berghaus, B. and A. Bücher (2016). Goodness-of-fit tests for multivariate copula-based time series models. *Econometric Theory FirstView*, 1–39.

Billingsley, P. (1979). *Probability and measure*. John Wiley & Sons, New York-Chichester-Brisbane. Wiley Series in Probability and Mathematical Statistics.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J. 30*(1), 69–81.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv. 2*, 107–144. Update of, and a supplement to, the 1986 original.

Bücher, A. and J. Segers (2015). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *ArXiv e-prints*.

Darsow, W. F., B. Nguyen, and E. T. Olsen (1992). Copulas and Markov processes. *Illinois J. Math. 36*(4), 600–642.

de Haan, L., S. I. Resnick, H. Rootzén, and C. G. de Vries (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Process. Appl. 32*(2), 213–224.

Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pp. 3–113. Boston, MA: Birkhäuser Boston.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab. 10*(4), 1274–1301.

Drees, H. (2002). Tail empirical processes under mixing conditions. In *Empirical process techniques for dependent data*, pp. 325–342. Birkhäuser Boston, Boston, MA.

Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist. 38*(4), 2145–2186.

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol. 65*(2), 545–556.

Hsing, T. (1984). *Point Processes Associated with Extreme Value Theory*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–The University of North Carolina at Chapel Hill.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist. 21*(4), 2043–2071.

Hsing, T., J. Hüsler, and M. R. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields 78*(1), 97–112.

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

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete 65*(2), 291–306.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Leadbetter, M. R. and H. Rootzén (1988). Extremal theory for stochastic processes. *Ann. Probab. 16*(2), 431–478.

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

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab. 15*(1), 281–291.

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab. 4*(2), 529–548.

Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semiparametric models for multivariate time series. *J. Multivariate Anal. 110*, 30–42.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, Volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York.

Robert, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist. 37*(1), 271–310.

Robert, C. Y., J. Segers, and C. A. T. Ferro (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*(2), 468–490.

Segers, J. (2005). Approximate distributions of clusters of extremes. *Statist. Probab. Lett. 74*(4), 330–336.

Smith, R. L. and I. Weissman (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B 56*(3), 515–528.

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

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge: Cambridge University Press.

Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab. 11*(4), 750–783.

Weissman, I. and S. Y. Novak (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inference 66*(2), 281–288.

SUPPLEMENTARY MATERIAL ON "WEAK CONVERGENCE OF A PSEUDO MAXIMUM LIKELIHOOD ESTIMATOR FOR THE EXTREMAL INDEX"

BETINA BERGHAUS AND AXEL BÜCHER

Abstract. This supplementary material contains the remaining lemmas needed for the proof of the sliding blocks version of Theorem 3.1 (Section A) and the proofs of Lemmas 4.1 and 6.1 from the main paper (Section B).

Appendix A. Auxiliary Lemmas for the proof of Theorem 3.1 – sliding blocks

Lemma A.1 (Approximation by an integral with bounded support – sliding blocks)**.** *Under Condition 2.1, for all $\delta > 0$,*

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

Proof. The proof is similar to the proof of Lemma 9.1, whence we only give a sketch proof. For some $\varepsilon \in (c_1, c_2)$ let $A'_n = A'_n(\varepsilon)$ denote the event $\{\min_{t=1}^{n-b_n+1} N_{nt} > 1-\varepsilon\}$. Note that $\mathbb{P}(A'_n) \to 0$ by Condition 2.1(v). Recalling the definition of f from the beginning of the proof of Lemma 9.1, we may then write $D^{\mathrm{sl}}_n - D^{\mathrm{sl}}_{n,\ell} = R^{\mathrm{sl}}_{n,\ell}\,\mathbf{1}_{A'_n} + o_{\mathbb{P}}(1)$, where

$$R^{\mathrm{sl}}_{n,\ell} = k_n^{-3/2} \sum_{i=1}^{k_n-1} \sum_{j=1}^{k_n} \sum_{s\in I_j} b_n^{-1} \sum_{t\in I_i} f(U_s, Z^{\mathrm{sl}}_{nt})\,\mathbf{1}(Z^{\mathrm{sl}}_{nt} \ge \ell).$$

Now, decompose $R^{\mathrm{sl}}_{n,\ell} = R^{\mathrm{sl}}_{n,\ell,2} + R^{\mathrm{sl}}_{n,\ell,3}$ according to whether the second sum over j is such that $|j-i| \le 2$ or $|j-i| \ge 3$, respectively. Similar as in the proof of Lemma 9.1, it can be shown that $R^{\mathrm{sl}}_{n,\ell,2}\,\mathbf{1}_{A'_n} = o_{\mathbb{P}}(1)$ and that $\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{P}(|R^{\mathrm{sl}}_{n,\ell,3}\,\mathbf{1}_{A'_n}| > \delta) = 0$. □

Lemma A.2 (Approximation by a Lebesgue integral – sliding blocks)**.** *Suppose Condition 2.1 is met. Then, as $n \to \infty$,*

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

Proof. As in the proof of Lemma 9.2 the result follows if we can show that $\mathrm{Var}\{\hat H^{\mathrm{sl}}_{k_n}(x)\} = o(1)$ for any $x \in [0,\ell]$. This in turn follows from similar arguments as in the proof of Proposition 3.1 in Robert et al. (2009). □

Lemma A.3 (Joint convergence of fidis – sliding blocks)**.** *Let*

$$G^{\mathrm{sl}}_n = \sqrt{k_n}(T^{\mathrm{sl}}_n - \mathrm{E}\,T^{\mathrm{sl}}_n), \qquad T^{\mathrm{sl}}_n = \frac{1}{n-b_n+1} \sum_{t=1}^{n-b_n+1} Z^{\mathrm{sl}}_{nt}.$$

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

$$\Big(e_n(x_1), \dots, e_n(x_m), G^{\mathrm{sl}}_n\Big)' \rightsquigarrow \Big(e(x_1), \dots, e(x_m), G^{\mathrm{sl}}\Big)',$$

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

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

where r *and* h *are defined in Lemma 9.3.*

Proof. For notational convenience, we will only show the joint weak convergence of $(e_n(x), G_n^{\mathrm{sl}})$ for some fixed $x>0$; the general case can be shown analogously. Let $A_n' = \{\min_{t=1}^{n-b_n+1} N_{nt}^{\mathrm{sl}} > 1-\varepsilon\}$, where $\varepsilon\in(0,c_1\wedge c_2)$ and note that $\mathbb{P}(A_n')\to 1$ as $n\to\infty$. Due to the Cramér-Wold device it suffices to prove that, for any $\lambda_1,\lambda_2\in\mathbb{R}$,

$$\{\lambda_1 e_n(x)+\lambda_2 G_n^{\mathrm{sl}}\}\,\mathbf{1}_{A_n'} \rightsquigarrow \lambda_1 e(x)+\lambda_2 G^{\mathrm{sl}}.$$

