

# Method of moments estimators for the extremal index of a stationary time series

Axel Bücher and Tobias Jennessen

*Heinrich-Heine-Universität Düsseldorf  
Universitätsstr. 1  
40225 Düsseldorf*

*e-mail:* [axel.buecher@hhu.de](mailto:axel.buecher@hhu.de); [tobias.jennessen@hhu.de](mailto:tobias.jennessen@hhu.de)

**Abstract:** The extremal index  $\theta$ , a number in the interval  $[0, 1]$ , is known to be a measure of primal importance for analyzing the extremes of a stationary time series. New rank-based estimators for  $\theta$  are proposed which rely on the construction of approximate samples from the exponential distribution with parameter  $\theta$  that is then to be fitted via the method of moments. The new estimators are analyzed both theoretically as well as empirically through a large-scale simulation study. In specific scenarios, in particular for time series models with  $\theta \approx 1$ , they are found to be superior to recent competitors from the literature.

Received December 2019.

## Contents

1	Introduction . . . . .	3104
2	Definition of estimators . . . . .	3106
2.1	Approximate $\text{Exp}(\theta)$ -samples based on disjoint blocks maxima . . . . .	3106
2.2	Approximate $\text{Exp}(\theta)$ -samples based on sliding blocks maxima . . . . .	3106
2.3	Preliminaries on the exponential distribution . . . . .	3107
2.4	Definition of the estimators . . . . .	3107
3	Mathematical preliminaries . . . . .	3108
4	Asymptotic results . . . . .	3112
5	Example: max-autoregressive process . . . . .	3115
5.1	Asymptotic variances for the ARMAX-model . . . . .	3116
5.2	Regularity conditions for the ARMAX-model . . . . .	3117
6	Finite-sample results . . . . .	3117
6.1	Comparison of the introduced estimators . . . . .	3118
6.2	Comparison with other estimators for the extremal index . . . . .	3121
7	Conclusion . . . . .	3123
A	Proofs of Theorems 4.1–4.3 . . . . .	3124
A.1	Proof of Theorem 4.1 . . . . .	3125
A.2	Proof of Theorem 4.2 . . . . .	3127

<b>A.3 Proof of Theorem 4.3</b> . . . . .	3129
<b>B Auxiliary results for the proof of Theorem 4.1</b> . . . . .	3131
<b>B.1 Auxiliary lemmas – disjoint blocks</b> . . . . .	3131
<b>B.2 Auxiliary lemmas – sliding blocks</b> . . . . .	3141
<b>B.3 Further auxiliary lemmas</b> . . . . .	3152
<b>Acknowledgements</b> . . . . .	3154
<b>References</b> . . . . .	3154

## 1. Introduction

The statistical analysis of the extremal behavior of a stationary time series is important in many fields of application, such as in hydrology, meteorology, finance or actuarial science [1]. Such an analysis typically consists of two steps: (1) assessing the tail of the marginal law and (2) assessing the serial dependence of the extremes, that is, the tendency that extreme observations occur in clusters. The present work is concerned with step (2). The most common and simplest mathematical object capturing the serial dependence between the extremes is provided by the extremal index  $\theta \in [0, 1]$ . In a suitable asymptotic framework, the extremal index can be interpreted as the reciprocal of the expected size of a cluster of extreme observations. The underlying probabilistic theory was worked out in [18, 19, 23, 17, 20].

Estimating the extremal index based on a finite stretch of observations from the time series has been extensively studied in the literature. An early overview is provided in Section 10.3.4 in [1], where the estimators are classified into three groups: estimators based on the blocks method, the runs method or the inter-exceedance time method. Respective references are [16, 31, 13, 32, 27, 22, 12, 11, 5], among many others. The proposed estimators typically depend on two or, arguably preferable, one parameter to be chosen by the statistician. The present paper is on a class of method of moments estimators (based on the blocks method), which improves upon a recent estimator proposed by Paul Northrop in [22] and analyzed theoretically in [3].

Some notations and assumptions are necessary for the motivation of the new class of estimators. Throughout the paper,  $X_1, X_2, \dots$  denotes a stationary sequence of real-valued random variables with continuous cumulative distribution function (c.d.f.)  $F$ . The sequence is assumed to have an extremal index  $\theta \in (0, 1]$ , i.e., for any  $\tau > 0$ , there exists a sequence  $u_b = u_b(\tau)$ ,  $b \in \mathbb{N}$ , such that  $\lim_{b \rightarrow \infty} b\bar{F}(u_b) = \tau$  and

$$\lim_{b \rightarrow \infty} \mathbb{P}(M_{1:b} \leq u_b) = e^{-\theta\tau}, \quad (1.1)$$

where  $\bar{F} = 1 - F$  and  $M_{1:b} = \max\{X_1, \dots, X_b\}$ . Next, define a sequence of standard uniform random variables by  $U_s = F(X_s)$  and let

$$Y_{1:b} = -b \log(N_{1:b}), \quad N_{1:b} = F(M_{1:b}) = \max\{U_1, \dots, U_b\}. \quad (1.2)$$

Since  $b\bar{F}\{F^{\leftarrow}(e^{-x/b})\} = b(1 - e^{-x/b}) \rightarrow x$  for  $b \rightarrow \infty$ , it follows from (1.1) that, for any  $x > 0$ ,

$$\mathbb{P}(Y_{1:b} \geq x) = \mathbb{P}(M_{1:b} \leq F^{\leftarrow}(e^{-x/b})) \rightarrow e^{-\theta x}, \quad (1.3)$$

where  $F^{\leftarrow}(z) = \inf\{y \in \mathbb{R} : F(y) \geq z\}$  denotes the generalized inverse of  $F$  evaluated at  $z \in \mathbb{R}$ . In other words, for large block length  $b$ ,  $Y_{1:b}$  approximately follows an exponential distribution with parameter  $\theta$ , denoted by  $\text{Exp}(\theta)$  throughout. This inspired [22] and [3] to estimate  $\theta$  by the maximum likelihood estimator for the exponential distribution; see Section 2 below for details on how to arrive at an observable (rank-based) approximate sample from the  $\text{Exp}(\theta)$ -distribution based on an observed stretch of length  $n$  from the time series  $(X_s)_{s \in \mathbb{N}}$ .

The idea of transforming observations into a sample of exponentially distributed observations is actually not new within extreme value statistics: it is also, among many others, the main motivation for the Pickands estimator in multivariate extremes [25, 14]. More precisely, if  $(X, Y)$  is a bivariate random vector from a multivariate extreme value distribution with Pickands function  $A = (A(w))_{w \in [0,1]}$ , then  $\xi(w) = \min\{-\log F_X(X)/(1-w), -\log F_Y(Y)/w\}$  is exponentially distributed with parameter  $A(w)$ . Given a sample of size  $n$  from  $(X, Y)$ , we may replace  $F_X$  and  $F_Y$  by their empirical counterparts and arrive at an approximate sample of size  $n$  from the  $\text{Exp}(A(w))$ -distribution, to be, for instance, estimated by the maximum likelihood estimator.

The present paper is now motivated by the following observation: while the maximum likelihood estimator is asymptotically efficient in the ideal situation of observing an i.i.d. sample from the exponential distribution, it was shown in [14] for rank-based estimators of the Pickands function that it is in fact more efficient to consider alternative estimators based on the method of moments, such as a rank-based version of the CFG-estimator [6]. Given that Northrop's blocks estimator is also rank-based, the main motivation of this work is to consider CFG-type estimators for the extremal index  $\theta$ . Alongside, we will also investigate other moment-based estimators, including one that is closely connected to the madogram estimator in [21]. We will show that, depending on the true value of  $\theta$ , the new estimators may either exhibit a smaller or a larger asymptotic variance than Northrop's maximum likelihood estimator. In particular, we will show that the CFG-type estimator's variance is substantially smaller for  $\theta$  close to one, i.e., for time series with little clustering of extremes.

The remaining parts of this paper are organized as follows: in Section 2, we collect some results about certain useful moments of the exponential distribution and use those to introduce the new estimators for  $\theta$ . Regularity assumptions needed to prove asymptotic results are summarized and discussed in Section 3. The paper's main results are then presented in Section 4, alongside with a discussion of certain aspects of the derived asymptotic variance formulas. Section 5 is about a particular time series model, for which we show that all regularity conditions imposed in Section 3 are met. The finite-sample performance of the new estimators is investigated in a Monte-Carlo simulation study in Section 6. Finally, all proofs are postponed to Section A.

## 2. Definition of estimators

Recall the definition of  $Y_{1:b}$  in (1.2), where  $b \in \mathbb{N}$ . Similarly, let

$$Z_{1:b} = b(1 - N_{1:b}), \quad N_{1:b} = F(M_{1:b}) = \max\{U_1, \dots, U_b\},$$

and note that, as  $b \rightarrow \infty$  and for any  $x > 0$ ,

$$\mathbb{P}(Z_{1:b} \geq x) = \mathbb{P}(M_{1:b} \leq F^{\leftarrow}(1 - x/b)) \rightarrow e^{-\theta x} \quad (2.1)$$

by similar arguments as for  $Y_{1:b}$ . The convergence relations in (1.3) and (2.1) serve as a basis for the method of moments estimators defined below.

Subsequently, let  $X_1, \dots, X_n$  denote a finite stretch of observations from the stationary sequence  $(X_s)_{s \geq 1}$ . Within Section 2.1 and 2.2, we start by using (1.3) and (2.1) to derive some observable, approximate samples from the  $\text{Exp}(\theta)$ -distribution. In Section 2.3, we collect some moment equations for the exponential distribution, which will then be used to motivate new estimators for the extremal index in Section 2.4.

### 2.1. Approximate $\text{Exp}(\theta)$ -samples based on disjoint blocks maxima

Divide the sample  $X_1, \dots, X_n$  into  $k_n$  successive blocks of size  $b_n$ , and for simplicity assume that  $n = b_n k_n$  (otherwise, the last block of less than  $b_n$  observations should be deleted). For  $i = 1, \dots, k_n$ , let

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

denote the maximum of the  $X_s$  in the  $i$ th block of observations and let

$$Y_{ni} = -b_n \log N_{ni}, \quad Z_{ni} = b_n(1 - N_{ni}), \quad N_{ni} = F(M_{ni}).$$

Due to relations (1.3) and (2.1), if the block size  $b = b_n$  is sufficiently large, the (unobservable) random variables  $Y_{ni}$  and  $Z_{ni}$  are approximately exponentially distributed with parameter  $\theta$ . Observable counterparts are obtained by replacing  $F$  by the (slightly adjusted) empirical c.d.f.  $\hat{F}_n(x) = (n+1)^{-1} \sum_{s=1}^n \mathbf{1}(X_s \leq x)$ , giving rise to the definitions

$$\hat{Y}_{ni} = -b_n \log \hat{N}_{ni}, \quad \hat{Z}_{ni} = b_n(1 - \hat{N}_{ni}), \quad \hat{N}_{ni} = \hat{F}_n(M_{ni}).$$

Both the samples  $\mathcal{Y}_n^{\text{db}} = \{\hat{Y}_{ni} : i = 1, \dots, k_n\}$  and  $\mathcal{Z}_n^{\text{db}} = \{\hat{Z}_{ni} : i = 1, \dots, k_n\}$  will be used later to define *disjoint blocks estimators for  $\theta$*  (note that both samples are dependent over  $i$  due to the use of  $\hat{F}_n$ , which complicates the asymptotic analysis).

### 2.2. Approximate $\text{Exp}(\theta)$ -samples based on sliding blocks maxima

As in the previous paragraph, let  $n$  denote the sample size and  $b_n$  denote a block length parameter (the assumption that  $k_n = n/b_n \in \mathbb{N}$  is not needed, no discarding is necessary). For  $t = 1, \dots, n - b_n + 1$ , let

$$M_{nt}^{\text{sb}} = M_{t:t+b_n-1} = \max\{X_t, \dots, X_{t+b_n-1}\}$$

denote the maximum of the  $X_s$  in a block of length  $b_n$  starting at observation  $t$ . Define

$$\begin{aligned} Y_{nt}^{\text{sb}} &= -b_n \log N_{nt}^{\text{sb}}, & Z_{nt}^{\text{sb}} &= b_n(1 - N_{nt}^{\text{sb}}), & N_{nt}^{\text{sb}} &= F(M_{nt}^{\text{sb}}), \\ \hat{Y}_{nt}^{\text{sb}} &= -b_n \log \hat{N}_{nt}^{\text{sb}}, & \hat{Z}_{nt}^{\text{sb}} &= b_n(1 - \hat{N}_{nt}^{\text{sb}}), & \hat{N}_{nt}^{\text{sb}} &= \hat{F}_n(M_{nt}^{\text{sb}}). \end{aligned}$$

By the same heuristics as before, the observable samples  $\mathcal{Y}_n^{\text{sb}} = \{\hat{Y}_{nt}^{\text{sb}} : t = 1, \dots, n - b_n + 1\}$  and  $\mathcal{Z}_n^{\text{sb}} = \{\hat{Z}_{nt}^{\text{sb}} : t = 1, \dots, n - b_n + 1\}$  are approximate samples from the exponential distribution and will be used later to define *sliding blocks estimators for  $\theta$*  (both samples are heavily dependent over  $i$  due to the use of  $\hat{F}_n$  and the use of overlapping blocks).

### 2.3. Preliminaries on the exponential distribution

Some important moment equations, valid for a random variable  $\xi$ , which is  $\text{Exp}(\theta)$ -distributed, are collected. First,

$$E[\log \xi] = -\log \theta - \gamma =: \varphi_{(C)}(\theta), \tag{CFG}$$

where  $\gamma = -\int_0^\infty \log(x)e^{-x} dx \approx 0.577$  denotes the Euler-Mascheroni-constant. Equation (CFG) is the basis for motivating the CFG-estimator, see [6, 14] and the details in Section 1. Next, note that

$$E[\exp(-\xi)] = \frac{\theta}{1 + \theta} =: \varphi_{(M)}(\theta), \tag{MAD}$$

which serves as a basis for the madogram, see [21]. A further choice, including (CFG) as a limit, is provided by

$$E[\xi^{1/p}] = \theta^{-1/p} \Gamma(1 + 1/p) =: \varphi_{(R),p}(\theta), \tag{ROOT}$$

where  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$  denotes the Gamma function and where  $p > 0$ . The moment estimator in case of  $p = 1$  will turn out to coincide with Northrop's maximum likelihood estimator. Also note that the previous equation is equivalent to

$$E\left[\frac{\xi^{1/p} - 1}{1/p}\right] = \frac{\theta^{-1/p} \Gamma(1 + 1/p) - 1}{1/p} =: \tilde{\varphi}_{(R),p}(\theta), \tag{2.2}$$

and taking the limits for  $p \rightarrow \infty$  on both sides (interchanging the limit and the expectation on the left) exactly yields Equation (CFG).

### 2.4. Definition of the estimators

Let  $\chi_m = \{\xi_1, \dots, \xi_m\}$  denote a generic sample (not necessarily independent) from the  $\text{Exp}(\theta)$ -distribution. Replacing the moments in Equations (CFG),

(MAD) and (ROOT) by their empirical counterparts and solving the equation for  $\theta$ , we obtain the following three estimators for  $\theta$ :

$$\begin{aligned}\hat{\theta}_{\text{CFG}}(\chi_m) &= e^{-\gamma} \exp \left\{ -\frac{1}{m} \sum_{i=1}^m \log(\xi_i) \right\}, \\ \hat{\theta}_{\text{MAD}}(\chi_m) &= \frac{\frac{1}{m} \sum_{i=1}^m \exp(-\xi_i)}{1 - \frac{1}{m} \sum_{i=1}^m \exp(-\xi_i)}, \\ \hat{\theta}_{\text{R},p}(\chi_m) &= \Gamma(1 + 1/p)^p \left( \frac{1}{m} \sum_{i=1}^m \xi_i^{1/p} \right)^{-p},\end{aligned}$$

where  $p > 0$ . It may be verified that  $\lim_{p \rightarrow \infty} \hat{\theta}_{\text{R},p}(\chi_m) = \hat{\theta}_{\text{CFG}}(\chi_m)$ , see also (2.2) for another relationship between the two estimators. Next, replacing  $\chi_m$  by any of the four samples  $\mathcal{Y}_n^{\text{db}}, \mathcal{Z}_n^{\text{db}}, \mathcal{Y}_n^{\text{sb}}$  or  $\mathcal{Z}_n^{\text{sb}}$  defined in Sections 2.1 and 2.2, we finally arrive at 12 method of moments estimators for  $\theta$ . We use the suggestive notations

$$\hat{\theta}_{\text{db,CFG}}^{y_n} = \hat{\theta}_{\text{CFG}}(\mathcal{Y}_n^{\text{db}}), \quad \hat{\theta}_{\text{sb,MAD}}^{z_n} = \hat{\theta}_{\text{MAD}}(\mathcal{Z}_n^{\text{sb}})$$

to, e.g., denote the disjoint blocks CFG-estimator based on the  $\hat{Y}_{ni}$  and the sliding blocks madogram-estimator based on the  $\hat{Z}_{ni}$ , respectively. Note that the four estimators of the form  $\hat{\theta}_{\text{m,R},1}^{y_n}, \hat{\theta}_{\text{m,R},1}^{z_n}, \text{m} \in \{\text{db}, \text{sb}\}$ , are the (pseudo) maximum likelihood (PML) estimators considered in [3].