We may write

$$\begin{aligned}\lambda_1 e_n(x)+\lambda_2 G_n^{\mathrm{sl}} &= \tfrac{\lambda_1}{k_n^{1/2}}\sum_{s=1}^{n}\{\mathbf{1}(U_s>1-\tfrac{x}{b_n})-\tfrac{x}{b_n}\} + \tfrac{\lambda_2 k_n^{1/2}}{n-b_n+1}\sum_{s=1}^{n-b_n+1}\{Z_{ns}^{\mathrm{sl}}-\mathrm{E}[Z_{n1}^{\mathrm{sl}}]\} \\ &= \sum_{j=1}^{k_n-1}\sum_{s\in I_j}\Big[\tfrac{\lambda_1}{k_n^{1/2}}\{\mathbf{1}(U_s>1-\tfrac{x}{b_n})-\tfrac{x}{b_n}\} + \tfrac{\lambda_2 k_n^{1/2}}{n-b_n+1}\{Z_{ns}^{\mathrm{sl}}-\mathrm{E}[Z_{n1}^{\mathrm{sl}}]\}\Big] + o_{\mathbb{P}}(1),\end{aligned}$$

where the $o_{\mathbb{P}}$ is due to omitting summands from the last block. Choose some integer sequence $k_n^*<k_n$ such that $k_n^*\to\infty$ and $k_n^*=o(k_n^{\delta/\{2(1+\delta)\}})$ as $n\to\infty$, where δ is defined in Condition 2.1(ii). Moreover, set $q_n^*=\lfloor k_n/(k_n^*+2)\rfloor$. For $j=1,\dots,q_n^*$, define

$$J_j^+ = \textstyle\bigcup_{i=(j-1)(k_n^*+2)+1}^{j(k_n^*+2)-2} I_i \qquad\text{and}\qquad J_j^- = I_{j(k_n^*+2)-1}\cup I_{j(k_n^*+2)},$$

i.e., we combine k_n^* consecutive I_i-blocks in one big block J_j^+ of size $k_n^* b_n$ and each of the big blocks is separated by a small block J_j^- of size $2b_n$, formed by merging two consecutive I_i-blocks. With this notation we obtain

$$\lambda_1 e_n(x)+\lambda_2 G_n^{\mathrm{sl}} = H_n^+ + H_n^- + o_{\mathbb{P}}(1), \qquad H_n^{\pm} = \frac{1}{\sqrt{q_n^*}}\sum_{j=1}^{q_n^*} S_{nj}^{\pm},$$

where, for $j=1,\dots,q_n^*$,

$$S_{nj}^{\pm} = \sqrt{\frac{q_n^*}{k_n}}\sum\nolimits_{s\in J_j^{\pm}}\Big[\lambda_1\{\mathbf{1}(U_s>1-\tfrac{x}{b_n})-\tfrac{x}{b_n}\} + \frac{\lambda_2 n}{n-b_n+1}\frac{1}{b_n}\{Z_{ns}^{\mathrm{sl}}-\mathrm{E}[Z_{n1}^{\mathrm{sl}}]\}\Big].$$

First, we will show that $H_n^-\,\mathbf{1}_{A_n'}=o_{\mathbb{P}}(1)$. As in the proof of Lemma 9.3 we have $H_n^-\,\mathbf{1}_{A_n'}=\tilde H_n^-\,\mathbf{1}_{A_n'}+o_{\mathbb{P}}(1)=\tilde H_n^- + o_{\mathbb{P}}(1)$, where $\tilde H_n^-$ is defined exactly as H_n^-, but with S_{nj}^- replaced by

$$S_{nj}^{\varepsilon -} = \sqrt{\frac{q_n^*}{k_n}}\sum\nolimits_{s\in J_j^{\pm}}\Big[\lambda_1\{\mathbf{1}(U_s^{\varepsilon}>1-\tfrac{x}{b_n})-\tfrac{x}{b_n}\} + \frac{\lambda_2 n}{n-b_n+1}\frac{1}{b_n}\{Z_{ns}^{\varepsilon,\mathrm{sl}}-\mathrm{E}[Z_{n1}^{\mathrm{sl}}]\}\Big],$$

with $Z_{ns}^{\varepsilon,\mathrm{sl}} = b_n(1 - \max_{s=t}^{t+b_n-1} U_s^\varepsilon)$. By an inequality similar to (9.4) and the argumentation subsequent to that inequality, it suffices to show that $\|S_{n1}^{\varepsilon-}\|_p = o(1)$ for some $p \in (2, 2+\delta)$ and that $\sum_{j=2}^{q_n^*} |\operatorname{Cov}(S_{nj}^{\varepsilon-}, S_{n,1+j}^{\varepsilon-})| = o(1)$. The first assertion follows from

$$\|S_{1j}^{\varepsilon-}\|_p \le 2\sqrt{\frac{q_n^*}{k_n}}\Big\{\lambda_1\|N_{b_n k_n^*}^{(x)}(E)\|_p + \lambda_2\|Z_{n1}^{\varepsilon,\mathrm{sl}} - \mathrm{E}[Z_{n1}^{\mathrm{sl}}]\|_p\Big\} = O(1/\sqrt{k_n^*}) = o(1),$$

by Condition 2.1(ii) and (vi) and the definition of q_n^*. For the second assertion, note that $S_{nj}^{\varepsilon-}$ is $\mathcal{B}^\varepsilon_{\{(jk_n^*+2j-2)b_n+1\}:\{j(k_n^*+2)b_n\}}$-measurable, whence

$$|\operatorname{Cov}(S_{nj}^{\varepsilon-}, S_{n,1+j}^{\varepsilon-})| \le 10\|S_{n1}^{\varepsilon-}\|_p^2.\alpha_{c_2}(jk_n^*b_n)^{1-2/p}$$

By Condition 2.1(iii) the sum $\sum_{j=2}^{q_n^*}\alpha_{c_2}(jk_n^*b_n)^{1-2/p}$ converges to 0, which implies the assertion.

It remains to be shown $H_n^+\mathbf{1}_{A_n'}$ converges to a normal distribution with the claimed covariance. As in the proof of Lemma 9.3, we can write

$$H_n^+\mathbf{1}_{A_n'} = \frac{1}{\sqrt{q_n^*}}\sum_{j=1}^{q_n^*}\tilde S_{nj}^+ + o_{\mathbb{P}}(1), \qquad \tilde S_{nj}^+ = S_{nj}^+\mathbf{1}(\max_{s\in J_j^+} Z_{ns}^{\mathrm{sl}} < \varepsilon b_n).$$

For $i \ne j$, the observations $\tilde S_{nj}^+$ and $\tilde S_{ni}^+$ are separated by at least one block of size b_n and measurable with respect to the $\mathcal{B}^\varepsilon_{\cdot:\cdot}$-sigma fields. Further, by Condition 2.1(iii), $q_n^*\alpha_{c_2}(b_n) \le k_n\alpha_{c_2}(b_n) = o(1)$. A standard argument for the characteristic function then shows that the weak limit of $(q_n^*)^{-1/2}\sum_{j=1}^{q_n^*}\tilde S_{nj}^+$ is the same as if the sample $(\tilde S_{nj}^+)_{j=1,\dots,q_n^*}$ was independent, which we will assume subsequently. By arguments as before, we can then pass back to an independent sample $(S_{nj}^+)_{j=1,\dots,q_n^*}$, and weak convergence follows from the classical central limit theorem for rowwise iid triangular arrays.

By Condition 2.1(ii) and (vi) and Minkowski's inequality, we have $\mathrm{E}[|S_{nj}^+|^{2+\delta}] = O(k_n^{*\,(2+\delta)/2})$. Hence,

$$\frac{\sum_{j=1}^{q_n^*}\mathrm{E}[|S_{nj}^+|^{2+\delta}]}{\big(\sum_{j=1}^{q_n^*}\mathrm{E}[||S_{nj}^+|^2]\big)^{\frac{2+\delta}{2}}} = q_n^{*\,-\delta/2}\frac{\mathrm{E}[|S_{nj}^+|^{2+\delta}]}{\mathrm{E}[|S_{nj}^+|^2]^{\frac{2+\delta}{2}}} = O(k_n^{-\delta/2}k_n^{*\,1+\delta})0 = o(k_n^{-\delta/2+\delta/2}) = o(1),$$

by the definition of k_n^*, provided that $\lim_{n\to\infty}\mathrm{E}[(S_{n1}^+)^2]$ exists (which we will show below). Therefore, Ljapunov's condition is satisfied and $\lambda_1 e_n(x) + \lambda_2 G_n^{\mathrm{sl}}$ converges weakly to a normal distribution with variance $\lim_{n\to\infty}\mathrm{E}[(S_{n1}^+)^2]$. Hence, it remains to be shown that

$$\lim_{n\to\infty}\mathrm{E}[(S_{n1}^+)^2] = \lambda_1^2 r(x,x) + 2\lambda_1\lambda_2 h(x) + \lambda_2^2\tfrac{2(\log(4)-1)}{\theta^2}$$

and this in turn follows from the proof of Theorem 4.1 in Robert (2009) (for the first summand in the latter display) and Lemma A.4, A.5 and A.6 below (note that, with $n^* = k_n^* b_n$, we can write $S_{n1}^+ = \lambda_1 e_{n^*} + \lambda_2 G_{n^*}^{\mathrm{sl}} + o_{\mathbb{P}}(1)$ and that all assumptions in Condition 2.1 are satisfied if n and k_n are replaced by n^* and k_n^*). □

Lemma A.4. *Suppose Conditions 2.1(ii), (iii) and (vi) are met. Then, for any $x \in [0,\infty)$, as $n \to \infty$,*

$$\operatorname{Cov}(e_n(x), G_n^{\mathrm{sl}}) \to h_{sl}(x),$$

where $h_{\mathrm{sl}}(0) = 0$ *and, for* $x \neq 0$,

$$h_{\mathrm{sl}}(x) = \frac{2}{\theta}\Big[\sum_{i=1}^{\infty} i \int_0^1 \Big\{\theta \int_0^x \sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) e^{-\theta \xi y}\,\mathrm{d}y + p^{(\xi x)}(i) e^{-\theta x}\Big\}\,\mathrm{d}\xi - x\Big],$$

where p_2 *is defined in Lemma 9.3 and where, for* $x > 0$,

$$p^{(x)}(i) = \mathbb{P}\big(N_E^{(x)} = i\big), \quad N_E^{(x)} = \sum_{i=1}^{\eta} \xi_i$$

with $\eta \sim$ *Poisson*(θx) *independent of iid random variables* $\xi_i \sim \pi, i \in \mathbb{N}$.

Proof. For the sake of a clear exposition, we will assume that both U_s and Z_{nt}^{sl} are measurable with respect to the $\mathcal{B}^{\varepsilon}_{:,:}$-sigma fields; the general case follows by multiplication with suitable indicator functions as in the previous proofs. Introduce the notation $A_j = \sum_{s \in I_j} \mathbf{1}(U_s > 1 - x/b_n)$ and $B_j = \sum_{s \in I_j} Z_{nt}^{\mathrm{sl}}$. We can write

$$\mathrm{Cov}(e_n(x), G_n^{\mathrm{sl}}) = \frac{1}{n - b_n + 1} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n - 1} \mathrm{Cov}(A_i, B_j) + \frac{1}{n - b_n + 1} \sum_{i=1}^{k_n} \mathrm{Cov}(A_i, Z_{n,n-b_n+1}^{\mathrm{sl}}).$$

The second sum on the right hand-side is negligible, since both $\|A_j\|_2 = \|N_{b_n}^{(x)}(E)\|_2 = O(1)$ and $\|Z_{n,n-b_n+1}^{\mathrm{sl}}\|_2 = O(1)$ by Condition 2.1(ii) and (vi). Regarding the first sum, by stationarity, we can write

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n - 1} \mathrm{Cov}(A_i, B_j) \\
&\quad = \frac{1}{n} \sum_{i=1}^{k_n - 1} \sum_{j=1}^{k_n - 1} \mathrm{Cov}(A_i, B_j) + O(b_n/n) \\
&\quad = \frac{k_n - 1}{n} \mathrm{Cov}(A_1, B_1) + \sum_{h=2}^{k_n - 1} \frac{k_n - h}{n} \big\{\mathrm{Cov}(A_1, B_h) + \mathrm{Cov}(A_h, B_1)\big\} + o(1).
\end{aligned}$$

Split the right-hand side according to whether $\mathrm{Cov}(A_i, B_j)$ is such that either $i - j \in \{0, 1\}$, or $i - j \in \{-1, 2\}$ or $i - j \in \{-k_n + 2, \dots, k_n - 2\} \setminus \{-1, 0, 1, 2\}$. Up to negligible terms, this allows to write the right-hand side of the previous display as $R_{n1} + R_{n2} + R_{n3}$, where $R_{n1} = b_n^{-1} \mathrm{Cov}(A_2, B_1 + B_2)$, $R_{n2} = b_n^{-1} \mathrm{Cov}(A_3, B_1 + B_4)$ and

$$R_{n3} = \sum_{h=3}^{k_n - 1} \frac{k_n - h}{n} \mathrm{Cov}(A_1, B_h) + \sum_{h=4}^{k_n - 1} \frac{k_n - h}{n} \mathrm{Cov}(A_h, B_1).$$

Both sums in R_{n3} converge to 0: first, $\|A_j\|_{2+\delta} = O(1)$ and $\|B_j\|_{2+\delta} = O(b_n)$. Second, the variables defining A_1 and B_h are at least $(h-1)b_n$-observations apart, while the variables defining A_1 and B_h are at least $(h-2)b_n$-observations apart. As a consequence, by Lemma 3.11 in Dehling and Philipp (2002),

$$|R_{n3}| \le C \sum_{h=1}^{k_n} \alpha_{c_2}^{\delta/(2+\delta)}(h b_n) \le C b_n^{-\eta} \sum_{h=1}^{\infty} h^{-\eta} = o(1).$$

The term R_{n2} is also negligible: we have

$$b_n^{-1}\operatorname{Cov}(A_3,B_4)=b_n^{-1}\operatorname{Cov}(A_1,B_2)=b_n^{-1}\sum_{t=b_n+1}^{2b_n}\operatorname{Cov}\{\textstyle\sum_{s=1}^{b_n}\mathbf{1}(U_s>1-x/b_n),Z_{nt}^{\mathrm{sl}}\}.$$

The covariance on the right-hand side can be bounded by a multiple of $\alpha_{c_2}(t-b_n)^{\delta/(2+\delta)}$. The remaining sum over the mixing-coefficients converges, such that $b_n^{-1}\operatorname{Cov}(A_3,B_4)=O(b_n^{-1})$. The covariance $b_n^{-1}\operatorname{Cov}(A_3,B_1)$ can be treated similarly.