### 3. Mathematical preliminaries

Further mathematical details are necessary before we can state asymptotic results about the estimators defined in the previous section. The asymptotic framework and the conditions are mostly similar as in Section 2 in [3], but will be repeated here for the sake of completeness.

The serial dependence of the time series  $(X_s)_{s \in \mathbb{N}}$  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  $\{X_s \mathbf{1}(X_s > F^{\leftarrow}(1 - \varepsilon_n)) : s \leq \ell\}$  and  $\{X_s \mathbf{1}(X_s > F^{\leftarrow}(1 - \varepsilon_n)) : s \geq \ell + k\}$ , where  $\varepsilon_n \rightarrow 0$  is some sequence reflecting the fact that only the dependence in the tail needs to be restricted (see, e.g., 29). As in [3], we need a slightly stronger condition, that also controls the dependence between the smallest of all block maxima. More precisely, for  $-\infty \leq p < q \leq \infty$  and  $\varepsilon \in (0, 1]$ , let  $\mathcal{B}_{p,q}^\varepsilon$  denote the sigma algebra generated by  $U_s^\varepsilon := U_s \mathbf{1}(U_s > 1 - \varepsilon)$  with  $s \in \{p, \dots, q\}$  and define, for  $\ell \geq 1$ ,

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

In Condition 3.1(iii) below, we will impose a condition on the decay of the mixing coefficients for small values of  $\varepsilon$ . Note that the coefficients are bounded by the standard alpha-mixing coefficients of the sequence  $U_s$ , which can be retrieved for  $\varepsilon = 1$ .

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

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

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

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

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

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

By Theorems 4.1 and 4.2 in [17], 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  $\pi$  on  $\mathbb{N}$ , that is,

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

where  $q_n$  is some  $\Delta(u_n(x))$ -separating sequence. Moreover, in that case, the convergence in the last display holds for any  $\Delta(u_n(x))$ -separating sequence  $q_n$ , and the weak limit of  $N_n^{(x)}$  is a compound poisson process  $\text{CP}(\theta x, \pi)$ . If the  $\Delta(u_n(x))$ -condition holds for any  $x > 0$ , then  $\pi$  does not depend on  $x$  (17, Theorem 5.1).

A multivariate version of the latter results is stated in [24], see also the summary in [27], page 278, and the thesis [15]. Suppose that the extremal index exists and that the  $\Delta(u_n(x_1), u_n(x_2))$ -condition is met for any  $x_1 \geq x_2 \geq 0$ ,  $x_1 \neq 0$ . Moreover, assume that there exists a family of probability measures  $\{\pi_2^{(\sigma)} : \sigma \in [0, 1]\}$  on  $\mathcal{J} = \{(i, j) : i \geq j \geq 0, i \geq 1\}$ , such that, for all  $(i, j) \in \mathcal{J}$ ,

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

where  $q_n$  is some  $\Delta(u_n(x_1), u_n(x_2))$ -separating sequence. In that case, the two-level point process  $\mathbf{N}_n^{(x_1, x_2)} = (N_n^{(x_1)}, N_n^{(x_2)})$  converges in distribution to a point process with characterizing Laplace transform explicitly stated in [27] on top of page 278. Note that

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

Finally, we will need the tail empirical process

$$e_n(x) = \frac{1}{\sqrt{k_n}} \sum_{s=1}^n \left\{ \mathbb{1}\left(U_s > 1 - \frac{x}{b_n}\right) - \frac{x}{b_n} \right\}, \quad x \geq 0, \quad (3.1)$$

where  $U_s = F(X_s)$ , see, e.g., [10, 29].

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

**Condition 3.1.**

- (i) The stationary time series  $(X_s)_{s \in \mathbb{N}}$  has an extremal index  $\theta \in (0, 1]$  and the above assumptions guaranteeing convergence of the one- and two-level point process of exceedances are satisfied.
- (ii) There exists  $\delta > 0$  such that, for any  $m > 0$ , there exists a constant  $\tilde{C}_m$  such that, for all  $0 \leq x_1 \leq x_2 \leq m, n \in \mathbb{N}$ ,

$$\mathbb{E} [|N_n^{(x_1)}(E) - N_n^{(x_2)}(E)|^{2+\delta}] \leq \tilde{C}_m(x_2 - x_1).$$

- (iii) There exist constants  $c_2 \in (0, 1)$  and  $C_2 > 0$  such that

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

for some  $\eta \geq 3(2 + \delta)/(\delta - \mu) > 3$ , where  $0 < \mu < \min(\delta, 1/2)$  and  $\delta > 0$  is from Condition (ii). The block size  $b_n$  converges to infinity and satisfies

$$k_n = o(b_n^2), \quad n \rightarrow \infty.$$

Further, there exists a sequence  $\ell_n \rightarrow \infty$  with  $\ell_n = o(b_n^{2/(2+\delta)})$  and  $k_n \alpha_{c_2}(\ell_n) = o(1)$  as  $n \rightarrow \infty$ .

- (iv) There exist constants  $c_1 \in (0, 1)$  and  $C_1 > 0$  such that, for any  $y \in (0, c_1)$  and  $n \in \mathbb{N}$ ,

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

- (v) For any  $c \in (0, 1)$ , one has

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

where  $N'_{ni} = \max\{U_s, s \in [(i-1)b_n/2 + 1, \dots, ib_n/2]\}$  for  $i = 1, \dots, 2k_n$ .



(vi) For any  $x > 0$ ,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left( N_{m:b_n} > 1 - \frac{x}{n} \mid U_1 \geq 1 - \frac{x}{n} \right) = 0.$$

**Condition 3.2** (Integrability).

(i) With  $\delta > 0$  from Condition 3.1(ii), one has

$$\limsup_{n \rightarrow \infty} \mathbb{E} [ |\log(Z_{1:n})|^{2+\delta} ] < \infty.$$

(ii) Fix  $p > 0$ . With  $\delta > 0$  from Condition 3.1(ii), one has

$$\limsup_{n \rightarrow \infty} \mathbb{E} [ Z_{1:n}^{(2+\delta)/p} ] < \infty.$$

**Condition 3.3** (Bias Condition). Recall  $\varphi_{(C)}$ ,  $\varphi_{(M)}$  and  $\varphi_{(R),p}$  defined in Equations (CFG), (MAD) and (ROOT), respectively.

(i) As  $n \rightarrow \infty$ ,  $\mathbb{E}[\log(Z_{1:b_n})] = \varphi_{(C)}(\theta) + o(k_n^{-1/2})$ .

(ii) As  $n \rightarrow \infty$ ,  $\mathbb{E}[\exp(-Z_{1:b_n})] = \varphi_{(M)}(\theta) + o(k_n^{-1/2})$ .

(iii) Fix  $p > 0$ . As  $n \rightarrow \infty$ ,  $\mathbb{E} [ Z_{1:b_n}^{1/p} ] = \varphi_{(R),p}(\theta) + o(k_n^{-1/2})$ .

**Condition 3.4** (Technical Condition for the CFG-type estimator).

(i) For some  $q > 1/2$ , we have  $b_n = O(k_n^q)$  as  $n \rightarrow \infty$ .

(ii) For some  $\tau \in (0, 1/2)$ , we have, as  $n \rightarrow \infty$ ,

$$\left\{ \frac{e_n(x)}{x^\tau} \right\}_{x \in [0,1]} \xrightarrow{d} \left\{ \frac{e(x)}{x^\tau} \right\}_{x \in [0,1]} \quad \text{in } D([0, 1]),$$

the càglàd space of functions on  $[0, 1]$ , where  $e_n$  denotes the tail empirical process defined in (3.1) and where  $e$  is a centered Gaussian process with continuous sample paths and covariance as given in Lemma B.1.

(iii) For any  $c > 0$ , we have, as  $n \rightarrow \infty$ ,

$$\max_{Z_{ni} \geq c} \left| \frac{e_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right| = o_{\mathbb{P}}(1).$$

(iv) For any  $c > 0$ , there exists  $\mu = \mu_c \in (1/2, 1/\{2(1 - \tau)\})$  with  $\tau$  from (ii) such that, as  $n \rightarrow \infty$ ,

$$\mathbb{P}(Z_{n1} < ck_n^{-\mu}) - \mathbb{P}(\xi < ck_n^{-\mu}) = o(\log(n)^{-1}k_n^{-1/2}), \quad \text{where } \xi \sim \text{Exp}(\theta).$$

The items of Condition 3.1 are the same as Condition 2.1(i)-(v) and (2.2) in [3] and are discussed in great detail in that reference. Condition 3.2 is needed for uniform integrability of the sequences  $Z_{n1}^{2/p}$  and  $\log^2 Z_{n1}$ , respectively. It implies

$$\lim_{n \rightarrow \infty} \text{Var}(Z_{n1}^{1/p}) = \text{Var}(\xi^{1/p}), \quad \lim_{n \rightarrow \infty} \text{Var}(\log Z_{n1}) = \text{Var}(\log \xi),$$

respectively, where  $\xi$  denotes an exponentially distributed random variable with parameter  $\theta$ . Condition 3.3 is a bias condition requiring the approximation of the first moment of  $f(Z_{n1})$  by  $E[f(\xi)]$  to be sufficiently accurate, where  $f(x) \in \{x^{1/p}, \exp(-x), \log x\}$ .

Condition 3.4 is a technical condition which is only needed for deriving the asymptotics of the CFG-estimator. The Condition 3.4(i) requires  $b$  to be not too large. Sufficient conditions for Condition 3.4(ii) in terms of beta mixing coefficients can be found in [10]. A sufficient condition for Condition 3.4(iii) is for instance strong mixing with polynomial rate  $\alpha_1(n) = O(n^{-(1+\sqrt{2})-\varepsilon}), n \rightarrow \infty$ , for some  $\varepsilon > 0$ , together with Condition 3.4(i) being met with  $q < 1/(\sqrt{2}-1) \approx 2.41$ . Indeed, for any  $x \geq c$  and  $\eta > 0$ , one can write

$$\frac{e_n(x)}{x} = \frac{1}{\sqrt{k_n}} \sum_{s=1}^n \left\{ \mathbb{1}\left(U_s > 1 - \frac{x}{b_n}\right) - \frac{x}{b_n} \right\} \frac{1}{x} = -b_n^{1/2-\eta} \mathbb{U}_{n,\eta} \left(1 - \frac{x}{b_n}\right) \frac{1}{x^{1-\eta}},$$

where

$$\mathbb{U}_{n,\eta}(u) = \frac{\frac{1}{\sqrt{n}} \sum_{s=1}^n \{\mathbb{1}(U_s \leq u) - u\}}{(1-u)^\eta} \mathbb{1}_{(0,1)}(u).$$

By Theorem 2.2 in [30], we have  $\sup_{x \geq 0} |\mathbb{U}_{n,\eta}(1 - x/b_n)| = O_{\mathbb{P}}(1)$  for all  $\eta \leq 1 - 2^{-1/2} \approx 0.29$ . Hence, by Condition 3.4(i),

$$\max_{Z_{ni} \geq c} \left| \frac{e_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right| = O_{\mathbb{P}} \left( \frac{b_n^{1/2-\eta}}{\sqrt{k_n}} \right) = O_{\mathbb{P}} \left( k_n^{q(1/2-\eta)-1/2} \right).$$

The expression on the right-hand side is  $o_{\mathbb{P}}(1)$  if we choose  $\eta \in (1/2 - 1/\{2q\}, 1 - 2^{-1/2}]$ ; note that the latter interval is non-empty since  $q < 1/(\sqrt{2}-1)$ . Finally, Condition 3.4(iv) is another technical condition requiring the approximation of the law of  $Z_{n1}$  by the exponential distribution to be sufficiently accurate in the lower tail.

### 4. Asymptotic results

We present asymptotic results on all estimators defined in Section 2. For simplicity, all results are stated and proved for the  $\hat{Z}_{ni}$ -versions only. As in Theorem 3.1 in [3], it may be verified that the respective versions based on  $\hat{Y}_{ni}$  show the same asymptotic behavior as the  $\hat{Z}_{ni}$ -versions. Throughout, for  $z \in (0, 1)$ , let  $(\xi_1^{(z)}, \xi_2^{(z)}) \sim \pi_2^{(z)}$ .

**Theorem 4.1.** *Under Condition 3.1, 3.2(i), 3.3(i) and 3.4, we have*

$$\sqrt{k_n}(\hat{\theta}_{m,\text{CFG}}^{z_n} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_{m,C}^2)$$

for  $m \in \{\text{db}, \text{sb}\}$  and as  $n \rightarrow \infty$ , where

$$\begin{aligned} \sigma_{\text{db},C}^2 &= 2\theta^3 \int_0^1 \frac{\theta E[\xi_1^{(z)} \xi_2^{(z)}] - E[\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} > 0)]}{z(1+z)} dz + \{\pi^2/6 - 2 \log(2)\} \theta^2, \\ \sigma_{\text{sb},C}^2 &= \sigma_{\text{db},C}^2 - \{\pi^2/6 - 8 \log(2) + 4\} \theta^2. \end{aligned}$$

**Theorem 4.2.** Under Condition 3.1 and 3.3(ii), we have

$$\sqrt{k_n}(\hat{\theta}_{m,\text{MAD}}^{z_n} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_{m,\text{M}}^2)$$

for  $m \in \{\text{db}, \text{sb}\}$  and as  $n \rightarrow \infty$ , where

$$\begin{aligned} \sigma_{\text{db},\text{M}}^2 &= 4\theta^2(1 + \theta) \int_0^1 \frac{\theta \mathbb{E}[\xi_1^{(z)} \xi_2^{(z)}] - \mathbb{E}[\xi_1^{(z)} \mathbf{1}(\xi_2^{(z)} > 0)]}{(1 + z)^3} dz + \frac{\theta^2(1 + \theta)}{2(2 + \theta)} \\ \sigma_{\text{sb},\text{M}}^2 &= \sigma_{\text{db},\text{M}}^2 - \frac{3\theta^2 + 4\theta - 4(1 + \theta)(2 + \theta) \log\{2(1 + \theta)/(2 + \theta)\}}{\theta(2 + \theta)(1 + \theta)^2}. \end{aligned}$$

**Theorem 4.3.** Fix  $p > 0$ . Under Condition 3.1, 3.2(ii) and 3.3(iii),

$$\sqrt{k_n}(\hat{\theta}_{m,\text{R},p}^{z_n} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_{m,p}^2)$$

for  $m \in \{\text{db}, \text{sb}\}$  and as  $n \rightarrow \infty$ , where

$$\begin{aligned} \sigma_{\text{db},p}^2 &= \frac{4p\theta^3}{\text{B}(1/p, 1/p)} \int_0^1 \frac{\theta \mathbb{E}[\xi_1^{(z)} \xi_2^{(z)}] + \mathbb{E}[\xi_1^{(z)} \mathbf{1}(\xi_2^{(z)} = 0)]z^{\frac{1}{p}-1}}{(1 + z)^{1+\frac{2}{p}}} dz \\ &\quad + \left\{ \frac{2p^3}{\text{B}(1/p, 1/p)} - p^2 - 2p \right\} \theta^2, \\ \sigma_{\text{sb},p}^2 &= \sigma_{\text{db},p}^2 - \left[ p^2 + \frac{2p^3}{\text{B}(1/p, 1/p)} \right. \\ &\quad \left. - \frac{4p}{\Gamma(1/p)^2} \int_0^\infty (1 - e^{-z})z^{1/p-2}\Gamma(1/p, z) dz \right] \theta^2, \end{aligned}$$

where  $\text{B}(x, y) = \int_0^1 t^{x-1}(1 - t)^{y-1} dt$  denotes the beta function and  $\Gamma(x, z) = \int_x^\infty t^{z-1}e^{-t} dt$  is the incomplete gamma function.

It is worthwhile to mention that the imposed conditions in each theorem are exactly the same for the disjoint and the sliding blocks version. Furthermore, apart from the different bias conditions, the conditions regarding  $k_n$  are exactly the same in Theorem 4.2 and 4.3, and slightly stronger for Theorem 4.1 in that the additional technical Condition 3.4 is imposed.