It remains to be shown that

$$R_{n1}=\frac{1}{b_n}\operatorname{Cov}(A_2,B_1+B_2)=\frac{1}{b_n}\sum_{t=1}^{2b_n}\operatorname{Cov}\Big\{\sum_{s\in I_2}\mathbf{1}(U_s>1-\tfrac{x}{b_n}),Z_{nt}^{\mathrm{sl}}\Big\}$$

converges to $h_{\mathrm{sl}}(x)$. To this end, define functions $f_n,g_n:[0,1]\to\mathbb{R}$ by

$$f_n(\xi)=\sum_{t=1}^{b_n}\mathrm{E}\Big[\sum_{s\in I_2}\mathbf{1}(U_s>1-\tfrac{x}{b_n})Z_{nt}^{\mathrm{sl}}\Big]\mathbf{1}\{\xi\in[\tfrac{t-1}{b_n},\tfrac{t}{b_n})\},$$

$$g_n(\xi)=\sum_{t=b_n+1}^{2b_n}\mathrm{E}\Big[\sum_{s\in I_2}\mathbf{1}(U_s>1-\tfrac{x}{b_n})Z_{nt}^{\mathrm{sl}}\Big]\mathbf{1}\{\xi\in[\tfrac{t-b_n-1}{b_n},\tfrac{t-b_n}{b_n})\}.$$

With this notation, we obtain

$$\operatorname{Cov}(e_n(x),G_n^{\mathrm{sl}})=\int_0^1\{f_n(\xi)+g_n(\xi)\}\,\mathrm{d}\xi-2x\,\mathrm{E}[Z_{n1}^{\mathrm{sl}}]+o(1).$$

By uniform integrability of Z_{n1}^{sl} we have $\mathrm{E}[Z_{n1}^{\mathrm{sl}}]\to\theta^{-1}$, as $n\to\infty$. Furthermore, for any n, f_n and g_n are uniformly bounded by $\|\sum_{s\in I_1}\mathbf{1}(U_s>1-\frac{x}{b_n})\|_2\times\|Z_{n1}^{\mathrm{sl}}\|_2$, which again is uniformly bounded in n by Condition 2.1(ii) and (vi), i.e., $\sup_n(\|f_n\|_\infty+\|g_n\|_\infty)<\infty$. Hence, by dominated convergence, the lemma follows if we show that, for any $\xi\in(0,1)$,

$$\begin{aligned}\lim_{n\to\infty}f_n(1-\xi)&=\lim_{n\to\infty}g_n(\xi)\\&=\sum_{i=1}^{\infty}i\int_0^x\sum_{l=0}^{i}p^{(\xi x)}(l)p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0)e^{-\theta\xi y}\,\mathrm{d}y+\theta^{-1}p^{(\xi x)}(i)e^{-\theta x}.\end{aligned}\tag{A.1}$$

We only do this for g_n, as f_n can be treated similarly. Fix $\xi\in(0,1)$ and note that

$$g_n(\xi)=\mathrm{E}\Big[\sum_{s\in I_2}\mathbf{1}(U_s>1-\tfrac{x}{b_n})Z_{n,(\lfloor(1+\xi)b_n\rfloor+1)}^{\mathrm{sl}}\Big].$$

Let us first show joint weak convergence of the two variables inside this expectation, and for that purpose consider

$$\begin{aligned}F_n(i,y)&:=\mathbb{P}\Big(\textstyle\sum_{s=b_n+1}^{2b_n}\mathbf{1}(U_s>1-\frac{x}{b_n})=i,Z_{n,(\lfloor(1+\xi)b_n\rfloor+1)}^{\mathrm{sl}}\ge y\Big)\\&=\mathbb{P}\Big(\textstyle\sum_{s=b_n+1}^{2b_n}\mathbf{1}(U_s>1-\frac{x}{b_n})=i,\sum_{s=\lfloor(1+\xi)b_n\rfloor+1}^{\lfloor(1+\xi)b_n\rfloor+b_n}\mathbf{1}(U_s>1-\frac{y}{b_n})=0\Big)\end{aligned}$$

For $y \in (0, x]$, we can write $F_n(i, y) = \sum_{l=0}^{i} A_n(l, i)$, where

$$A_n(l,i) = \mathbb{P}\Big(\sum_{s=b_n+1}^{\lfloor (1+\xi)b_n \rfloor} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = l, \sum_{s=\lfloor (1+\xi)b_n \rfloor+1}^{2b_n} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = i - l, \\ \sum_{s=\lfloor (1+\xi)b_n \rfloor+1}^{2b_n} \mathbf{1}(U_s > 1 - \tfrac{y}{b_n}) = 0, \sum_{s=2b_n+1}^{\lfloor (2+\xi)b_n \rfloor} \mathbf{1}(U_s > 1 - \tfrac{y}{b_n}) = 0 \Big).$$

Let us show that we can manipulate any sum inside this probability by adding or subtracting r_n summands, where r_n is some integer sequence with $r_n = o(b_n)$. Indeed, for any fixed $x > 0$ and sufficiently large n:

$$\mathbb{P}\big(\sum_{s=1}^{r_n} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = 0 \big) \geq 1 - r_n \mathbb{P}(U_1 > 1 - \tfrac{x}{b_n}) = 1 - \tfrac{x r_n}{b_n} \to 1, \qquad n \to \infty.$$

Now, by omitting the last r_n summands of the first sum inside the probability defining $A_n(l,i)$, this sum becomes asymptotically independent of the remaining sums in the probability (at the cost of an additive $\alpha_{c_2}(r_n)$-error). The same can be done for the last sum and we obtain

$$A_n(l,i) = \mathbb{P}\Big(\sum_{s=b_n+1}^{\lfloor (1+\xi)b_n \rfloor} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = l \Big) \times \mathbb{P}\Big(\sum_{s=2b_n+1}^{\lfloor (2+\xi)b_n \rfloor} \mathbf{1}(U_s > 1 - \tfrac{y}{b_n}) = 0 \Big) \\ \times \mathbb{P}\Big(\sum_{s=\lfloor (1+\xi)b_n \rfloor+1}^{2b_n} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = i - l, \sum_{s=\lfloor (1+\xi)b_n \rfloor+1}^{2b_n} \mathbf{1}(U_s > 1 - \tfrac{y}{b_n}) = 0 \Big) \\ + O(\alpha_{c_2}(r_n)) + O(r_n/b_n).$$

This expression converges to $p^{(\xi x)}(l) p^{(\xi y)}(0) p_2^{((1-\xi)x,(1-\xi)y)}(i-l, 0)$ by Theorem 4.1 in Robert (2009). As a consequence,

$$F_n(i,y) \to \sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) p^{(\xi y)}(0).$$

In the case $y > x$ similar arguments imply that

$$F_n(i,y) \to p^{(\xi x)}(i) p^{(y)}(0) = p^{(\xi x)}(i) e^{-\theta y}.$$

Since both $\sum_{s \in I_2} \mathbf{1}(U_s > 1 - \frac{x}{b_n})$ and $Z^{\mathrm{sl}}_{n(\lfloor (1+\xi)b_n \rfloor+1)}$ are in $L_{2+\delta}(\mathbb{P})$, weak convergence implies convergence of moments, whence

$$g_n(\xi) = \sum_{i=1}^{\infty} i \int_0^{\infty} \mathbb{P}\Big(\sum_{s=b_n+1}^{2b_n} \mathbf{1}(U_s > 1 - \tfrac{x}{b_n}) = i, Z^{\mathrm{sl}}_{n(\lfloor (1+\xi)b_n \rfloor+1)} \geq y \Big) \mathrm{d}y \\ \to \sum_{i=1}^{\infty} i \int_0^{x} \sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) e^{-\theta \xi y} \, \mathrm{d}y + \int_x^{\infty} p^{(\xi x)}(i) e^{-\theta y} \, \mathrm{d}y.$$