The proofs are provided in Section A and bear some similarities with the one of Theorem 3.2 in [3]. In particular, they rely on the delta method, Wichura's theorem and empirical process theory to adequately handle the asymptotic contribution of the rank transformation. The most sophisticated proof is the one of Theorem 4.1, which is essentially due to the fact that  $\mathbb{E}[\log \xi] = \int_0^\infty \log(t)\theta e^{-\theta t} dt$  is an improper integral both at zero and at infinity (see also [14] for similar technical difficulties with the CFG-estimator for the Pickands dependence function in multivariate extremes).

It is worth to mention that the difference

$$\text{AsyVar}(\sqrt{k_n}\hat{\theta}_{\text{db},\text{CFG}}^{z_n}/\theta) - \text{AsyVar}(\sqrt{k_n}\hat{\theta}_{\text{sb},\text{CFG}}^{z_n}/\theta) = (\sigma_{\text{db},\text{C}}^2 - \sigma_{\text{sb},\text{C}}^2)/\theta^2 \approx 0.0977$$

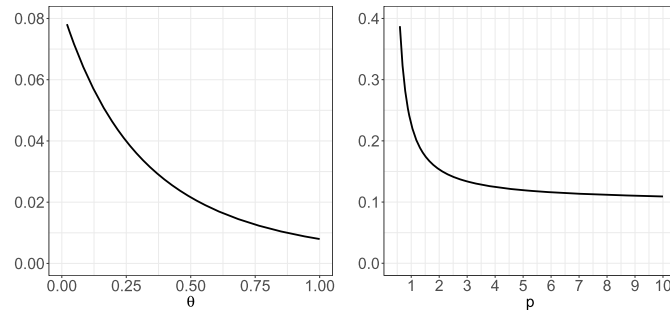


FIG 1. Graph of the functions  $\theta \mapsto (\sigma_{\text{db},M}^2 - \sigma_{\text{sb},M}^2)/\theta^2$  (left) and  $p \mapsto (\sigma_{\text{db},p}^2 - \sigma_{\text{sb},p}^2)/\theta^2$  (right).

is a universal constant independent of any properties of the observed time series. The same holds true for the Root-estimator with a constant depending in a complicated way on the parameter  $p$  (the graph of  $p \mapsto (\sigma_{\text{db},p}^2 - \sigma_{\text{sb},p}^2)/\theta^2$  is depicted in Figure 1, with a value of approximately 0.2274 for the PML-estimator). For the Madogram-estimator, this difference depends on  $\theta$  (see Figure 1 for the graph of  $\theta \mapsto (\sigma_{\text{db},M}^2 - \sigma_{\text{sb},M}^2)/\theta^2$ ); it is non-negative and decreasing with value  $1/12 \approx 0.083$  for  $\theta \rightarrow 0$  and approximately 0.0079 for  $\theta = 1$ . In that regard, the use of sliding blocks over disjoint blocks is least beneficial for the Madogram-estimator.

*Example 4.4.* In the case that the time series is serially independent, the cluster size distributions are given by  $\pi(i) = \mathbf{1}(i = 1)$  and  $\pi_2^{(z)}(i, j) = (1 - z)\mathbf{1}(i = 1, j = 0) + z\mathbf{1}(i = 1, j = 1)$ , which implies

$$\theta = 1, \quad E[\xi_1^{(z)}\xi_2^{(z)}] = z \quad \text{and} \quad E[\xi_1^{(z)}\mathbf{1}(\xi_2^{(z)} = 0)] = 1 - z.$$

It can be seen that these formulas hold true whenever  $\theta = 1$ . Consequently, the limiting variances in Theorem 4.1 and 4.2 are equal to

$$\begin{aligned} \sigma_{\text{db},C}^2 &= \frac{\pi^2}{6} - 2 \log(2) \approx 0.2586, & \sigma_{\text{sb},C}^2 &= 6 \log(2) - 4 \approx 0.1588, \\ \sigma_{\text{db},M}^2 &= 1/3, & \sigma_{\text{sb},M}^2 &\approx 0.32536. \end{aligned}$$

It is remarkable that the asymptotic variances are substantially smaller than those of the maximum likelihood estimator, see Example 3.1 in [3], which are equal to  $1/2$  and  $0.2726$  for the disjoint and sliding blocks version, respectively.

The limiting variance in the case of the Root-estimator is given by

$$\begin{aligned} \sigma_{\text{db},p}^2 &= \frac{2p}{B(\frac{1}{p}, \frac{1}{p})} \left[ p^2 + 2^{-2/p}p \right] - p^2 - p, \\ \sigma_{\text{sb},p}^2 &= \sigma_{\text{db},p}^2 - \left[ p^2 + \frac{2p^3}{B(\frac{1}{p}, \frac{1}{p})} - \frac{4p}{\Gamma(\frac{1}{p})^2} \int_0^\infty (1 - e^{-z})z^{1/p-2}\Gamma(\frac{1}{p}, z) dz \right]. \end{aligned}$$

Some values are

$$\begin{aligned} \sigma_{\text{db},1/2}^2 &= \frac{15}{16}, & \sigma_{\text{db},1}^2 &= \frac{1}{2}, & \sigma_{\text{db},2}^2 &\approx 0.3662, \\ \sigma_{\text{sb},1/2}^2 &= \frac{7}{16}, & \sigma_{\text{sb},1}^2 &\approx 0.2726, & \sigma_{\text{sb},2}^2 &\approx 0.212909. \end{aligned}$$

It can further be shown that  $\lim_{p \rightarrow \infty} \sigma_{m,p}^2 = \sigma_{m,C}^2$  for  $m \in \{\text{db}, \text{sb}\}$ .

*Remark 4.5.* Instead of working with  $\hat{F}_n$  in the definition of  $\hat{Z}_{ni} = b_n\{1 - \hat{F}_n(M_{ni})\}$ , one may alternatively use the empirical c.d.f. of  $(X_s)_{s \notin I_i}$  multiplied by  $(n - b_n)/(n - b_n + 1)$  for  $I_i = \{(i - 1)b_n + 1, \dots, ib_n\}$ , denoted by  $\hat{F}_{n,-i}$ , and define  $\tilde{Z}_{ni} = b_n\{1 - \hat{F}_{n,-i}(M_{ni})\}$  and  $\tilde{\theta} = \hat{\theta}(\tilde{Z}_{n1}, \dots, \tilde{Z}_{nk_n})$ . This modification has been motivated as a bias reduction scheme in [22]. Since

$$\tilde{Z}_{ni} = b_n\{1 - \hat{F}_{n,-i}(M_{ni})\} = b_n\{1 - \hat{F}_n(M_{ni})\} \frac{n + 1}{n - b_n + 1} = \hat{Z}_{ni} \frac{n + 1}{n - b_n + 1},$$

some simple calculations show that, for instance for the CFG-estimator,

$$e^{-\gamma} \exp \left\{ -\frac{1}{k_n} \sum_{i=1}^{k_n} \log(\tilde{Z}_{ni}) \right\} = \frac{n - b_n + 1}{n + 1} \hat{\theta}_{\text{db,CFG}}^{z_n},$$

showing that the modification is asymptotically negligible. It is however beneficial in finite-sample situations, whence it has been applied throughout the finite-sample situations considered in Section 6. Obviously, similar adaptations can be applied to the sliding blocks version and the other moment based estimators.

### 5. Example: max-autoregressive process

In this section, we exemplarily discuss the new estimators when applied to a max-autoregressive process, defined by the recursion

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

where  $\alpha \in [0, 1)$  and where  $(Z_s)_{s \in \mathbb{Z}}$  is an i.i.d. sequence of Fréchet(1)-distributed random variables. A stationary solution of the above recursion is

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

such that the stationary solution is again Fréchet(1)-distributed. Note that a model with an arbitrary stationary c.d.f.  $F$  may be obtained by considering  $\tilde{X}_s = F^{\leftarrow}\{\exp(-1/X_s)\}$  and that all subsequent results are also valid for  $(\tilde{X}_s)_s$ .

We start by explicitly calculating the asymptotic variances of the estimators in Section 5.1, and then show in Section 5.2 that all regularity conditions from Section 3 are met.

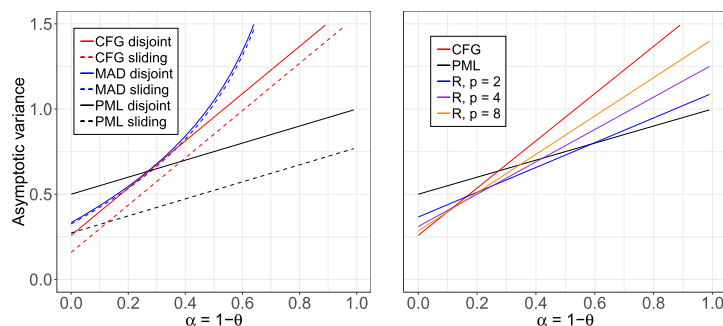


FIG 2. Asymptotic variance of  $\sqrt{k_n}(\hat{\theta}_n/\theta - 1)$  in the ARMAX-model. The estimators in the right figure rely on disjoint blocks.

### 5.1. Asymptotic variances for the ARMAX-model

Recall that the ARMAX-model has extremal index  $\theta = 1 - \alpha$  and that the corresponding cluster size distribution is geometric, that is,  $\pi(j) = \alpha^{j-1}(1 - \alpha)$ ,  $j \geq 1$ , see, e.g., Chapter 10 in [1]. From Example 6.1 in [3], one further has

$$\begin{aligned} \mathbb{E}[\xi_1^{(z)} \xi_2^{(z)}] &= \frac{\alpha^{w+1} + z + zw(1 - \alpha)}{(1 - \alpha)^2}, \\ \mathbb{E}[\xi_1^{(z)} \mathbf{1}(\xi_2^{(z)} = 0)] &= \frac{1 - \alpha^{w+1}}{1 - \alpha} - z(w + 1), \end{aligned}$$

where  $w = \lfloor \log(z)/\log(\alpha) \rfloor$  and  $(\xi_1^{(z)}, \xi_2^{(z)}) \sim \pi_2^{(z)}$ . This allows to calculate the limiting variances in Theorem 4.1–4.3 explicitly. For the CFG-type estimator, some tedious but straightforward calculations imply

$$\frac{\sigma_{\text{db,C}}^2}{\theta^2} = \frac{\pi^2}{6} + 2 \log(2)(\alpha - 1) \quad \text{and} \quad \frac{\sigma_{\text{sb,C}}^2}{\theta^2} = 2 \log(2)(3 + \alpha) - 4,$$

see also Figure 2 for a picture of the graph of these functions. Next, we compare these variances with the disjoint and sliding blocks variances of the PML-estimator in [3], which are given by  $\sigma_{\text{db},1}^2$  and  $\sigma_{\text{sb},1}^2$  and satisfy

$$\frac{\sigma_{\text{db},1}^2}{\theta^2} = \frac{1}{2}(1 + \alpha) \quad \text{and} \quad \frac{\sigma_{\text{sb},1}^2}{\theta^2} = \frac{8 \log(2) - 5 + \alpha}{2},$$

respectively. Thus,  $\sigma_{\text{db,C}}^2 \leq \sigma_{\text{db},1}^2$  iff  $\alpha \leq \{1 + 4 \log(2) - \pi^2/3\}/\{4 \log(2) - 1\} \approx 0.2723$  and  $\sigma_{\text{sb,C}}^2 \leq \sigma_{\text{sb},1}^2$  iff  $\alpha \leq \{3 - 4 \log(2)\}/\{4 \log(2) - 1\} \approx 0.128$ .

Further comparisons can be drawn from Figure 2, where the asymptotic variances of  $\sqrt{k_n}(\hat{\theta}_n/\theta - 1)$  are additionally illustrated for the Madogram- and the Root-estimators.

**5.2. Regularity conditions for the ARMAX-model**

Recall that  $X_s$  is Fréchet(1)-distributed, i.e., the stationary c.d.f.  $F$  is given by  $F(x) = \exp(-1/x), x > 0$ , with inverse  $F^{-1}(x) = -\log(x)^{-1}$ .

The assumptions in Condition 3.1 are satisfied as shown in [3], page 2322, provided  $b_n$  and  $k_n$  are chosen to satisfy the conditions in item (iii). Next, by induction,

$$\mathbb{P}\left(\max_{s=1,\dots,b} X_s \leq x\right) = F(x)^{1+\theta(b-1)},$$

which implies that the c.d.f. of  $Z_{1:b} = b\{1 - F(M_{1:b})\}$  is given by

$$\mathbb{P}(Z_{1:b} \leq x) = \begin{cases} 1, & x \geq b, \\ 1 - \left(1 - \frac{x}{b}\right)^{1+\theta(b-1)}, & x \in [0, b], \\ 0, & b \leq 0. \end{cases} \tag{5.1}$$

A tedious but straightforward calculation then shows that the assumptions in Condition 3.2 and 3.3 are met, provided  $k_n/b_n^2 = o(1)$ , cf. Condition 3.1(iii). Condition 3.4(i) is a condition on the choice of  $b_n$ , that is under the control of the statistician. Conditions 3.4(ii) and 3.4(iii) are consequences of mixing properties of  $(X_s)_s$  as argued at the end of Section 3. It remains to show that Condition 3.4(iv) is satisfied. By (5.1) and with  $\xi \sim \text{Exp}(\theta)$ , we have

$$\begin{aligned} \mathbb{P}(Z_{n1} < ck_n^{-\mu}) - \mathbb{P}(\xi < ck_n^{-\mu}) &= \exp(-\theta ck_n^{-\mu}) - \left(1 - \frac{ck_n^{-\mu}}{b_n}\right)^{1+\theta(b_n-1)} \\ &= o(k_n^{-1/2}(\log n)^{-1}), \quad n \rightarrow \infty, \end{aligned}$$

for any  $\mu > 1/2$ , where the final estimate follows from Taylor’s theorem and Condition 3.4(i).

**6. Finite-sample results**

A Monte-Carlo simulation study was performed to assess the finite-sample performance of the introduced estimators and to compare them with competing estimators from the literature. The data is simulated from the following four time series models that were also investigated in [3]:

- The **ARMAX-model**:

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

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**:

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

where  $\lambda \in (0, 1)$  and where  $(Z_s)_s$  denotes an i.i.d. sequence of standard normal random variables. We consider  $\lambda = 0.1, 0.5, 0.9, 0.99$  for which the simulated values  $\theta = 0.997, 0.727, 0.460, 0.422$  were obtained, respectively; see Table 3.1 in [8].

- The **ARCH-model**:

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

where  $\lambda \in (0, 1)$  and where  $(Z_s)_s$  denotes an i.i.d. sequence of standard normal random variables. We consider  $\lambda = 0.1, 0.5, 0.7, 0.99$  for which the simulated values  $\theta = 0.999, 0.835, 0.721, 0.571$  were obtained, respectively; see Table 3.2 in [8].

- The **Markovian Copula-model** ([7]):

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

Here,  $F^{\leftarrow}$  is the left-continuous quantile function of some arbitrary continuous c.d.f.  $F$ ,  $(U_s)_s$  is a stationary Markovian time series of order 1 and  $C_{\vartheta}$  denotes the Survival Clayton Copula with parameter  $\vartheta > 0$ . We consider choices  $\vartheta = 0.23, 0.41, 0.68, 1.06, 1.90$  such that (approximately)  $\theta = 0.95, 0.8, 0.6, 0.4, 0.2$  [3] and fix  $F$  as the standard uniform c.d.f. (the results are independent of this choice, as the estimators are rank-based). Algorithm 2 in [26] allows to simulate from this model.

In each case, the sample size is fixed to  $n = 2^{13} = 8192$  and the block size is chosen from  $b = b_n \in \{2^2, \dots, 2^9\}$ . The performance is assessed based on  $N = 3000$  simulation runs each.

### 6.1. Comparison of the introduced estimators

We start by comparing the finite-sample properties of the proposed sliding blocks estimators  $\hat{\theta}_{m,\text{CFG}}^x$ ,  $\hat{\theta}_{m,\text{MAD}}^x$  and  $\hat{\theta}_{m,\text{R,p}}^x$  for  $p \in \{0.5, 0.75, 1, 2, 4, 8, 16\}$  for  $x \in \{z_n, y_n\}$  and for  $m \in \{\text{sb}, \text{db}\}$ .