Calculating the integral on the right-hand side explicitly yields (A.1). □

Lemma A.5. *Suppose Conditions 2.1(iii) and (vi) are met, then, as $n \to \infty$,*

$$\mathrm{Var}(G_n^{\mathrm{sl}}) \to \frac{2(\log(4) - 1)}{\theta^2}.$$

Proof. As in proof of Lemma A.4 we will assume that the Z^{sl}_{nt} are measurable with respect to the $\mathcal{B}^{\varepsilon}_{:,:}$-sigma fields. Similar as in the beginning of the proof of Lemma A.4, one can show that

$$\operatorname{Var}(G_n^{\mathrm{sl}}) = \frac{2}{b_n}\sum_{t=1}^{b_n} \operatorname{Cov}(Z_{n1}^{\mathrm{sl}} Z_{n,(1+t)}^{\mathrm{sl}}) + o(1) = 2\int_0^1 h_n(\xi)\, \mathrm{d}\xi - 2\,\mathrm{E}[Z_{n1}^{\mathrm{sl}}]^2 + o(1),$$

where $h_n : [0,1] \to \mathbb{R}$ is defined as

$$h_n(\xi) = \sum_{t=1}^{b_n} \mathrm{E}[Z_{n1}^{\mathrm{sl}} Z_{n,(1+t)}^{\mathrm{sl}}]\, \mathbf{1}\{\xi \in [\tfrac{t-1}{b_n}, \tfrac{t}{b_n})\} = \mathrm{E}[Z_{n1}^{\mathrm{sl}} Z_{n,(\lfloor b_n\xi\rfloor+1)}^{\mathrm{sl}}].$$

Condition 2.1(vi) implies $\mathrm{E}[Z_{n1}^{\mathrm{sl}}] \to \theta^{-1}$. The limit of the integral over h_n can deduced from pointwise convergence and the dominated convergence theorem. To see this, note that $\sup_n \|h_n\|_\infty \le \sup_n \mathrm{E}[{Z_{n1}^{\mathrm{sl}}}^2] < \infty$, due to Condition 2.1(vi). Regarding the pointwise convergence, suppose we have shown that, for any $\xi \in (0,1)$, there exists some random vector $(X^{(\xi)}, Y^{(\xi)})$ with dirtybution function depending on ξ, such that

$$(Z_{n1}^{\mathrm{sl}}, Z_{n,(\lfloor b_n\xi\rfloor+1)}^{\mathrm{sl}}) \rightsquigarrow (X^{(\xi)}, Y^{(\xi)}). \tag{A.2}$$

In that case, $h_n(\xi) = \mathrm{E}[Z_{n1}^{\mathrm{sl}} Z_{n,(\lfloor b_n\xi\rfloor+1)}^{\mathrm{sl}}]$ converges to $\mathrm{E}[X^{(\xi)}Y^{(\xi)}]$ by Condition 2.1(vi). Let us show (A.2). Fix $x, y \in \mathbb{R}^+$ and write

$$\begin{aligned}\bar F_n(x,y) &= \mathbb{P}(Z_{n1}^{\mathrm{sl}} > x, Z_{n,(\lfloor b_n\xi\rfloor+1)}^{\mathrm{sl}} > y) \\ &= \mathbb{P}(N_{1:\lfloor b_n\xi\rfloor} < 1 - \tfrac{x}{b_n}, N_{(\lfloor b_n\xi\rfloor+1):b_n} < 1 - \tfrac{x\vee y}{b_n}, N_{(b_n+1):\lfloor b_n(\xi+1)\rfloor} < 1 - \tfrac{y}{b_n}).\end{aligned}$$

Now, if r_n is an integer sequence such that $r_n = o(b_n)$, then, for sufficiently large n,

$$\mathbb{P}(N_{1:r_n} > 1 - \tfrac{x}{b_n}) \le \tfrac{x r_n}{b_n} \to 0, \qquad n \to \infty,$$

which is why we can omit or add r_n observations in the maximum without changing the limit of its distribution. Similar as in the proof of Lemma A.4 this gives

$$\begin{aligned}\bar F_n(x,y) = \mathbb{P}(N_{1:\lfloor b_n\xi\rfloor} < 1 - \tfrac{x}{b_n}) &\times \mathbb{P}(N_{(\lfloor b_n\xi\rfloor+1):b_n} < 1 - \tfrac{x\vee y}{b_n}) \\ &\times \mathbb{P}(N_{(b_n+1):\lfloor b_n(\xi+1)\rfloor} < 1 - \tfrac{y}{b_n}) + O(\alpha_{c_2}(r_n)) + O(\tfrac{r_n x\vee y}{b_n}),\end{aligned}$$

which, by (1.3), converges to

$$\bar F_\xi(x,y) = \exp(-\theta\xi x)\exp(-\theta(1-\xi)(x\vee y))\exp(-\theta\xi y) = \exp\{-\theta(\xi(x\wedge y) + x\vee y)\}.$$

This implies (A.2), with $(X^{(\xi)}, Y^{(\xi)})$ being defined by its joint survival function $\bar F_\xi : [0,\infty)^2 \to [0,1]$. Now, it is easy to see that

$$\lim_{n\to\infty} h_n(\xi) = \mathrm{E}[X^{(\xi)}Y^{(\xi)}] = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \bar F_\xi(x,y)\mathrm{d}x\mathrm{d}y = \frac{2}{\theta^2(1+\xi)}.$$

Finally, putting everything together, we obtain

$$\lim_{n\to\infty} \operatorname{Var}(G_n^{\mathrm{sl}}) = 2\int_0^1 \lim_{n\to\infty} h_n(\xi)\mathrm{d}\xi - \frac{2}{\theta^2} = \frac{2}{\theta^2}\Big(\int_0^1 \frac{2}{1+\xi}\mathrm{d}\xi - 1\Big) = \frac{2\{\log(4) - 1\}}{\theta^2}$$

as asserted. □

Lemma A.6. *Under the above conditions, $h_{\mathrm{sl}} = h$, where h_{sl} and h are defined in Lemma A.4 and Lemma 9.3, respectively.*

Proof. By the definition of $p^{(x)}$ and $p_2^{(x,y)}$ in Lemma A.4 and Lemma 9.3 we obtain that

$$\sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) = \mathbb{P}\Big(\sum_{j=1}^{\eta_1}\zeta_j + \sum_{j=1}^{\eta_2}\zeta_{j1}^{(y/x)} = i, \sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big),$$

with independent random variables $\eta_1 \sim \text{Poisson}(\xi\theta x)$, $\eta_2 \sim \text{Poisson}((1-\xi)\theta x)$, $\zeta_i \sim \pi, i \in \mathbb{N}$, and $(\zeta_{i1}^{(y/x)}, \zeta_{i2}^{(y/x)}) \sim \pi_2^{(y/x)}, i \in \mathbb{N}$. For this reason, we can write

$$\begin{aligned}\sum_{i=1}^{\infty} i \sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) &= \mathrm{E}\left[\Big\{\sum_{j=1}^{\eta_1}\zeta_j + \sum_{j=1}^{\eta_2}\zeta_{j1}^{(y/x)}\Big\}\mathbf{1}\Big(\sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big)\right] \\ &= \mathrm{E}\Big[\sum_{j=1}^{\eta_1}\zeta_j\Big]\mathbb{P}\Big(\sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big) + \mathrm{E}\left[\sum_{j=1}^{\eta_2}\zeta_{j1}^{(y/x)}\mathbf{1}\Big(\sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big)\right].\end{aligned}$$