As the simulation results are, to a large extent, similar among the different models and estimators, they are only partially reported, with a particular view on highlighting selected interesting qualitative features. We begin by a detailed investigation of the variance, the squared bias and the mean squared error (MSE) as a function of the block size parameter  $b$ . In Figure 3, we present results for the disjoint and sliding blocks version of the CFG- and the PML-estimator in a representative ARMAX-model with  $\theta = 0.75$ . Similarly as in [3] and as to be expected from the asymptotic results, the bias of the disjoint and the sliding blocks version are almost identical, while the variance is uniformly smaller for the sliding blocks version (in particular for large values of  $b_n$ ). Since this qualitative behavior holds uniformly over all models and estimators, we omit the disjoint blocks estimator from the subsequent discussions and write  $\hat{\theta}_{\text{CFG}}^x = \hat{\theta}_{\text{sb,CFG}}^x$  etc. for simplicity.



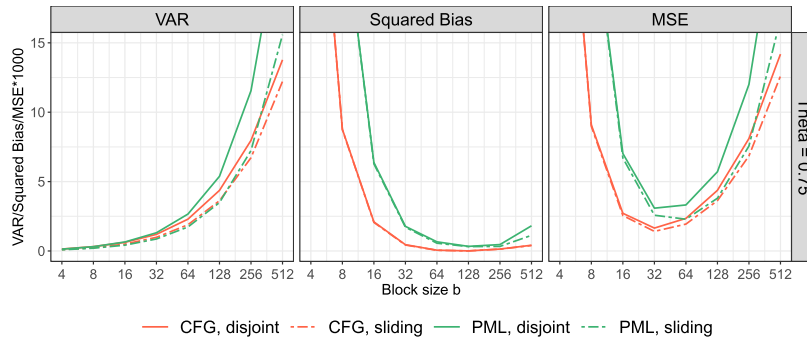


FIG 3. Comparison of variance, squared bias and MSE, multiplied by  $10^3$ , of the disjoint and sliding blocks CFG- and PML-estimator in the ARMAX-model.

Next, we compare the different moment estimators. For illustrative purposes, we begin by restricting the presentation to the  $z_n$ -versions and the ARCH-model. The corresponding results are depicted in Figure 4 (for the CFG-, the Madogram- and three selected Root-estimators). In general, as to be expected from the underlying theory, the variance curves are increasing in  $b$ , while the squared bias curves are (mostly) decreasing in  $b$ , resulting in a typical U-shape for the MSE curves. The hierarchy of the estimators with regard to the considered performance measures is similar among the considered values of  $\theta$ . In terms of the MSE, up to an intermediate block size, the CFG- and Madogram-estimator are superior to the other estimators (especially to the PML-estimator), while for large block sizes the Madogram-estimator has a relatively high MSE, but the CFG-estimator partly remains superior. The Root-estimators are, as expected, ordered in  $p$  and located between the PML- and CFG-estimator.

Next, a comparison between the  $z_n$ - and  $y_n$ -versions of the estimators is drawn in Figure 5; for illustrative purposes, attention is restricted to six different models and two estimators. Remarkably, there are many models, especially for smaller values of  $\theta$ , in which the MSE-curves of the  $y_n$ -versions lie uniformly below the ones of the  $z_n$ -versions. In the remaining models, neither version can be said to be strictly preferable. Furthermore, it is remarkable that, for  $\theta$  close to one, the MSE-curves of the  $y_n$ -versions are often no longer U-shaped, but increasing in the block size instead. The latter behavior may be explained by the proximity to the i.i.d. case, since in that case, we have

$$\mathbb{P}(Y_{1:b} \geq y) = \mathbb{P}(N_{1:b} \leq e^{-y/b}) = \mathbb{P}(U_1 \leq e^{-y/b})^b = e^{-y}$$

for all  $b \in \mathbb{N}$ , such that there is real equality in relation (1.3), resulting in a vanishing bias.

Next, we investigate the dependence of the performance of the Root-estimators on the parameter  $p$ ; recall that  $p = 1$  yields the PML-estimator, while ‘ $p = \infty$ ’ yields the CFG-estimator. In Figure 6, the MSE-curves are depicted

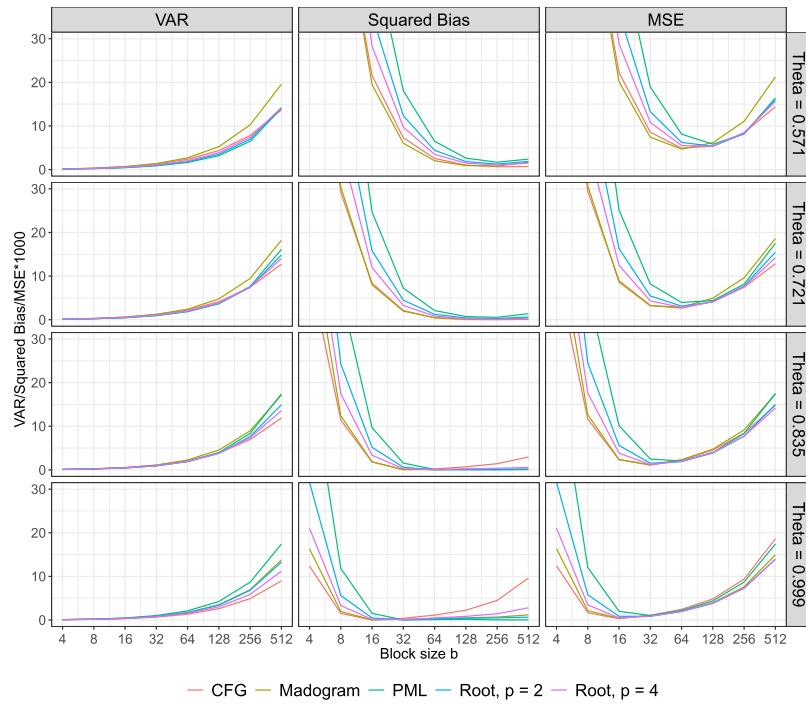


FIG 4. Variance, squared bias and MSE, multiplied by  $10^3$ , for the estimation of  $\theta$  within the ARCH-model for four values of  $\theta$ .

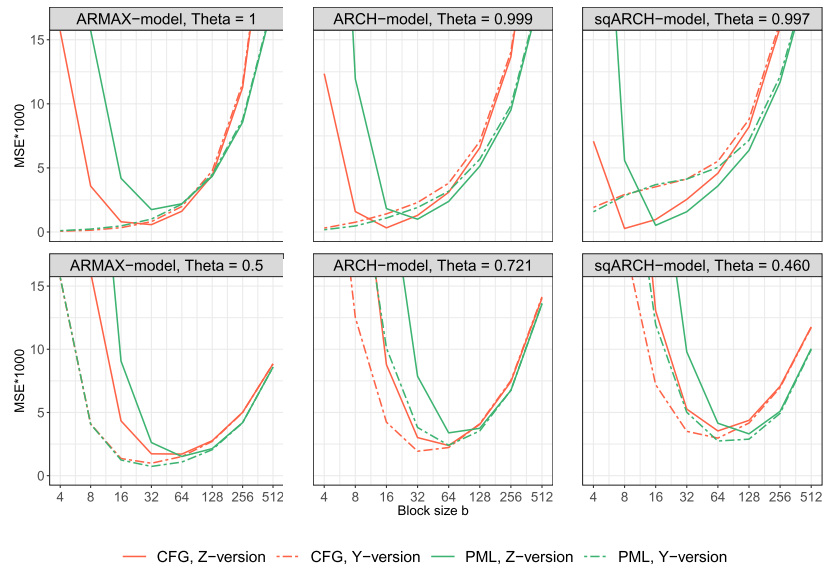


FIG 5. Comparison of the MSE multiplied by  $10^3$  of the  $z_n$ - and  $y_n$ -versions of the estimators.

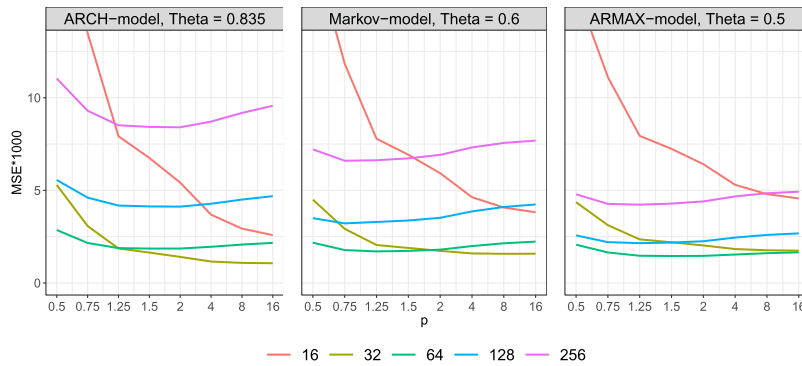


FIG 6. Mean Squared Error multiplied by  $10^3$  of the Root-estimators as a function of the parameter  $p$  for block sizes  $b \in \{16, 32, 64, 128, 256\}$  and three different models.

TABLE 1

Identification of the Root-estimator  $p$  with the minimum MSE for the ARCH- and ARMAX-model and every considered block size  $b$ . The  $p$  with the minimum MSE over all block sizes is presented in the last line.

Model	ARCH				ARMAX				
	Theta	0.999	0.835	0.721	0.571	1	0.75	0.5	0.25
$b = 4$		$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
8		$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
16		$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
32		2	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$
64		2	2	$\infty$	$\infty$	16	8	1.5	2
128		2	1.5	4	4	8	4	1	1
256		2	4	$\infty$	1.25	4	8	1	0.75
512		2	8	$\infty$	$\infty$	4	$\infty$	1	0.75
$\min_b$		$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	1.5	1

as a function of  $p$  for various fixed block sizes and for three selected models. It can be seen that choices of  $p < 1$  lead to a poor behavior of the corresponding estimators. At the same time, the results do not allow to identify some ‘optimal’ choice of  $p \geq 1$  which is valid uniformly over all models. A similar conclusion can be drawn from Table 1, which presents, for the ARCH- and ARMAX-model and every block size  $b$ , the value of  $p$  for which the Root-estimator attains the minimal MSE ( $p = \infty$  corresponds to the CFG-estimator). One can see that most values of  $p$  are represented, with  $p = \infty$  appearing most often, but that there is no optimal choice of  $p$  universally over all models.

### 6.2. Comparison with other estimators for the extremal index

In this section, we compare the performance of the introduced new estimators with the following estimators: the bias-reduced sliding blocks estimator from [28] (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 [27], the intervals

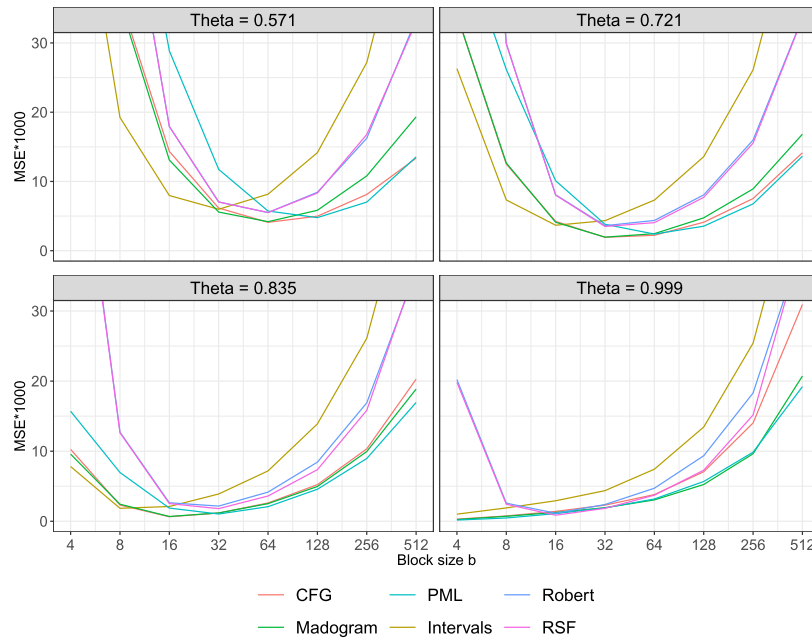


FIG 7. Mean Squared Error multiplied by  $10^3$  for the estimation of  $\theta$  within the ARCH-model for four values of  $\theta$ .

estimator from [13] and the ML-estimator from [32]. The parameters  $\sigma$  and  $\phi$  for the Robert-estimator (cf. page 276 of [27]) are chosen as  $\sigma = 0.7$  and  $\phi = 1.3$ . In the case of the intervals- and Suveges-estimator, the choice of a threshold  $u$  is required, which is here chosen as the  $1 - 1/b_n$  empirical quantile of the observed data. With regard to our estimators, we present results for the sliding-blocks, bias-reduced and  $z_n$ -versions, if not indicated otherwise.

In Figure 7, we depict the MSE as a function of the block size  $b$ . For most models, the MSE-curves of the estimators from the literature are again U-shaped due to the bias-variance tradeoff already described in Section 6.1. It can further be seen that no estimator is uniformly best in any model under consideration. The method-of-moment estimators do however compare quite well to the competitors.

The minimum values of the MSE-curves in Figure 7 are of particular interest. Due to the large amount of estimators and models under consideration (in total 26 estimators and 17 models) we try to simplify possible comparisons by the following aggregation, summarized in Table 2. First, in the first four columns of the table, we calculate for each time series model and each estimator under consideration, the sum (sum over all values of  $\theta$  considered for the specific model) of the minimum MSE-values (minimum over  $b$ ). Second, in the last four columns of the table, we present the sum of the minimum MSE-values (minimum over  $b$ ) over all models, for which the extremal index  $\theta$  lies in the in-

TABLE 2  
 Sum of minimal Mean Squared Error multiplied by  $10^3$  over different models and  $\theta_1 \in (0, 0.3], \theta_2 \in (0.3, 0.6], \theta_3 \in (0.6, 0.8]$  and  $\theta_4 \in (0.8, 1]$ . The three smallest values per column are in boldface.

Estimator	armax	arch	arch <sup>2</sup>	markov	(0, .3]	(.3, .6]	(.6, .8]	(.8, 1]
CFG, Z	4.80	8.54	8.46	11.19	5.84	19.08	5.46	<b>2.61</b>
CFG, Y	2.56	<b>6.98</b>	8.41	12.63	5.08	15.45	<b>3.56</b>	6.49
Madogram, Z	5.17	8.87	7.92	10.77	5.66	18.12	5.68	3.27
Madogram, Y	3.00	<b>7.08</b>	8.62	12.65	5.10	15.72	<b>3.59</b>	6.94
PML, Z	6.18	11.74	7.99	10.89	4.44	18.62	7.37	6.38
PML, Y	<b>1.96</b>	8.40	7.45	10.99	<b>3.73</b>	14.83	4.04	6.21
R, p = 0.5, Z	9.64	17.37	11.57	12.11	4.90	24.18	11.25	10.35
R, p = 0.5, Y	2.33	11.99	8.49	10.94	3.90	18.14	5.66	6.05
R, p = 0.75, Z	7.08	13.33	8.83	10.99	4.44	19.80	8.79	7.20
R, p = 0.75, Y	2.03	9.26	7.63	10.74	<b>3.66</b>	15.53	4.41	6.06
R, p = 1.25, Z	5.77	11.02	7.82	10.80	4.56	18.33	6.61	5.89
R, p = 1.25, Y	<b>1.96</b>	8.06	<b>7.37</b>	11.04	3.74	<b>14.47</b>	3.90	6.32
R, p = 1.5, Z	5.54	10.48	7.86	<b>10.47</b>	4.72	18.38	6.21	5.04
R, p = 1.5, Y	1.98	7.93	<b>7.32</b>	11.10	3.76	<b>14.32</b>	3.84	6.40
R, p = 2, Z	5.22	9.82	8.11	<b>10.22</b>	4.84	18.67	5.76	4.10
R, p = 2, Y	2.03	7.88	<b>7.34</b>	11.16	3.84	<b>14.34</b>	3.72	6.51
R, p = 4, Z	4.84	9.10	8.40	<b>10.14</b>	5.07	18.81	5.39	3.20
R, p = 4, Y	2.20	7.52	7.64	11.58	4.21	14.53	3.67	6.52
R, p = 8, Z	4.76	8.88	8.42	10.48	5.37	18.96	5.36	<b>2.85</b>
R, p = 8, Y	2.35	7.31	7.95	12.02	4.56	14.91	3.68	6.48
R, p = 16, Z	4.76	8.69	8.41	10.78	5.58	18.99	5.39	<b>2.68</b>
R, p = 16, Y	2.45	<b>7.14</b>	8.16	12.32	4.80	15.18	<b>3.61</b>	6.47
Intervals	3.49	12.53	11.72	21.86	<b>3.60</b>	15.55	11.46	18.98
ML Süveges	<b>1.90</b>	22.67	8.70	25.20	14.93	30.46	4.95	8.13
Robert	8.54	12.45	9.97	13.61	6.46	22.42	8.34	7.34
RSF	8.09	11.68	9.77	15.85	7.28	23.52	7.52	7.06

terval  $(0, 0.3], (0.3, 0.6], (0.6, 0.8]$  or  $(0.8, 1]$ , respectively. It can be seen that the CFG-estimator wins thrice, the Madogram- and PML-estimator wins twice, the Süveges and the Intervals-estimator wins once, and that the remaining smallest values are covered by a version of the Root-estimator. Also note that for large values of  $\theta \in (0.8, 1]$  (last column), the CFG-estimator and the Root-estimator for  $p \in \{8, 16\}$  are the best performing estimators. As a final interesting observation, note that the  $y$ -versions of the moment estimators mostly outperform the  $z$ -version, except for the column corresponding to  $\theta \in (0.8, 1]$  and some entries in the columns ‘Markov’ and ‘sqARCH’. A more refined analysis showed that these differences were almost exclusively attributable to the two specific models ‘Markov( $\theta = 0.95$ )’ and ‘sqARCH( $\theta = 0.997$ )’, which appear to be rather difficult to estimate for all estimators under consideration.

### 7. Conclusion

Estimating the extremal index is a classical problem in extreme value analysis for univariate stationary time series, with many ad-hoc solutions based on diverse motivations. This paper considers a new approach that is based on certain rescaled samples of ranks of block maxima and the method of moment

principle. The underlying samples have also been used by [22] and [3] to define explicit (pseudo) maximum likelihood estimators for the extremal index. Using the method of moment principle instead results in a large variety of alternative estimators. Studying their properties was initially motivated by the fact that a similar approach in multivariate extremes (the rank-based CFG-estimator for the Pickands function) was found to yield a more efficient estimator than the (pseudo) maximum likelihood method [14].

The method of moment principle being a rather universal principle, the present paper goes far beyond only considering a CFG-type estimator. In fact, based on natural moment equations for the exponential distribution (see Section 2.3), three classes of method of moment estimators were considered, which may each be based on (1) either disjoint or sliding block maxima, and (2) on certain  $y$ - or  $z$ -transformations of the block maxima. The sliding blocks version was always found to be more efficient than the disjoint blocks version. The  $y$ - and  $z$ -version share a similar behavior in terms of their asymptotic variances, but their bias may differ substantially depending on the underlying data generating process. The initial conjecture derived from [14] was partially confirmed: for  $\theta$  in an explicit neighbourhood of 1, the asymptotic variance of the CFG-type estimator is always smaller than the one of the ML-type estimator. A comparison between the various method of moment estimators is more cumbersome, with no universal answer, neither theoretically nor in terms of simulated finite sample results. If one were to come up with a single proposal, then the simulation study overall suggests to use the sliding blocks  $y$ -version of the root-estimator with an intermediate choice of  $p$ , say,  $p = 1.25$ .

In comparison with many other estimators for the extremal index, the proposed estimators have the advantage of being based on only one parameter to be chosen by the statistician, namely the block size  $b$ . Moreover, the estimators perform equally well or even better in some typical finite sample situations.

Finally, this work leaves some interesting questions for future research: (1) what is the minimal asymptotic variance that can be achieved by estimators based on the considered rank-based samples? (2) More generally, are there estimators for the extremal index that are semiparametrically efficient? (3) Can the sliding blocks method be used to derive more efficient estimators for the cluster size distribution, for instance by generalizing the disjoint blocks versions in [27]?

## Appendix A: Proofs of Theorems 4.1–4.3

The proofs of Theorems 4.1–4.3 are actually quite similar in that each proof will be decomposed into a sequence of similar intermediate lemmas. Occasionally, those lemmas will be hardest to prove for Theorem 4.1 and easiest to prove for Theorem 4.2; this is also reflected by the larger number of conditions required for the proof of Theorem 4.1. The proof of Theorem 4.3 in turn is quite similar to the one in [3], and of intermediate difficulty. For the above reasons, we will carry out the proof of Theorem 4.1 in great detail (Section A.1), and skip

parts of the technical arguments needed for Theorem 4.2 and 4.3 where possible (Sections A.2 and A.3). Intermediate, but less central results for the proof of Theorem 4.1 are given in Sections B.1, B.2 and B.3.

All convergences are for  $n \rightarrow \infty$  if not stated otherwise.

**A.1. Proof of Theorem 4.1**

The following notations will be used throughout:

$$\begin{aligned} \hat{S}_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \log(\hat{Z}_{ni}), & S_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \log(Z_{ni}), \\ \hat{S}_n^{\text{sb}} &= \frac{1}{n - b_n + 1} \sum_{i=1}^{n - b_n + 1} \log(\hat{Z}_{ni}^{\text{sb}}), & S_n^{\text{sb}} &= \frac{1}{n - b_n + 1} \sum_{i=1}^{n - b_n + 1} \log(Z_{ni}^{\text{sb}}). \end{aligned}$$

Note that  $\hat{\theta}_{\text{db,CFG}}^{z_n} = \varphi_{(C)}^{-1}(\hat{S}_n)$  and  $\hat{\theta}_{\text{sb,CFG}}^{z_n} = \varphi_{(C)}^{-1}(\hat{S}_n^{\text{sb}})$ , where  $\varphi_{(C)}^{-1}(x) = \exp\{-(x + \gamma)\}$ . Observing that  $(\varphi_{(C)}^{-1})' \{\varphi_{(C)}(\theta)\} = \theta$ , the two assertions of the theorem are a consequence of the delta-method and Proposition A.1 and Proposition A.2, respectively.  $\square$

**Proposition A.1.** *Under Condition 3.1, 3.2(i), 3.3(i) and 3.4, we have*

$$\sqrt{k_n} \{\hat{S}_n - \varphi_{(C)}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{db,C}}^2 / \theta^2) \quad \text{as } n \rightarrow \infty.$$

*Proof.* We may decompose

$$\sqrt{k_n} \{\hat{S}_n - \varphi_{(C)}(\theta)\} = A_n + B_n + C_n,$$

where

$$A_n = \sqrt{k_n} \{\hat{S}_n - S_n\}, \quad B_n = \sqrt{k_n} \{S_n - \mathbb{E}(S_n)\}, \quad C_n = \sqrt{k_n} \{\mathbb{E}(S_n) - \varphi_{(C)}(\theta)\}.$$

We have  $C_n = o(1)$  by Condition 3.3(i). For the treatment of  $A_n$ , recall the tail empirical process defined in (3.1). Further, let  $\tilde{N}_{ni} = (n + 1)/n \times \hat{N}_{ni}$ , and note that

$$\begin{aligned} 1 - \tilde{N}_{ni} &= \frac{1}{n} \sum_{s=1}^n \mathbb{1}(X_s > M_{ni}) \\ &= \frac{1}{n} \sum_{s=1}^n \mathbb{1}\left(U_s > 1 - \frac{Z_{ni}}{b_n}\right) \\ &= \frac{\sqrt{k_n}}{n} \frac{1}{\sqrt{k_n}} \sum_{s=1}^n \left\{ \mathbb{1}\left(U_s > 1 - \frac{Z_{ni}}{b_n}\right) - \frac{Z_{ni}}{b_n} \right\} + \frac{Z_{ni}}{b_n} \\ &= \frac{\sqrt{k_n}}{n} e_n(Z_{ni}) + \frac{Z_{ni}}{b_n}. \end{aligned} \tag{A.1}$$

Finally, let

$$\hat{H}_{k_n}(x) := \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbf{1}(Z_{ni} \leq x) \quad (\text{A.2})$$

denote the empirical c.d.f. of  $Z_{n1}, \dots, Z_{nk_n}$ . By Equation (A.1), we obtain

$$\begin{aligned} A_n &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log(1 - \hat{N}_{ni}) - \log(Z_{ni} b_n^{-1}) \\ &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \left\{ \frac{n}{n+1} \left( \frac{1}{n} + 1 - \hat{N}_{ni} \right) \right\} - \log \left( \frac{Z_{ni}}{b_n} \right) \\ &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left[ \log \left\{ \frac{1}{n} + \frac{\sqrt{k_n}}{n} e_n(Z_{ni}) + \frac{Z_{ni}}{b_n} \right\} - \log \left( \frac{Z_{ni}}{b_n} \right) + \log \left( \frac{n}{n+1} \right) \right] \\ &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \log \left\{ 1 + \frac{\sqrt{k_n} b_n}{n} \cdot \frac{e_n(Z_{ni})}{Z_{ni}} + \frac{b_n}{n Z_{ni}} \right\} + \sqrt{k_n} \log \left( \frac{n}{n+1} \right) \\ &= \int_0^\infty W_n(x) d\hat{H}_{k_n}(x) + o(1), \end{aligned} \quad (\text{A.3})$$

where

$$W_n(x) = \sqrt{k_n} \log \left\{ 1 + \frac{1}{\sqrt{k_n}} \left( \frac{e_n(x)}{x} + \frac{1}{\sqrt{k_n} x} \right) \right\}.$$

Heuristically,  $\hat{H}_{k_n}(x) \approx 1 - \exp(-\theta x)$  and  $W_n(x) \approx e(x)/x$  (where  $e$  denotes the limit of the tail empirical process), whence the tentative limit of  $A_n$  should be

$$A = \int_0^\infty \frac{e(x)}{x} \theta e^{-\theta x} dx.$$

For a rigorous treatment of  $A_n + B_n$ , let

$$\begin{aligned} E_n &= \int_0^\infty W_n(x) d\hat{H}_{k_n}(x), & E_{n,m} &= \int_{1/m}^m W_n(x) d\hat{H}_{k_n}(x), \\ E'_m &= \int_{1/m}^m \frac{e(x)}{x} \theta e^{-\theta x} dx \end{aligned}$$

and let  $B$  be defined as in Lemma B.1 below. As shown above,  $A_n = E_n + o(1)$ . The proposition is hence a consequence of Wichura's theorem ([4], Theorem 25.5) and the following items:

- (i) For all  $m \in \mathbb{N}$ :  $E_{n,m} + B_n \xrightarrow{d} E'_m + B$  as  $n \rightarrow \infty$ .
- (ii)  $E'_m + B \xrightarrow{d} A + B \sim \mathcal{N}(0, \sigma_{\text{db},C}^2/\theta^2)$  as  $m \rightarrow \infty$ .
- (iii) For all  $\delta > 0$ :  $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|E_n - E_{n,m}| > \delta) = 0$ .

The assertion in (i) is proven in Lemma B.4. The assertion in (ii) follows from the fact that  $E'_m + B$  is normally distributed with variance  $\tau_m^2$  as specified in Lemma B.4, and the fact that  $\tau_m^2 \rightarrow \sigma_{\text{db},C}^2/\theta^2$  as  $m \rightarrow \infty$  by Lemma B.5. Finally, Lemma B.6 proves (iii).  $\square$



**Proposition A.2.** *Under Condition 3.1, 3.2(i), 3.3(i) and 3.4, we have*

$$\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - \varphi_{(\text{C})}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{sb,C}}^2/\theta^2) \quad \text{as } n \rightarrow \infty.$$

*Proof.* The proof is very similar to the proof of Proposition A.1. Decompose

$$\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - g(\theta)\} = A_n^{\text{sb}} + B_n^{\text{sb}} + C_n^{\text{sb}},$$

where

$$\begin{aligned} A_n^{\text{sb}} &:= \sqrt{k_n}\{\hat{S}_n^{\text{sb}} - S_n^{\text{sb}}\}, & B_n^{\text{sb}} &:= \sqrt{k_n}\{S_n^{\text{sb}} - E[S_n^{\text{sb}}]\}, \\ C_n^{\text{sb}} &:= \sqrt{k_n}\{E[S_n^{\text{sb}}] - \varphi_{(\text{C})}(\theta)\}. \end{aligned}$$

Again, we have  $C_n^{\text{sb}} = o(1)$  by Condition 3.3(i). A similar calculation as in (A.3) in the case of the disjoint blocks shows that  $A_n^{\text{sb}}$  can be written in the following way

$$A_n^{\text{sb}} = \int_0^\infty W_n(x) \, d\hat{H}_n^{\text{sb}}(x) + o(1),$$

where

$$\hat{H}_n^{\text{sb}}(x) = \frac{1}{n - b_n + 1} \sum_{t=1}^{n-b_n+1} \mathbb{1}(Z_{nt}^{\text{sb}} \leq x)$$

denotes the empirical c.d.f. of  $Z_{n1}^{\text{sb}}, \dots, Z_{n,n-b_n+1}^{\text{sb}}$ . We may now treat  $A_n^{\text{sb}} + B_n^{\text{sb}}$  exactly as  $A_n + B_n$  in the proof of Proposition A.1, with  $E_n, E_{n,m}$  and Lemma B.4, B.5 and B.6 replaced by

$$E_n^{\text{sb}} = \int_0^\infty W_n(x) \, d\hat{H}_n^{\text{sb}}(x), \quad E_{n,m}^{\text{sb}} = \int_{1/m}^m W_n(x) \, d\hat{H}_n^{\text{sb}}(x),$$

and Lemma B.10, B.11 and B.12, respectively. □

### A.2. Proof of Theorem 4.2

The following notation will be used throughout:

$$\begin{aligned} \hat{S}_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \exp(-\hat{Z}_{ni}), & S_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \exp(-Z_{ni}), \\ \hat{S}_n^{\text{sb}} &= \frac{1}{n - b_n + 1} \sum_{i=1}^{n-b_n+1} \exp(-\hat{Z}_{ni}^{\text{sb}}), & S_n^{\text{sb}} &= \frac{1}{n - b_n + 1} \sum_{i=1}^{n-b_n+1} \exp(-Z_{ni}^{\text{sb}}). \end{aligned}$$

Note that  $\hat{\theta}_{\text{db,MAD}}^{z_n} = \varphi_{(\text{M})}^{-1}(\hat{S}_n)$  and  $\hat{\theta}_{\text{sb,MAD}}^{z_n} = \varphi_{(\text{M})}^{-1}(\hat{S}_n^{\text{sb}})$ , where  $\varphi_{(\text{M})}(x) = x/(1+x)$ . The assertion follows from the delta-method and Proposition A.3 and A.5. □

**Proposition A.3.** *Under Condition 3.1 and 3.3(ii), we have*

$$\sqrt{k_n}\{\hat{S}_n - \varphi_{(M)}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{db},M}^2/(1+\theta)^4) \quad \text{as } n \rightarrow \infty.$$

*Proof.* Write  $\sqrt{k_n}\{\hat{S}_n - \varphi_{(M)}(\theta)\} = A_n + B_n + C_n$ , where

$$A_n = \sqrt{k_n}\{\hat{S}_n - S_n\}, \quad B_n = \sqrt{k_n}\{S_n - E[S_n]\}, \quad C_n = \sqrt{k_n}\{E[S_n] - \varphi_{(M)}(\theta)\}.$$

The term  $C_n$  is asymptotically negligible by Condition 3.3(ii). A straightforward calculation shows that the summand  $A_n$  can be written in terms of the tail empirical process  $e_n$  as

$$A_n = \int_0^\infty W_n(x) \, d\hat{H}_{k_n}(x), \quad W_n(x) = \sqrt{k_n}e^{-x} \left[ \exp(-e_n(x)k_n^{-1/2}) - 1 \right],$$

where  $\hat{H}_{k_n}$  is the empirical c.d.f. of  $Z_{n1}, \dots, Z_{nk_n}$ , see (A.2). The asymptotic normality of  $A_n + B_n$  can now be shown as in the proof of Proposition A.1. The corresponding key result is given by Lemma A.4; whose proof is similar (but easier) as for the CFG-estimator (Lemma B.1) and is omitted for the sake of brevity.  $\square$

**Lemma A.4.** (a) *For any  $x_1, \dots, x_m \in [0, \infty)$ , as  $n \rightarrow \infty$ ,*

$$(e_n(x_1), \dots, e_n(x_m), B_n) \xrightarrow{d} (e(x_1), \dots, e(x_m), B) \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1}),$$

with

$$\Sigma_{m+1} = \begin{pmatrix} r(x_1, x_1) & \dots & r(x_1, x_m) & f(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \dots & r(x_m, x_m) & f(x_m) \\ f(x_1) & \dots & f(x_m) & \frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2} \end{pmatrix},$$

where the covariance function  $r$  is given as in Lemma B.1 and

$$f(x) = \sum_{i=1}^{\infty} i \int_0^1 p^{(x)}(i) - p_2^{(x, -\log(y))}(i, 0) \mathbf{1}(x \geq -\log(y)) \, dy - x\varphi_{(M)}(\theta).$$

(b) *For any  $x_1, \dots, x_m \in [0, \infty)$ , as  $n \rightarrow \infty$ ,*

$$(W_n(x_1), \dots, W_n(x_m), B_n) \xrightarrow{d} (-e^{-x_1}e(x_1), \dots, -e^{-x_m}e(x_m), B).$$