By Wald's identity, we have $\mathrm{E}\big[\sum_{j=1}^{\eta_1}\zeta_j\big] = \xi x$. Independence of η_2 and $\zeta_{j2}^{(y/x)}, j \in \mathbb{N}$, further implies

$$\mathbb{P}\Big(\sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big) = \sum_{k=0}^{\infty}\mathbb{P}(\zeta_{12}^{(y/x)} = 0)^k\mathbb{P}(\eta_2 = k) = e^{-\theta(1-\xi)y},$$

where we used that $\mathbb{P}(\zeta_{12}^{(y/x)} = 0) = 1 - y/x$, see (9.9). Finally, (9.7) implies that

$$\mathrm{E}\left[\sum_{j=1}^{\eta_2}\zeta_{j1}^{(y/x)}\mathbf{1}\Big(\sum_{j=1}^{\eta_2}\zeta_{j2}^{(y/x)} = 0\Big)\right] = \mathrm{E}\big[\zeta_{11}^{(y/x)}\mathbf{1}(\zeta_{12}^{(y/x)} = 0)\big]\theta(1-\xi)xe^{-(1-\xi)\theta y}.$$

Altogether, we obtain

$$\begin{aligned}&\sum_{i=1}^{\infty} i \sum_{l=0}^{i} p^{(\xi x)}(l) p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0) \\ &\qquad = \xi x e^{-\theta(1-\xi)y} + \mathrm{E}[\zeta_{11}^{(y/x)}\mathbf{1}(\zeta_{12}^{(y/x)} = 0)]\theta(1-\xi)xe^{-(1-\xi)\theta y}.\end{aligned}$$

Now, noting that $\sum_{i=1}^{\infty} ip^{(\xi x)}(i) = \mathrm{E}\big[\sum_{j=1}^{\eta_1}\zeta_j\big] = \xi x$, we can rewrite h_{sl} as follows

$$\begin{aligned}h_{\text{sl}}(x) &= \frac{2}{\theta}\left[\sum_{i=1}^{\infty} i \int_0^1 \left\{\theta\int_0^x \sum_{l=0}^{i} p^{(\xi x)}(l)p_2^{((1-\xi)x,(1-\xi)y)}(i-l,0)e^{-\theta\xi y}\,\mathrm{d}y + p^{(\xi x)}(i)e^{-\theta x}\right\}\mathrm{d}\xi - x\right] \\ &= 2\int_0^x\int_0^1 \xi x e^{-\theta y}\,\mathrm{d}\xi\,\mathrm{d}y + 2\int_0^x\int_0^1 \mathrm{E}[\zeta_{11}^{(y/x)}\mathbf{1}(\zeta_{12}^{(y/x)} = 0)]\theta(1-\xi)xe^{-\theta y}\,\mathrm{d}\xi\,\mathrm{d}y \\ &\qquad + \frac{2}{\theta}\int_0^1 \xi x e^{-\theta x}\,\mathrm{d}\xi - \frac{2x}{\theta} \\ &= \int_0^x xe^{-\theta y}\,\mathrm{d}y + \int_0^x \mathrm{E}[\zeta_{11}^{(y/x)}\mathbf{1}(\zeta_{12}^{(y/x)} = 0)]\theta x e^{-\theta y}\,\mathrm{d}y + \frac{x}{\theta}e^{-\theta x} - \frac{2x}{\theta} \\ &= \int_0^1 \mathrm{E}[\zeta_{11}^{(\sigma)}\mathbf{1}(\zeta_{12}^{(\sigma)} = 0)]\theta x^2 \exp^{-\theta\sigma x}\,\mathrm{d}\sigma - \frac{x}{\theta}.\end{aligned}$$

From (9.8) we finally obtain that $h_{\text{sl}}(x) = h(x)$. □

Appendix B. Additional proofs

Proof of Lemma 4.1. We begin with the disjoint blocks estimator and write $(\hat T_n, T_n) = (\hat T_n^{\mathrm{dj}}, T_n^{\mathrm{dj}})$. Recalling (3.6), we can write $k_n \operatorname{E}[\hat T_n - T_n] = S_{n1} + S_{n2} + S_{n3} + S_{n4}$, where

$$S_{n1} = \sum_{s=1}^{b_n} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n1}}{b_n}) - \tfrac{Z_{n1}}{b_n}]$$

$$S_{n2} = \frac{k_n - 1}{k_n} \sum_{s=1}^{b_n} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n}],$$

$$S_{n3} = \frac{k_n - 1}{k_n} \sum_{s=b_n+1}^{2b_n} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n1}}{b_n}) - \tfrac{Z_{n1}}{b_n}]$$

$$S_{n4} = \sum_{i=3}^{k_n} \frac{k_n - i + 1}{k_n} \Big\{ \sum_{s \in I_1} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{ni}}{b_n}) - \tfrac{Z_{ni}}{b_n}] + \sum_{s \in I_i} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n1}}{b_n}) - \tfrac{Z_{n1}}{b_n}] \Big\}.$$

Note that $S_{n1} = -\operatorname{E}[Z_{n1}] \to -\theta^{-1}$, as $n \to \infty$, by Condition 2.1 (vi). Hence, it remains to be shown that S_{n2}, S_{n3} and S_{n4} vanish as $n \to \infty$.

Consider S_{n2}. Choose some integer $l \in \mathbb{N}$ and let n be sufficiently large such that $b_n > l$. Write $S_{n2} = (k_n - 1)/k_n \{S_{n2}^+ + S_{n2}^-\}$, where

$$S_{n2}^+ = \sum_{s=1}^{b_n - l} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n}], \qquad S_{n2}^- = \sum_{s=b_n-l+1}^{b_n} \operatorname{E}[\mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n}].$$

The absolute value of S_{n2}^- can be bounded by

$$\frac{l}{b_n} \operatorname{E}[|Z_{n1}|] + l\, \mathbb{P}(\max_{s=1}^{l} U_s > \max_{s=l+1}^{l+b_n} U_s)$$

which goes to 0 as $n \to \infty$ for any fixed l by Condition 2.1 (vi) and similar reasons as in the proof of Lemma 9.3, see (9.5). For the treatment of S_{n2}^+ fix $q > 0$ such that $q < \lim_{n\to\infty} \|Z_{n1}\|_2 = \sqrt{2}/\theta$. Then, for sufficiently large n, we can use the coupling construction leading to (9.3) (with $X = U_s$ and $Y = Z_{n2}$) to find a random variable Z_{n2}^* that has the same distribution as Z_{n2}, is in dependent of U_s and satisfies

$$\mathbb{P}(|Z_{n2} - Z_{n2}^*| > q) \le 18(\|Z_{n2}\|_2/q)^{2/5} \alpha(\sigma(U_s), \sigma(U_{n2}))^{4/5}.$$

By a monotonicity argument, we have

$$\begin{aligned}
&\big| \operatorname{E}\big[\big\{ \mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n} \big\} \mathbf{1}(|Z_{n2} - Z_{n2}^*| \le q)\big] \big| \\
&\qquad \le \big| \operatorname{E}\big[\big\{ \mathbf{1}(U_s > 1 - \tfrac{Z_{n2}^* + q}{b_n}) - \tfrac{Z_{n2}^* + q}{b_n} \big\} \mathbf{1}(|Z_{n2} - Z_{n2}^*| \le q)\big] \big| \\
&\qquad\qquad + \big| \operatorname{E}\big[\big\{ \mathbf{1}(Us > 1 - \tfrac{Z_{n2}^* - q}{b_n}) - \tfrac{Z_{n2}^* - q}{b_n} \big\} \mathbf{1}(|Z_{n2} - Z_{n2}^*| \le q)\big] \big| + \frac{2q}{b_n}.
\end{aligned}$$