**Proposition A.5.** *Under Condition 3.1 and 3.3(ii), we have*

$$\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - \varphi_{(M)}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{sb},M}^2/(1+\theta)^4) \quad \text{as } n \rightarrow \infty.$$

*Proof.* The proof is similar to the proof of Proposition A.3. We may decompose  $\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - \varphi_{(M)}(\theta)\} = A_n^{\text{sb}} + B_n^{\text{sb}} + C_n^{\text{sb}}$ , where

$$A_n^{\text{sb}} = \sqrt{k_n}\{\hat{S}_n^{\text{sb}} - S_n^{\text{sb}}\}, \quad B_n^{\text{sb}} = \sqrt{k_n}\{S_n^{\text{sb}} - E[S_n^{\text{sb}}]\},$$

$$C_n^{\text{sb}} = \sqrt{k_n} \{E[S_n^{\text{sb}}] - \varphi_{(\text{M})}(\theta)\}.$$

Again, we have  $C_n^{\text{sb}} = o(1)$  by Condition 3.3(ii) and

$$A_n^{\text{sb}} = \int_0^\infty W_n(x) \, d\hat{H}_n^{\text{sb}}(x),$$

where  $\hat{H}_n^{\text{sb}}$  denotes the empirical c.d.f. of  $Z_{n1}^{\text{sb}}, \dots, Z_{n,n-b_n+1}^{\text{sb}}$ . The sum  $A_n^{\text{sb}} + B_n^{\text{sb}}$  can now be treated as in proof of Proposition A.2. The corresponding key result, Lemma B.7, needs to be replaced by Lemma A.6; whose proof is again omitted for the sake of brevity.  $\square$

**Lemma A.6.** (a) For any  $x_1, \dots, x_m \in [0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(e_n(x_1), \dots, e_n(x_m), B_n^{\text{sb}}) \xrightarrow{d} (e(x_1), \dots, e(x_m), B^{\text{sb}}) \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1}^{\text{sb}}),$$

where all entries of  $\Sigma_{m+1}^{\text{sb}}$  are the same as those of  $\Sigma_{m+1}$  in Lemma A.4 except for the entry at position  $(m+1, m+1)$ , which needs to be replaced by

$$v(\theta) = 2 - \frac{4}{\theta+1} + 4 \frac{\log(\theta+1) - \log(\theta+2) + \log(2)}{\theta(\theta+1)} - \frac{2\theta^2}{(\theta+1)^2}.$$

(b) For any  $x_1, \dots, x_m \in [0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(W_n(x_1), \dots, W_n(x_m), B_n^{\text{sb}}) \xrightarrow{d} (-e^{-x_1}e(x_1), \dots, -e^{-x_m}e(x_m), B^{\text{sb}}).$$

### A.3. Proof of Theorem 4.3

For fixed  $p > 0$ , define

$$\begin{aligned} \hat{S}_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} \hat{Z}_{ni}^{1/p}, & S_n &= \frac{1}{k_n} \sum_{i=1}^{k_n} Z_{ni}^{1/p}, \\ \hat{S}_n^{\text{sb}} &= \frac{1}{n-b_n+1} \sum_{i=1}^{n-b_n+1} \hat{Z}_{ni}^{1/p}, & S_n^{\text{sb}} &= \frac{1}{n-b_n+1} \sum_{i=1}^{n-b_n+1} Z_{ni}^{1/p}. \end{aligned}$$

Note that  $\hat{\theta}_{\text{db,R},p}^{z_n} = \varphi_{(\text{R}),p}^{-1}(\hat{S}_n)$  and  $\hat{\theta}_{\text{sb,R},p}^{z_n} = \varphi_{(\text{R}),p}^{-1}(\hat{S}_n^{\text{sb}})$ , where  $\varphi_{(\text{R}),p}(x) = x^{-1/p}\Gamma(1+1/p)$ . By the delta-method, the assertion follows from Proposition A.7 and A.9.  $\square$

**Proposition A.7.** Under Condition 3.1, 3.2(ii) and 3.3(iii), we have

$$\sqrt{k_n} \{\hat{S}_n - \varphi_{(\text{R}),p}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{db},p}^2 \psi_p(\theta)) \text{ as } n \rightarrow \infty,$$

where  $\psi_p(\theta) = \Gamma(1+1/p)^2 p^{-2} \theta^{-(2+2/p)}$ .

*Proof.* Decompose  $\sqrt{k_n}\{\hat{S}_n - \varphi_{(R),p}(\theta)\} = A_n + B_n + C_n$ , where

$$\begin{aligned} A_n &= \sqrt{k_n}\{\hat{S}_n - S_n\}, & B_n &= \sqrt{k_n}\{S_n - E[S_n]\}, \\ C_n &= \sqrt{k_n}\{E[S_n] - \varphi_{(R),p}(\theta)\}. \end{aligned}$$

By Condition 3.3(iii), the term  $C_n$  converges to zero. A straightforward calculation shows that the term  $A_n$  can be written as

$$A_n = \int_0^\infty W_n(x) d\hat{H}_{k_n}(x), \quad W_n(x) = \sqrt{k_n} \left\{ \left[ \frac{e_n(x)}{\sqrt{k_n}} + x \right]^{1/p} - x^{1/p} \right\}.$$

The asymptotic normality of  $A_n + B_n$  can be shown as in the proof of Proposition A.1 by an application of Wichura's theorem. Here, Lemma B.1 needs to be replaced by Lemma A.8, whose proof is similar but easier and therefore omitted for the sake of brevity.  $\square$

**Lemma A.8.** (a) For any  $x_1, \dots, x_m \in (0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(e_n(x_1), \dots, e_n(x_m), B_n) \xrightarrow{d} (e(x_1), \dots, e(x_m), B) \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1})$$

with

$$\Sigma_{m+1} = \begin{pmatrix} r(x_1, x_1) & \dots & r(x_1, x_m) & f_p(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \dots & r(x_m, x_m) & f_p(x_m) \\ f_p(x_1) & \dots & f_p(x_m) & v_p(\theta) \end{pmatrix},$$

where the covariance function  $r$  is defined as in Lemma B.1 and

$$\begin{aligned} f_p(x) &= \sum_{i=1}^\infty i \int_0^\infty p_2^{(x, y^p)}(i, 0) \mathbb{1}(x \geq y^p) dy - x\varphi_{(R),p}(\theta), \\ v_p(\theta) &= \theta^{-\frac{2}{p}} \{ \Gamma(1 + 2/p) - \Gamma(1 + 1/p)^2 \}. \end{aligned}$$

(b) For any  $x_1, \dots, x_m \in (0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(W_n(x_1), \dots, W_n(x_m), B_n) \xrightarrow{d} (e(x_1)x_1^{\frac{1}{p}-1}p^{-1}, \dots, e(x_m)x_m^{\frac{1}{p}-1}p^{-1}, B).$$

**Proposition A.9.** Under Condition 3.1, 3.2(ii) and 3.3(iii), we have

$$\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - \varphi_{(R),p}(\theta)\} \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{sb},p}^2 \psi_p(\theta)) \text{ as } n \rightarrow \infty,$$

where  $\psi_p(\theta) = \Gamma(1 + 1/p)^2 p^{-2} \theta^{-(2+2/p)}$ .

*Proof.* The proof is similar to the proof of Proposition A.7. Write  $\sqrt{k_n}\{\hat{S}_n^{\text{sb}} - \varphi_{(R),p}(\theta)\} = A_n^{\text{sb}} + B_n^{\text{sb}} + C_n^{\text{sb}}$ , where

$$A_n^{\text{sb}} = \sqrt{k_n}\{\hat{S}_n^{\text{sb}} - S_n^{\text{sb}}\}, \quad B_n^{\text{sb}} = \sqrt{k_n}\{S_n^{\text{sb}} - E[S_n^{\text{sb}}]\},$$

$$C_n^{\text{sb}} = \sqrt{k_n} \{E[S_n^{\text{sb}}] - \varphi_{(R),p}(\theta)\}.$$

By Condition 3.3(iii),  $C_n^{\text{sb}} = o(1)$ , and a straightforward calculation yields

$$A_n^{\text{sb}} = \int_0^\infty W_n(x) \, d\hat{H}_n^{\text{sb}}(x),$$

where  $\hat{H}_n^{\text{sb}}$  denotes the empirical c.d.f. of  $Z_{n1}^{\text{sb}}, \dots, Z_{n,n-b_n+1}^{\text{sb}}$ . The sum  $A_n^{\text{sb}} + B_n^{\text{sb}}$  can be treated as in the proof of Proposition A.2, where the main result, Lemma B.7, needs to be replaced by Lemma A.10, whose proof is omitted for the sake of brevity.  $\square$

**Lemma A.10.** (a) For any  $x_1, \dots, x_m \in (0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(e_n(x_1), \dots, e_n(x_m), B_n^{\text{sb}}) \xrightarrow{d} (e(x_1), \dots, e(x_m), B^{\text{sb}}) \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1}^{\text{sb}}),$$

where all entries of  $\Sigma_{m+1}^{\text{sb}}$  are the same as those of  $\Sigma_{m+1}$  in Lemma A.8 except for the entry at position  $(m+1, m+1)$ , which needs to be replaced by

$$v_p^{\text{sb}}(\theta) = 4p^{-2}\theta^{-2/p} \int_0^\infty (1 - e^{-z})z^{1/p-2}\Gamma(1/p, z) \, dz - 2\theta^{-2/p}\Gamma(1 + 1/p)^2.$$

(b) For any  $x_1, \dots, x_m \in (0, \infty)$ , as  $n \rightarrow \infty$ ,

$$(W_n(x_1), \dots, W_n(x_m), B_n^{\text{sb}}) \xrightarrow{d} (e(x_1)x_1^{\frac{1}{p}-1}p^{-1}, \dots, e(x_m)x_m^{\frac{1}{p}-1}p^{-1}, B^{\text{sb}}).$$

## Appendix B: Auxiliary results for the proof of Theorem 4.1

### B.1. Auxiliary lemmas – disjoint blocks

Throughout this section, we assume that Condition 3.1, 3.2(i) and 3.3(i) are met.

**Lemma B.1.** For any  $x_1, \dots, x_m \in [0, \infty)$  and  $m \in \mathbb{N}$ , we have

$$(e_n(x_1), \dots, e_n(x_m), B_n)' \xrightarrow{d} (e(x_1), \dots, e(x_m), B)',$$

where  $(e(x_1), \dots, e(x_m), B)' \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1})$  with

$$\Sigma_{m+1} = \begin{pmatrix} r(x_1, x_1) & \dots & r(x_1, x_m) & f(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \dots & r(x_m, x_m) & f(x_m) \\ f(x_1) & \dots & f(x_m) & \pi^2/6 \end{pmatrix}.$$

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

$$r(x, y) = \theta x \sum_{i=1}^\infty \sum_{j=0}^i ij\pi_2^{(y/x)}(i, j), \quad f(x) = h(x) - x\varphi_{(C)}(\theta),$$

$$h(x) = \sum_{i=1}^{\infty} i \left[ \int_0^{\infty} \mathbf{1}(e^y \leq x) p_2^{(x, e^y)}(i, 0) \, dy - \int_{-\infty}^0 p^{(x)}(i) - \mathbf{1}(e^y \leq x) p_2^{(x, e^y)}(i, 0) \, dy \right]$$

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

$$p_2^{(x, y)}(i, j) = \mathbb{P}(N_E^{(x, y)} = (i, j)), \quad N_E^{(x, y)} = \sum_{i=1}^{\eta} (\xi_{i1}^{(y/x)}, \xi_{i2}^{(y/x)})$$

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

$$p^{(x)}(i) = \mathbb{P}(N_E^{(x)} = i), \quad N_E^{(x)} = \sum_{i=1}^{\eta_2} \xi_i$$

with  $\eta_2 \sim \text{Poisson}(\theta x)$  independent of i.i.d. random variables  $\xi_i \sim \pi$ ,  $i \in \mathbb{N}$ .

**Lemma B.2.** For any  $m \in \mathbb{N}$ , we have

$$\{(W_n(x), B_n)'\}_{x \in [1/m, m]} \xrightarrow{d} \left\{ \left( \frac{e(x)}{x}, B \right)' \right\}_{x \in [1/m, m]} \quad \text{in } D([1/m, m]) \times \mathbb{R},$$

where  $(e, B)'$  is a centered Gaussian process with continuous sample paths and with covariance functional as specified in Lemma B.1.

**Lemma B.3.** For any  $m \in \mathbb{N}$ , we have

$$E_{n,m} = E'_{n,m} + o_{\mathbb{P}}(1) \quad \text{as } n \rightarrow \infty,$$

where  $E'_{n,m} = \int_{1/m}^m W_n(x) \theta e^{-\theta x} \, dx$ .

**Lemma B.4.** For any  $m \in \mathbb{N}$ , we have

$$E_{n,m} + B_n \xrightarrow{d} E'_m + B \sim \mathcal{N}(0, \tau_m^2) \quad \text{as } n \rightarrow \infty,$$

where, with  $r$  and  $f$  defined as in Lemma B.1,

$$\tau_m^2 = \theta^2 \int_{1/m}^m \int_{1/m}^m r(x, y) \frac{1}{xy} e^{-\theta(x+y)} \, dx dy + 2\theta \int_{1/m}^m f(x) \frac{1}{x} e^{-\theta x} \, dx + \frac{\pi^2}{6}.$$

**Lemma B.5.** As  $m \rightarrow \infty$ ,  $\tau_m^2 \rightarrow \sigma_{\text{db},(C)}^2 / \theta^2$ , where  $\sigma_{\text{db},(C)}^2$  is specified in Theorem 4.1.

**Lemma B.6.** If, in addition to Condition 3.1, 3.2(i) and 3.3(i), Condition 3.4 holds, then, for all  $\delta > 0$ ,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(|E_{n,m} - E_n| > \delta) = 0.$$

*Proof of Lemma B.1.* We proceed similarly as in the proof of Lemma 9.3 in [3]. Weak convergence of the  $(e_n(x_1), \dots, e_n(x_m))'$  is a consequence of Theorem 4.1 in [27]. For the treatment of the joint convergence with  $B_n$ , we only consider the case  $m = 1$  and set  $x_1 = x$ ; the general case can be treated analogously. For  $i = 1, \dots, k_n$ , we decompose a block  $I_i = \{(i-1)b_n + 1, \dots, ib_n\}$  into a big block  $I_i^+$  and a small block  $I_i^-$ , where, recalling  $\ell_n$  from Condition 3.1(iii),

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

and set

$$e_n^+(x) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \sum_{s \in I_i^+} \left\{ \mathbb{1}\left(U_s > 1 - \frac{x}{b_n}\right) - \frac{x}{b_n} \right\},$$

$$B_n^+ = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \left\{ \log(Z_{ni}^+) - \mathbb{E}[\log(Z_{ni}^+)] \right\},$$

where  $Z_{ni}^+ = b_n(1 - N_{ni}^+)$ ,  $N_{ni}^+ = \max_{s \in I_i^+} U_s$ . Next, according to Lemma 6.6 in [27],

$$e_n^-(x) := e_n(x) - e_n^+(x) = o_{\mathbb{P}}(1).$$

It can further be shown by the same arguments as in the proof of Lemma 9.3 in [3] that

$$B_n^- := B_n - B_n^+ = o_{\mathbb{P}}(1).$$

Finally, for  $\varepsilon \in (0, c_1 \wedge c_2)$ , define  $A_n^+ = \{\min_{i=1, \dots, k_n} N_{ni}^+ > 1 - \varepsilon\}$ , and note that  $\mathbb{P}(A_n^+) \rightarrow 1$  by Condition 3.1(v). As a consequence of the previous three statements, it suffices to show that, using the Cramér-Wold device,

$$\{\lambda_1 e_n^+(x) + \lambda_2 B_n^+\} \mathbb{1}_{A_n^+} \xrightarrow{d} \lambda_1 e(x) + \lambda_2 B, \tag{B.1}$$

for any  $\lambda_1, \lambda_2 \in \mathbb{R}$ .

Now, the left-hand side of (B.1) can be written as

$$\{\lambda_1 e_n^+(x) + \lambda_2 B_n^+\} \mathbb{1}_{A_n^+} = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} \tilde{g}_{i,n} + o_{\mathbb{P}}(1),$$

where  $\tilde{g}_{i,n} = g_{i,n} \mathbb{1}(Z_{ni}^+ < \varepsilon b_n)$  and where

$$g_{i,n} = \lambda_1 \sum_{s \in I_i^+} \left\{ \mathbb{1}\left(U_s > 1 - \frac{x}{b_n}\right) - \frac{x}{b_n} \right\} + \lambda_2 \left\{ \log(Z_{ni}^+) - \mathbb{E}[\log(Z_{ni}^+)] \right\}.$$

Note, that  $\tilde{g}_{i,n}$  only depends on the block  $I_i^+$  and is  $\mathcal{B}_{(i-1)b_n+1:ib_n-\ell_n}^\varepsilon$ -measurable. In particular, the  $(\tilde{g}_{i,n})_{i=1, \dots, k_n}$  are each separated by a small block of length  $\ell_n$ . A standard argument based on characteristic functions and the assumption on alpha mixing may then be used to show that the weak limit of  $k_n^{-1/2} \sum_{i=1}^{k_n} \tilde{g}_{i,n}$  is the same as if the  $\tilde{g}_{i,n}$  were independent.

Next, we show that Ljapunov's condition ([4], Theorem 27.3) is satisfied. By Minkowski's inequality, for any  $p \in (2, 2 + \delta)$ , we have  $C_\infty = \sup_{n \in \mathbb{N}} \mathbb{E}[|\tilde{g}_{1,n}|^p] < \infty$  by Condition 3.1(ii) and 3.2(i). Further, by stationarity and independence, we get

$$\frac{\sum_{i=1}^{k_n} \mathbb{E}[|\tilde{g}_{i,n}|^p]}{\text{Var}(\sum_{i=1}^{k_n} \tilde{g}_{i,n})^{p/2}} = k_n^{1-p/2} \frac{\mathbb{E}[|\tilde{g}_{1,n}|^p]}{(\mathbb{E}[\tilde{g}_{1,n}^2])^{p/2}} \leq C_\infty \times k_n^{1-p/2} \mathbb{E}[\tilde{g}_{1,n}^2]^{-p/2}.$$

Hence, provided  $\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{g}_{1,n}^2]$  exists, the last expression converges to 0 and hence Ljapunov's condition is met. As a consequence,  $k_n^{-1/2} \sum_{i=1}^{k_n} \tilde{g}_{i,n}$  weakly converges to a centered normal distribution with variance  $\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{g}_{1,n}^2]$ .

Finally, since  $\lim_{n \rightarrow \infty} \mathbb{E}[\tilde{g}_{1,n}^2] = \lim_{n \rightarrow \infty} \mathbb{E}[g_{1,n}^2]$ , it remains to be shown that

$$\lim_{n \rightarrow \infty} \mathbb{E}[g_{1,n}^2] = \lambda_1^2 r(x, x) + 2\lambda_1 \lambda_2 h(x) + \lambda_2^2 \pi^2 / 6.$$

Since similar arguments as in the proof of  $B_n^- = o_{\mathbb{P}}(1)$  and  $e_n^- = o_{\mathbb{P}}(1)$  allow us to replace  $I_1^+$  by  $I_1$  and then  $b_n$  by  $n$ , this in turn is a consequence of

$$\lim_{n \rightarrow \infty} \text{Var}(N_n^{(x)}(E)) = r(x, x), \quad (\text{B.2})$$

$$\lim_{n \rightarrow \infty} \text{Cov}\{N_n^{(x)}(E), \log(Z_{1:n})\} = f(x), \quad (\text{B.3})$$

$$\lim_{n \rightarrow \infty} \text{Var}\{\log(Z_{1:n})\} = \pi^2 / 6. \quad (\text{B.4})$$

The assertion in (B.2) follows from Theorem 4.1 in [27]. Further, since  $Z_{1:n} \xrightarrow{d} \xi \sim \text{Exp}(\theta)$  and since  $|\log(Z_{1:n})|^2$  is uniformly integrable by Condition 3.2(i), we have

$$\lim_{n \rightarrow \infty} \text{Var}\{\log(Z_{1:n})\} = \text{Var}\{\log(\xi)\} = \frac{\pi^2}{6},$$

which is (B.4). Finally, note that  $\mathbb{E}[N_n^{(x)}(E)] = x$  and  $\mathbb{E}[\log(Z_{1:n})] \rightarrow \varphi_{(C)}(\theta)$  by similar arguments as given above. As a consequence, (B.3) follows from  $\lim_{n \rightarrow \infty} \mathbb{E}[N_n^{(x)}(E) \log(Z_{1:n})] = h(x)$ . The latter in turn can be seen as follows: first,

$$\mathbb{E}[N_n^{(x)}(E) \log(Z_{1:n})] = \sum_{i=1}^n i \mathbb{E}[\mathbf{1}(N_n^{(x)}(E) = i) \log(Z_{1:n})]. \quad (\text{B.5})$$

The expected value on the right-hand side can be written as

$$\begin{aligned} & \int_0^\infty \mathbb{P}(\mathbf{1}(N_n^{(x)}(E) = i) \log(Z_{1:n}) > y) \, dy \\ & \quad - \int_{-\infty}^0 1 - \mathbb{P}(\mathbf{1}(N_n^{(x)}(E) = i) \log(Z_{1:n}) > y) \, dy \\ & = \int_0^\infty \mathbb{P}(N_n^{(x)}(E) = i, Z_{1:n} > e^y) \, dy \end{aligned}$$



$$- \int_{-\infty}^0 \mathbb{P}(N_n^{(x)}(E) = i) - \mathbb{P}(N_n^{(x)}(E) = i, Z_{1:n} > e^y) \, dy.$$

Now,

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

and  $\mathbb{P}(N_n^{(x)}(E) = i) \rightarrow p^{(x)}(i)$ , see [24] and [27]. By uniform integrability we obtain that the expected value on the right-hand side of (B.5) converges to  $h(x)$ . The proof is finished.  $\square$

*Proof of Lemma B.2.* For fixed  $x > 0$ , consider the function

$$f_n : \mathbb{R} \rightarrow \mathbb{R}, f_n(z) = \sqrt{k_n} \log \left\{ 1 + \frac{1}{\sqrt{k_n}} \left( \frac{z}{x} + \frac{1}{\sqrt{k_n}x} \right) \right\}.$$

For  $z_n \rightarrow z$ , one has  $f_n(z_n) \rightarrow e(z)/z$ . Hence, since  $(e_n(x_1), \dots, e_n(x_m), B_n)'$  converges in distribution to  $(e(x_1), \dots, e(x_m), B)'$  for any  $x_1, \dots, x_m > 0$  and  $m \in \mathbb{N}$  by Lemma B.1, we can apply the extended continuous mapping theorem (Theorem 18.11 in [33]) to obtain  $(W_n(x_1), \dots, W_n(x_m), B_n)'$   $\rightarrow$   $(e(x_1)/x_1, \dots, e(x_m)/x_m, B)'$  in distribution. This is the fidi-convergence needed to prove Lemma B.2.

Asymptotic tightness of the tail empirical process  $e_n$  follows from Theorem 4.1 in [27]. Asymptotic tightness of  $B_n$  follows from its weak convergence. This implies asymptotic tightness of the vector  $(e_n, B_n)$ , for instance by a simple adaptation of Lemma 1.4.3 in [34].  $\square$

*Proof of Lemma B.3.* Let  $H(x) = 1 - e^{-\theta x}$  be the cdf of the  $\text{Exp}(\theta)$ -distribution. From the proof of Lemma 9.2 in [3], we have, for any  $m \in \mathbb{N}$ ,

$$\sup_{x \in [1/m, m]} |\hat{H}_{k_n}(x) - H(x)| = o_{\mathbb{P}}(1), \quad n \rightarrow \infty.$$

Since

$$E_{n,m} - E'_{n,m} = \int_{1/m}^m W_n(x) \, d(\hat{H}_{k_n} - H)(x),$$

the assertion follows from Lemma B.2, Lemma C.8 in [2] and the continuous mapping theorem.  $\square$

*Proof of Lemma B.4.* As a consequence of Lemma B.3, Lemma B.2 and the continuous mapping theorem, we have

$$\begin{aligned} E_{n,m} + B_n &= \int_{1/m}^m W_n(x) \, \theta e^{-\theta x} \, dx + B_n + o_{\mathbb{P}}(1) \\ &\xrightarrow{d} \int_{1/m}^m \frac{e(x)}{x} \theta e^{-\theta x} \, dx + B = E'_m + B \sim \mathcal{N}(0, \tau_m^2), \end{aligned}$$

where the variance  $\tau_m^2$  is given by

$$\begin{aligned} \tau_m^2 &= \text{Var} \left\{ \int_{1/m}^m e(x) \frac{1}{x} \theta e^{-\theta x} dx \right\} + 2 \text{Cov} \left\{ \int_{1/m}^m e(x) \frac{1}{x} \theta e^{-\theta x} dx, B \right\} \\ &\quad + \text{Var}(B) \\ &= \theta^2 \int_{1/m}^m \int_{1/m}^m r(x, y) \frac{1}{xy} e^{-\theta(x+y)} dx dy + 2\theta \int_{1/m}^m f(x) \frac{1}{x} e^{-\theta x} dx + \frac{\pi^2}{6} \end{aligned}$$

as asserted.  $\square$

*Proof of Lemma B.5.* By the definition of  $\tau_m^2$  in Lemma B.4

$$\lim_{m \rightarrow \infty} \tau_m^2 = \theta^2 \int_0^\infty \int_0^\infty r(x, y) \frac{1}{xy} e^{-\theta(x+y)} dx dy + 2\theta \int_0^\infty f(x) \frac{1}{x} e^{-\theta x} dx + \frac{\pi^2}{6}. \quad (\text{B.6})$$

For  $x > y$ , we have  $r(x, y) = \theta x \mathbb{E} [\xi_1^{(y/x)} \xi_2^{(y/x)}]$  with  $(\xi_1^{(y/x)}, \xi_2^{(y/x)}) \sim \pi_2^{(y/x)}$ . Hence, applying the transformation  $z = y/x$ , the first summand on the right-hand side of (B.6) can be written as

$$\begin{aligned} \theta^2 \int_0^\infty \int_0^\infty \frac{r(x, y)}{xy} e^{-\theta(x+y)} dx dy &= 2\theta^3 \int_0^\infty \int_0^x \frac{\mathbb{E} [\xi_1^{(y/x)} \xi_2^{(y/x)}]}{y} e^{-\theta(x+y)} dy dx \\ &= 2\theta^3 \int_0^\infty \int_0^1 \frac{\mathbb{E} [\xi_1^{(z)} \xi_2^{(z)}]}{z} e^{-\theta x(1+z)} dz dx \\ &= 2\theta^2 \int_0^1 \frac{\mathbb{E} [\xi_1^{(z)} \xi_2^{(z)}]}{z(z+1)} dz. \end{aligned} \quad (\text{B.7})$$

For the second summand on the right-hand side of (B.6), note that

$$\sum_{i=1}^\infty ip_2^{(x, e^y)}(i, 0) = \mathbb{E} [\xi_1^{(e^y/x)} \mathbf{1}(\xi_2^{(e^y/x)} = 0)] \theta x e^{-\theta e^y}, \quad (\text{B.8})$$

see Formula (A.7) in the proof of Lemma 9.6 in [3] and  $\sum_{i=1}^\infty ip^{(x)}(i) = \mathbb{E}[N_E^{(x)}] = x$ , see [27]. Therefore, we can rewrite  $h$  from Lemma B.1 as follows

$$\begin{aligned} h(x) &= \int_0^\infty \mathbf{1}(e^y \leq x) \mathbb{E} [\xi_1^{(e^y/x)} \mathbf{1}(\xi_2^{(e^y/x)} = 0)] \theta x e^{-\theta e^y} dy \\ &\quad - \int_{-\infty}^0 x - \mathbf{1}(e^y \leq x) \mathbb{E} [\xi_1^{(e^y/x)} \mathbf{1}(\xi_2^{(e^y/x)} = 0)] \theta x e^{-\theta e^y} dy. \\ &= x \int_{1/x}^\infty \mathbf{1}(z \leq 1) \mathbb{E} [\xi_1^{(z)} \mathbf{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta zx}}{z} dz \\ &\quad - x \int_0^{1/x} \frac{1}{z} - \mathbf{1}(z \leq 1) \mathbb{E} [\xi_1^{(z)} \mathbf{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta zx}}{z} dz, \end{aligned}$$

where we have used the transformation  $z = e^y/x$ . For  $0 < x \leq 1$ , the first integral is zero and we obtain

$$\begin{aligned} h(x) &= -x \int_0^1 \frac{1}{z} - \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta z x}}{z} dz - x \int_1^{1/x} \frac{1}{z} dz \\ &= -x \int_0^1 \frac{1}{z} - \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta z x}}{z} dz + x \log(x), \end{aligned}$$

while for  $x > 1$ ,

$$\begin{aligned} h(x) &= x \int_{1/x}^1 \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta z x}}{z} dz \\ &\quad - x \int_0^{1/x} \frac{1}{z} - \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} = 0)] \theta \frac{e^{-\theta z x}}{z} dz. \end{aligned}$$

As a consequence, writing  $g(z) = \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} = 0)]$ , we obtain

$$\begin{aligned} \int_0^\infty h(x) \frac{1}{x} e^{-\theta x} dx &= \int_0^1 \log(x) e^{-\theta x} dx - \int_0^1 e^{-\theta x} \int_0^1 \frac{1}{z} - g(z) \theta \frac{e^{-\theta z x}}{z} dz dx \\ &\quad + \int_1^\infty e^{-\theta x} \int_{1/x}^1 g(z) \theta \frac{e^{-\theta z x}}{z} dz dx \\ &\quad - \int_1^\infty e^{-\theta x} \int_0^{1/x} \frac{1}{z} - g(z) \theta \frac{e^{-\theta z x}}{z} dz dx. \end{aligned}$$

Next, some tedious calculations based on Fubini's theorem allow to rewrite the sum of the last three double integrals as

$$s = \int_0^1 \frac{e^{-\theta/z} - 1}{\theta z} + \frac{g(z)}{z(1+z)} dz.$$

Using the fact that  $g(z) = \frac{1}{\theta} - \mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} > 0)]$ , we thus obtain

$$\begin{aligned} &\int_0^\infty h(x) \frac{1}{x} e^{-\theta x} dx \\ &= \int_0^1 \log(z) e^{-\theta z} + \frac{e^{-\theta/z} - 1}{\theta z} + \frac{1}{\theta z(1+z)} - \frac{\mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} > 0)]}{z(1+z)} dz \\ &= \int_0^1 \log(z) e^{-\theta z} + \frac{e^{-\theta/z}}{\theta z} - \frac{1}{\theta(1+z)} - \frac{\mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} > 0)]}{z(1+z)} dz. \end{aligned}$$

Finally, one can show

$$\int_0^1 \log(z) e^{-\theta z} + \frac{e^{-\theta/z}}{\theta z} dz = -(\log \theta + \gamma)/\theta = \varphi_{(C)}(\theta)/\theta,$$

such that, assembling terms and recalling  $f(x) = h(x) - x\varphi_{(C)}(\theta)$ ,

$$\begin{aligned} \int_0^1 f(x) \frac{1}{x} e^{-\theta x} dx &= \int_0^1 h(x) \frac{1}{x} e^{-\theta x} dx - \varphi_{(C)}(\theta) \int_0^\infty e^{-\theta x} dx \\ &= -\log(2)/\theta - \int_0^1 \frac{\mathbb{E} [\xi_1^{(z)} \mathbb{1}(\xi_2^{(z)} > 0)]}{z(1+z)} dz. \end{aligned} \tag{B.9}$$

The lemma is now an immediate consequence of (B.6), (B.7) and (B.9).  $\square$

*Proof of Lemma B.6.* By Lemma B.3, it suffices to show the assertion with  $E_{n,m}$  replaced by  $E'_{n,m}$ . Define  $\tilde{e}_n(x) := e_n(x) + k_n^{-1/2}$ , such that, by Condition 3.4(iii), we have

$$\max_{Z_{ni} \geq c} \left| \frac{\tilde{e}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right| = o_{\mathbb{P}}(1)$$

for any constant  $c > 0$ . Fix  $m \in \mathbb{N}$ . By the previous display, for any  $\varepsilon > 0$ , the event

$$B_n = B_n(m, \varepsilon) = \left\{ \max_{Z_{ni} \geq m} \left| \frac{\tilde{e}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right| \leq \varepsilon \right\}$$

satisfies  $\mathbb{P}(B_n) \rightarrow 1$ . Next,

$$\begin{aligned} |E_{n,m} - E_n| &\leq \left| \int_0^\infty \log \left( 1 + \frac{\tilde{e}_n(x)}{x \sqrt{k_n}} \right) \sqrt{k_n} \mathbb{1}_{(0,1/m]}(x) d\hat{H}_{k_n}(x) \right| \\ &\quad + \left| \int_0^\infty \log \left( 1 + \frac{\tilde{e}_n(x)}{x \sqrt{k_n}} \right) \sqrt{k_n} \mathbb{1}_{[m,\infty)}(x) d\hat{H}_{k_n}(x) \right| \\ &=: |V_{n1}| + |V_{n2}|, \end{aligned}$$

such that

$$|E_{n,m} - E_n| = |E_{n,m} - E_n| \mathbb{1}_{B_n} + o_{\mathbb{P}}(1) \leq |V_{n1}| + |V_{n2}| \mathbb{1}_{B_n} + o_{\mathbb{P}}(1). \tag{B.10}$$