Furthermore, since Z_{n2}^* is independent of U_s,

$$\begin{aligned}
&\big| \operatorname{E}\big[\big\{ \mathbf{1}(U_s > 1 - \tfrac{Z_{n2}^* \pm q}{b_n}) - \tfrac{Z_{n2}^* \pm q}{b_n} \big\} \mathbf{1}(|Z_{n2} - Z_{n2}^*| \le q)\big] \big| \\
&\qquad = \big| \operatorname{E}\big[\big\{ \mathbf{1}(U_s > 1 - \tfrac{Z_{n2}^* \pm q}{b_n}) - \tfrac{Z_{n2}^* \pm q}{b_n} \big\} \mathbf{1}(|Z_{n2} - Z_{n2}^*| > q)\big] \big|.
\end{aligned}$$

Combining everything we obtain

$$|S_{n2}^{+}| \le \sum_{s=1}^{b_n-l} \big|\operatorname{E}\big[\big\{\mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n}\big\}\mathbf{1}(|Z_{n2} - Z_{n2}^{*}| \le q)\big]\big|$$
$$+ \sum_{s=1}^{b_n-l} \big|\operatorname{E}\big[\big\{\mathbf{1}(U_s > 1 - \tfrac{Z_{n2}}{b_n}) - \tfrac{Z_{n2}}{b_n}\big\}\mathbf{1}(|Z_{n2} - Z_{n2}^{*}| > q)\big]\big|$$
$$\le \frac{2q(b_n - l)}{b_n} + 54(\|Z_{n2}\|_2/q)^{2/5} \sum_{s=l+1}^{b_n} \alpha(s)^{4/5}.$$

As a consequence, since $\alpha(s) \le Cs^{-\eta} \le Cs^{-3}$ by Condition 2.1 (iii),

$$\limsup_{n\to\infty} |S_{n2}| \le 2q + 54C(\sqrt{2}/(\theta q))^{2/5} \sum_{s=l}^{\infty} s^{-12/5}$$

This bound in turn can be made arbitrarily small by first choosing q sufficiently small and then choosing l sufficiently large. Hence, $\lim_{n\to\infty} |S_{n2}| = 0$. Along the same lines, we obtain that $\lim_{n\to\infty} |S_{n3}| = 0$.

The term S_{n4} can also be treated by a coupling construction. Here, we choose $q = q_n = k_n^{-1-\varepsilon}$ for some $\varepsilon \in (0, 3/4)$. By similar arguments as before, we obtain that

$$|S_{n4}| \le 2\sum_{i=3}^{k_n} \Big\{2q_n + 54(\|Z_{n1}\|_2/q_n)^{2/5} b_n \alpha((i-2)b_n)^{4/5}\Big\}$$
$$\le 4k_n^{-\varepsilon} + 108 \cdot k_n^{2/5(1+\varepsilon)} b_n^{-7/5} \|Z_{n1}\|_2^{2/5} C \sum_{i=3}^{k_n} (i-2)^{-12/5}$$
$$= O((k_n/b_n^2)^{2/5(1+\varepsilon)} b_n^{-3/5+4/5\varepsilon}) = o(1),$$

by Condition 2.1 (iii) and by the choice of ε. The proof for the disjoint blocks estimator is finished.

Sliding Blocks. By the definition of $\hat{T}_n^{\mathrm{sl}}$ and T_n^{sl} we can write

$$k_n \operatorname{E}[\hat{T}_n^{\mathrm{sl}} - T_n^{\mathrm{sl}}] = S_{n1}^{\mathrm{sl}} + S_{n2}^{\mathrm{sl}} + S_{n3}^{\mathrm{sl}} + S_{n4}^{\mathrm{sl}} + S_{n5}^{\mathrm{sl}} + o(1),$$

as $n \to \infty$, where

$$S_{n1}^{\mathrm{sl}} = \frac{1}{b_n} \sum_{s=1}^{b_n} \sum_{t=1}^{b_n} \operatorname{E}\Big[\mathbf{1}\Big(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big]$$
$$S_{n2}^{\mathrm{sl}} = \frac{1}{b_n} \frac{k_n - 1}{k_n} \sum_{s=b_n+1}^{2b_n} \sum_{t=1}^{b_n} \operatorname{E}\Big[\mathbf{1}\Big(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big]$$
$$S_{n3}^{\mathrm{sl}} = \frac{1}{b_n} \frac{k_n - 2}{k_n} \sum_{s=1}^{b_n} \sum_{t=b_n+1}^{2b_n} \operatorname{E}\Big[\mathbf{1}\Big(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big]$$
$$S_{n4}^{\mathrm{sl}} = \frac{1}{b_n} \sum_{i=3}^{k_n-1} \frac{k_n - i}{k_n} \sum_{s\in I_1} \sum_{t\in I_i} \operatorname{E}\Big[\mathbf{1}\Big(U_s > 1 - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big) - \tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big]$$

$$S_{n5}^{\mathrm{sl}} = \frac{1}{b_n}\sum_{i=3}^{k_n}\frac{k_n-i+1}{k_n}\sum_{s\in I_i}\sum_{t\in I_1}\mathrm{E}\Big[\mathbf{1}\Big(U_s>1-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big)-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big].$$

S_{n3}^{sl} and $S_{n4}^{\mathrm{sl}}+S_{n5}^{\mathrm{sl}}$ are negligible by the same reasons as for the treatment of S_{n2} and S_{n4} above, respectively. Regarding S_{n1}^{sl}, we can write

$$\begin{aligned} S_{n1}^{\mathrm{sl}} &= \frac{1}{b_n}\sum_{s=1}^{b_n}\sum_{t=1}^{b_n}\mathrm{E}\Big[\mathbf{1}\Big(U_s>1-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big)-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big] \\ &= \frac{1}{b_n}\sum_{t=1}^{b_n}\sum_{s=1}^{t-1}\mathrm{E}\Big[\mathbf{1}\Big(U_s>1-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big)-\tfrac{Z_{nt}^{\mathrm{sl}}}{b_n}\Big]-\frac{1}{b_n^2}\sum_{t=1}^{b_n}\sum_{s=t}^{b_n}\mathrm{E}[Z_{nt}^{\mathrm{sl}}]. \end{aligned}$$

The first summand on the right-hand side vanishes by similar arguments as we used to show the negligibility of S_{n2} above. Furthermore, the second sum on the right-hand side converges to $-\frac{1}{2\theta}$ for $n\to\infty$, by Condition 2.1(vi). Hence, $\lim_{n\to\infty}S_{n1}^{\mathrm{sl}}=-\frac{1}{2\theta}$. Similarly, $\lim_{n\to\infty}S_{n2}^{\mathrm{sl}}=-\frac{1}{2\theta}$, which finishes the proof. □

Proof of Lemma 6.1. A function f is slowly varying with index $\alpha\in\mathbb{R}$, notationally $f\in RV_\alpha$, if $\lim_{t\to\infty}f(tx)/f(t)=x^\alpha$ for any $x>0$. Recall the Potter bounds (Bingham et al., 1987, Theorem 1.5.6): if $f\in RV_\alpha$, then, for any $\delta_1,\delta_2>0$, there exists some constant $t_0=t_0(\delta_1,\delta_2)$ such that, for any t and x with $t\ge t_0, tx\ge t_0$:

$$(1-\delta_1)x^\alpha\min(x^{\delta_2},x^{-\delta_2})\le\frac{f(tx)}{f(t)}\le(1+\delta_1)x^\alpha\max(x^{\delta_2},x^{-\delta_2}).$$

Let $U(z)=F^{\leftarrow}(1-1/z)=\{1/(1-F)\}^{\leftarrow}(z)$. Since $1-F(x)\sim cx^{-\kappa}$, the function $x\mapsto 1/(1-F(x))$ is regularly varying with index κ. We obtain that $U\in RV_{1/\kappa}$ by, e.g., Proposition 0.8 (v) in Resnick (1987),

For non-negative integers $j>i$ define

$$\Pi_{i+1:j}=\prod_{k=i+1}^{j}A_k,\qquad Y_{i+1:j}=\sum_{k=i+1}^{j}\Pi_{k+1:j}B_k.$$

Then $X_j=\Pi_{i+1:j}X_i+Y_{i+1:j}$ and $(\Pi_{i+1:j},Y_{i+1:j})$ is independent of X_i. We obtain that

$$\begin{aligned} \mathbb{P}(U_i>1-y,U_j>1-y) &= \mathbb{P}\{X_i>F^{\leftarrow}(1-y),\Pi_{i+1:j}X_i+Y_{i+1:j}>F^{\leftarrow}(1-y)\} \\ &\le P_{n1}+P_{n2} \end{aligned}$$

where

$$\begin{aligned} P_{n1}&=\mathbb{P}\{X_i>F^{\leftarrow}(1-y),\Pi_{i+1:j}X_i>F^{\leftarrow}(1-y)/2\}, \\ P_{n2}&=\mathbb{P}\{X_i>F^{\leftarrow}(1-y),Y_{i+1:j}>F^{\leftarrow}(1-y)/2\}. \end{aligned}$$

Consider P_{n2}. By independence of $Y_{i+1:j}$ and X_i, we get the bound

$$\begin{aligned} P_{n2}&\le\mathbb{P}\{X_i>F^{\leftarrow}(1-y)\}\mathbb{P}\{Y_{i+1:j}>F^{\leftarrow}(1-y)/2\} \\ &\le y\mathbb{P}\{X_j>F^{\leftarrow}(1-y)/2\} \\ &= y[1-F\{F^{\leftarrow}(1-y)/2\}] \\ &\le 2^{\kappa+2}y^2 \end{aligned}$$

The last inequality follows from the Potter bounds applied to $1-F$ ($\delta_1=\delta_2=1$): we may choose c_1 sufficiently small such that

$$1-F\{F^{\leftarrow}(1-y)/2\} \le 2(1/2)^{-\kappa-1}[1-F\{F^{\leftarrow}(1-y)\}] = 2^{\kappa+2}y \qquad \forall\, y\in(0,c_1).$$

Now consider P_{n1}. By Markov's inequality and a change of variable, for any $\xi\in(0,\kappa)$,

$$\begin{aligned} P_{n1} &= \int_{F^{\leftarrow}(1-y)}^{\infty} \mathbb{P}\left\{\Pi_{i+1:j}u > F^{\leftarrow}(1-y)/2\right\} F(\,\mathrm{d}u) \\ &\le \int_{F^{\leftarrow}(1-y)}^{\infty} \mathrm{E}[\Pi_{i+1:j}^{\xi}] \left\{\frac{U(1/y))}{2u}\right\}^{-\xi} F(\,\mathrm{d}u) \\ &= 2^{\xi}\,\mathrm{E}[A_1^{\xi}]^{j-i} \int_0^y \left\{\frac{U(1/v)}{U(1/y)}\right\}^{\xi} \mathrm{d}v \end{aligned}$$

By the Potter bounds applied to $U\in RV_{1/\kappa}$, with $\delta_1=1$ and $\delta_2\in(0,1/\xi-1/\kappa)$, we have, for all sufficiently large t and for all $x\ge 1$,

$$\frac{U(tx)}{U(t)} \le 2x^{\tau}, \qquad \text{where } \tau = 1/\kappa+\delta_2 < 1/\xi.$$

With $t=1/y\ge 1/c_1$ and $x=y/v\ge 1$ we obtain, after decreasing c_1 if necessary,

$$\int_0^y \left\{\frac{U(1/v)}{U(1/y)}\right\}^{\xi} \mathrm{d}v \le 2^{\xi}\int_0^y (y/v)^{\xi\tau}\,\mathrm{d}v = \frac{2^{\xi}}{1-\tau\xi}\cdot y$$

As a consequence, $P_{n1}\le 4^{\xi}/(1-\tau\xi)\,\mathrm{E}[A_1^{\xi}]^{j-i}y$.

The derived bounds on P_{n1} and P_{n2} directly yield the bound

$$\begin{aligned} \mathrm{E}\Big\{\sum_{i=1}^n \mathbf{1}(U_i>1-y)\Big\}^2 &= \sum_{i=1}^n \mathbb{P}(U_i>1-y) + 2\sum_{1\le i<j\le n}\mathbb{P}(U_i>1-y, U_j>1-y) \\ &\le ny + 2n^2\cdot 2^{\kappa+2}y^2 + 2n\frac{4^{\xi}}{1-\tau\xi}\Big(\sum_{s=1}^{\infty}\mathrm{E}[A_1^{\xi}]^s\Big)y. \end{aligned}$$

The assertion follows from the fact that $\mathrm{E}[A_1^{\xi}]<\mathrm{E}[A_1^{\kappa}]=1$ by condition (S). □

References

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*, Volume 112 of *Mathematical Centre Tracts*. Amsterdam: Mathematisch Centrum.

Berghaus, B. and A. Bücher (2016). Goodness-of-fit tests for multivariate copula-based time series models. *Econometric Theory FirstView*, 1–39.

Billingsley, P. (1979). *Probability and measure*. John Wiley & Sons, New York-Chichester-Brisbane. Wiley Series in Probability and Mathematical Statistics.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Cambridge: Cambridge University Press.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J. 30*(1), 69–81.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv. 2*, 107–144. Update of, and a supplement to, the 1986 original.

Bücher, A. and J. Segers (2015). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *ArXiv e-prints*.

Darsow, W. F., B. Nguyen, and E. T. Olsen (1992). Copulas and Markov processes. *Illinois J. Math. 36*(4), 600–642.

de Haan, L., S. I. Resnick, H. Rootzén, and C. G. de Vries (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Process. Appl. 32*(2), 213–224.

Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In *Empirical process techniques for dependent data*, pp. 3–113. Boston, MA: Birkhäuser Boston.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab. 10*(4), 1274–1301.

Drees, H. (2002). Tail empirical processes under mixing conditions. In *Empirical process techniques for dependent data*, pp. 325–342. Birkhäuser Boston, Boston, MA.

Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist. 38*(4), 2145–2186.

Ferro, C. A. T. and J. Segers (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol. 65*(2), 545–556.

Hsing, T. (1984). *Point Processes Associated with Extreme Value Theory.* ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–The University of North Carolina at Chapel Hill.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist. 21*(4), 2043–2071.

Hsing, T., J. Hüsler, and M. R. Leadbetter (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields 78*(1), 97–112.

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

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete 65*(2), 291–306.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York-Berlin.

Leadbetter, M. R. and H. Rootzén (1988). Extremal theory for stochastic processes. *Ann. Probab. 16*(2), 431–478.

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

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab. 15*(1), 281–291.

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab. 4*(2), 529–548.

Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semiparametric models for multivariate time series. *J. Multivariate Anal. 110*, 30–42.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, Volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York.

Robert, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist. 37*(1), 271–310.

Robert, C. Y., J. Segers, and C. A. T. Ferro (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*(2), 468–490.

Segers, J. (2005). Approximate distributions of clusters of extremes. *Statist. Probab. Lett. 74*(4), 330–336.

Smith, R. L. and I. Weissman (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B 56*(3), 515–528.

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

van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge: Cambridge University Press.

Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. in Appl. Probab. 11*(4), 750–783.

Weissman, I. and S. Y. Novak (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inference 66*(2), 281–288.

Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany

E-mail address: betina.berghaus@rub.de

Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany

E-mail address: axel.buecher@rub.de