We begin by treating the term  $|V_{n2}| \mathbb{1}_{B_n}$ . Since  $\log(1+x) = \int_0^1 x/(1+sx) ds$  for any  $x > -1$ , we have

$$\begin{aligned} V_{n2} \mathbb{1}_{B_n} &= \int_0^\infty \frac{\tilde{e}_n(x)}{x} \int_0^1 \frac{1}{1 + s \frac{\tilde{e}_n(x)}{x \sqrt{k_n}}} ds \mathbb{1}(x \geq m) d\hat{H}_{k_n}(x) \mathbb{1}_{B_n} \\ &= \frac{1}{k_n} \sum_{i=1}^{k_n} \frac{\tilde{e}_n(Z_{ni})}{Z_{ni}} \mathbb{1}(Z_{ni} \geq m) \int_0^1 \frac{1}{1 + s \frac{\tilde{e}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}}} ds \mathbb{1}_{B_n} \\ &= k_n^{-3/2} \sum_{i=1}^{k_n} \frac{\mathbb{1}(Z_{ni} \geq m)}{Z_{ni}} \int_0^1 \frac{1}{1 + s \frac{\tilde{e}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}}} ds \left\{ \sum_{t=1}^n f(U_t, Z_{ni}) + 1 \right\} \mathbb{1}_{B_n}, \end{aligned}$$

where

$$f(U_t, Z_{ni}) = \mathbb{1}(U_t > 1 - Z_{ni}/b_n) - Z_{ni}/b_n. \tag{B.11}$$

For given  $\varepsilon \in (0, c_1 \wedge c_2)$  with  $c_j$  as in Condition 3.1, let  $C_n = C_n(\varepsilon)$  denote the event  $\{\min_{i=1, \dots, k_n} N_{ni} > 1 - \varepsilon/2\} = \{\max_{i=1, \dots, k_n} Z_{ni} < \varepsilon b_n/2\}$ , which satisfies  $\mathbb{P}(C_n) \rightarrow 1$  by Condition 3.1(v). Hence, we can write  $V_{n2} \mathbb{1}_{B_n} = \bar{V}_{n2} \mathbb{1}_{C_n} + o_{\mathbb{P}}(1)$ , where

$$\begin{aligned} \bar{V}_{n2} &= k_n^{-3/2} \sum_{i=1}^{k_n} \frac{1}{Z_{ni}} \mathbb{1}(\varepsilon b_n/2 > Z_{ni} \geq m) \int_0^1 \frac{1}{1 + s \frac{\tilde{\varepsilon}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}}} ds \\ &\quad \times \left\{ \sum_{t=1}^n f(U_t, Z_{ni}) + 1 \right\} \mathbb{1}_{B_n}. \end{aligned}$$

We obtain

$$\begin{aligned} |\bar{V}_{n2}| &\leq \frac{1}{m} k_n^{-3/2} \sum_{i=1}^{k_n} \mathbb{1}(\varepsilon b_n/2 > Z_{ni} \geq m) \int_0^1 \frac{1}{\left| 1 + s \frac{\tilde{\varepsilon}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right|} ds \\ &\quad \times \left\{ \left| \sum_{t=1}^n f(U_t, Z_{ni}) \right| + 1 \right\} \mathbb{1}_{B_n}. \end{aligned}$$

On the event  $B_n$  the integral over  $s$  can be bounded as follows

$$\int_0^1 \frac{1}{\left| 1 + s \frac{\tilde{\varepsilon}_n(Z_{ni})}{Z_{ni} \sqrt{k_n}} \right|} ds \mathbb{1}_{B_n} \leq \int_0^1 \frac{1}{1 - s\varepsilon} ds \mathbb{1}_{B_n} \leq \frac{1}{1 - \varepsilon}.$$

The previous two displays imply that  $|\bar{V}_{n2}|$  is bounded by

$$\begin{aligned} &\frac{1}{m} \frac{1}{1 - \varepsilon} k_n^{-3/2} \sum_{i=1}^{k_n} \mathbb{1}(\varepsilon b_n/2 > Z_{ni} \geq m) \left\{ \left| \sum_{t=1}^n f(U_t, Z_{ni}) \right| + 1 \right\} \\ &= \frac{1}{m} \frac{1}{1 - \varepsilon} k_n^{-3/2} \sum_{i=1}^{k_n} \mathbb{1}(\varepsilon b_n/2 > Z_{ni} \geq m) \left| \sum_{t=1}^n f(U_t, Z_{ni}) \right| + O_{\mathbb{P}}(k_n^{-1/2}). \end{aligned}$$

The upper bound can now be treated exactly as in the proof of Lemma 9.1 in [3], finally yielding

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|V_{n2} \mathbb{1}_{B_n}| > \delta) = 0. \tag{B.12}$$

It remains to treat  $|V_{n1}|$ . Write

$$\begin{aligned} |V_{n1}| &\leq T_n(0, dk_n^{-1}) + T_n(dk_n^{-1}, dk_n^{-\mu}) + T_n(dk_n^{-\mu}, 1/m) \\ &=: T_{n1} + T_{n2} + T_{n3}, \end{aligned} \tag{B.13}$$

where, for some constant  $d > 0$  and  $\mu = \mu_d$  determined below,

$$T_n(a, b) = \sqrt{k_n} \int_0^\infty \mathbb{1}(x \in (a, b]) \left| \log \left( 1 + \frac{\tilde{\varepsilon}_n(x)}{x \sqrt{k_n}} \right) \right| d\hat{H}_{k_n}(x).$$

We start by covering the term  $T_{n1} = T_n(0, dk_n^{-1})$  and determining the constants  $d$  and  $\mu$ . Note that for the event  $J_n = \{\min_{i=1, \dots, k_n} Z_{ni} > dk_n^{-1}\}$  one has

$$\begin{aligned} \mathbb{P}(J_n) &= \mathbb{P}\left(k_n \min_{i=1, \dots, k_n} Z_{ni} > d\right) = \mathbb{P}\left(n \left(1 - \max_{i=1, \dots, k_n} N_{ni}\right) > d\right) \\ &= \mathbb{P}(Z_{1:n} > d) \rightarrow e^{-d\theta}. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{P}(T_{n1} > \delta) &= \mathbb{P}(T_{n1} \mathbb{1}_{J_n} + T_{n1} \mathbb{1}_{J_n^c} > \delta) \\ &\leq \mathbb{P}(T_{n1} \mathbb{1}_{J_n} > \delta/2) + \mathbb{P}(T_{n1} \mathbb{1}_{J_n^c} > \delta/2) \\ &\leq \mathbb{P}(J_n^c) \rightarrow 1 - \exp(-d\theta). \end{aligned}$$

Hence, for any given  $\varepsilon > 0$  we can choose  $d = d(\varepsilon) < -\log(1 - \varepsilon)/\theta$ , such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(T_{n1} > \delta) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(J_n^c) = 1 - \exp(-d\theta) < \varepsilon. \quad (\text{B.14})$$

Now, choose  $\mu = \mu_d \in (1/2, 1/\{2(1 - \tau)\})$  from Condition 3.4(iv), where  $\tau \in (0, 1/2)$  is from Condition 3.4(ii). Next, consider  $T_{n3} = T_n(dk_n^{-\mu}, 1/m)$  and note that, for  $x \in (dk_n^{-\mu}, 1/m]$ , we have

$$\left| \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right| = \left| \frac{\tilde{e}_n(x)}{x^\tau} \right| \frac{1}{x^{1-\tau}\sqrt{k_n}} \leq \frac{1}{d^{1-\tau}} \left| \frac{\tilde{e}_n(x)}{x^\tau} \right| k_n^{\mu(1-\tau)-1/2} = o_{\mathbb{P}}(1)$$

uniformly in  $x$ , by Condition 3.4(ii). As a consequence, the event

$$D_n = \left\{ \left| \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right| \leq \frac{1}{2} \right\}$$

satisfies  $\mathbb{1}_{D_n^c} = o_{\mathbb{P}}(1)$ , whence, recalling that  $x/(1+x) \leq \log(1+x) \leq x$  for any  $x > -1$ , we have

$$\begin{aligned} T_{n3} &= \sqrt{k_n} \int_{(dk_n^{-\mu}, 1/m]} \left| \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \right| \mathbb{1}_{D_n} d\hat{H}_{k_n}(x) + o_{\mathbb{P}}(1) \\ &\leq \int_{(dk_n^{-\mu}, 1/m]} \max \left\{ \left| \frac{\tilde{e}_n(x)}{x} \right|, \left| \frac{\tilde{e}_n(x)}{x} \right| \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right)^{-1} \right\} \mathbb{1}_{D_n} d\hat{H}_{k_n}(x) + o_{\mathbb{P}}(1) \\ &\leq 2 \int_{(dk_n^{-\mu}, 1/m]} \left| \frac{\tilde{e}_n(x)}{x^\tau} \right| \frac{1}{x^{1-\tau}} \mathbb{1}_{D_n} d\hat{H}_{k_n}(x) + o_{\mathbb{P}}(1). \end{aligned}$$

By Lemma B.15, Condition 3.4(ii) and the continuous mapping theorem, the last expression converges weakly to

$$T_3(m) = 2 \int_0^{1/m} \left| \frac{e(x)}{x^\tau} \right| \frac{1}{x^{1-\tau}} dH(x).$$

As a consequence,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(T_{n3} \geq \delta) \leq \lim_{m \rightarrow \infty} \mathbb{P}(T_3(m) > \delta) = 0. \quad (\text{B.15})$$

Finally, regarding  $T_{n2}$ , note that, for  $x \in (dk_n^{-1}, dk_n^{-\mu})$ ,

$$\begin{aligned} \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} &\leq \frac{1}{x} \left( \frac{1}{k_n} + \frac{1}{k_n} \sum_{i=1}^n \mathbb{1}(U_i > 1 - x/b_n) \right) \leq \frac{n+1}{d}, \\ \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} &\geq \frac{1}{x} \left( \frac{1}{k_n} - \frac{1}{k_n} \sum_{i=1}^n x/b_n \right) \geq \frac{1}{dk_n^{1-\mu}} - 1, \end{aligned}$$

which implies

$$\begin{aligned} &\left| \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \right| \\ &= \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \mathbb{1} \left( \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} > 0 \right) - \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \mathbb{1} \left( \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} < 0 \right) \\ &\leq \log \left( (n+1)d^{-1} + 1 \right) + \log \left( dk_n^{1-\mu} \right) \\ &\lesssim \log(n). \end{aligned}$$

As a consequence, the term  $T_{n2} = T_n(dk_n^{-1}, dk_n^{-\mu})$  can be bounded as follows

$$T_{n2} \lesssim \log(n) \sqrt{k_n} \int_{(dk_n^{-1}, dk_n^{-\mu}]} d\hat{H}_{k_n}(x) = \frac{\log(n)}{\sqrt{k_n}} \sum_{i=1}^{k_n} \mathbb{1}(Z_{ni} \in (dk_n^{-1}, dk_n^{-\mu})).$$

Hence, by Condition 3.4(iv),

$$\begin{aligned} E[T_{n2}] &\lesssim \log(n) \sqrt{k_n} \mathbb{P}(Z_{n1} < dk_n^{-\mu}) \\ &= \log(n) \sqrt{k_n} \{1 - \exp(-\theta dk_n^{-\mu})\} + o(1) \\ &= \theta d \log(n) k_n^{1/2-\mu} \{1 + o(1)\} + o(1) \\ &= O(\log(k_n) k_n^{1/2-\mu}) = o(1), \end{aligned} \tag{B.16}$$

where the last line follows from  $\log n = \log k_n + \log b_n \lesssim (1+q) \log k_n$  by Condition 3.4(i).

The assertion follows from (B.10), combined with (B.12), (B.13), (B.14), (B.15) and (B.16).  $\square$

### B.2. Auxiliary lemmas – sliding blocks

Throughout this section, we assume that Condition 3.1, 3.2(i) and 3.3(i) are met.

**Lemma B.7.** *For any  $x_1, \dots, x_m \in [0, \infty)$  and  $m \in \mathbb{N}$ , we have*

$$(e_n(x_1), \dots, e_n(x_m), B_n^{\text{sb}})' \xrightarrow{d} (e(x_1), \dots, e(x_m), B^{\text{sb}})',$$

where  $(e(x_1), \dots, e(x_m), B^{\text{sb}})' \sim \mathcal{N}_{m+1}(0, \Sigma_{m+1}^{\text{sb}})$  with

$$\Sigma_{m+1}^{\text{sb}} = \begin{pmatrix} r(x_1, x_1) & \dots & r(x_1, x_m) & f(x_1) \\ \vdots & \ddots & \vdots & \vdots \\ r(x_m, x_1) & \dots & r(x_m, x_m) & f(x_m) \\ f(x_1) & \dots & f(x_m) & 8 \log(2) - 4 \end{pmatrix}.$$

Here, the functions  $r$  and  $f$  are defined as in Lemma B.1.

**Lemma B.8.** For any  $m \in \mathbb{N}$ , we have

$$\{(W_n(x), B_n^{\text{sb}})'\}_{x \in [1/m, m]} \xrightarrow{d} \left\{ \left( \frac{e(x)}{x}, B^{\text{sb}} \right)' \right\}_{x \in [1/m, m]} \quad \text{in } D([1/m, m]) \times \mathbb{R},$$

where  $(e, B^{\text{sb}})'$  is a centered Gaussian process with continuous sample paths and with covariance functional as specified in Lemma B.7.

**Lemma B.9.** For any  $m \in \mathbb{N}$ , we have

$$E_{n,m}^{\text{sb}} = E'_{n,m} + o_{\mathbb{P}}(1) \quad \text{as } n \rightarrow \infty,$$

where  $E'_{n,m} = \int_{1/m}^m W_n(x) \theta e^{-\theta x} dx$  is as in Lemma B.3.

**Lemma B.10.** For any  $m \in \mathbb{N}$ , we have

$$E_{n,m}^{\text{sb}} + B_n^{\text{sb}} \xrightarrow{d} E'_m + B^{\text{sb}} \sim \mathcal{N}(0, \tau_{\text{sb},m}^2) \quad \text{as } n \rightarrow \infty,$$

where, with  $r$  and  $f$  defined as in Lemma B.1,

$$\begin{aligned} \tau_{\text{sb},m}^2 = \theta^2 \int_{1/m}^m \int_{1/m}^m r(x, y) \frac{1}{xy} e^{-\theta(x+y)} dx dy \\ + 2\theta \int_{1/m}^m f(x) \frac{1}{x} e^{-\theta x} dx + 8 \log(2) - 4. \end{aligned}$$

**Lemma B.11.** As  $m \rightarrow \infty$ ,  $\tau_{\text{sb},m}^2 \rightarrow \sigma_{\text{sb},(\text{C})}^2 / \theta^2$ , where  $\sigma_{\text{sb},(\text{C})}^2$  is specified in Theorem 4.1.

**Lemma B.12.** If in addition, Condition 3.4 holds, then, for all  $\delta > 0$ ,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P(|E_{n,m}^{\text{sb}} - E_n^{\text{sb}}| > \delta) = 0.$$

*Proof of Lemma B.7.* As in the proof of Lemma B.1 we only show joint weak convergence of  $(e_n(x), B_n^{\text{sb}})$  for some fixed  $x > 0$ ; the general case can be shown analogously. For given  $\varepsilon \in (0, c_1 \wedge c_2)$  let  $A'_n = \{\min_{t=1, \dots, n-b_n+1} N_{nt}^{\text{sb}} > 1 - \varepsilon\}$ , such that  $\mathbb{P}(A'_n) \rightarrow 1$  by Condition 3.1(v). By the Cramér-Wold device, it suffices to prove weak convergence of



$$\begin{aligned} \lambda_1 e_n(x) + \lambda_2 B_n^{\text{sb}} &= \sum_{j=1}^{k_n-1} \sum_{s \in I_j} \left[ \frac{\lambda_1}{\sqrt{k_n}} \left\{ \mathbb{1} \left( U_s > 1 - \frac{x}{b_n} \right) - \frac{x}{b_n} \right\} \right. \\ &\quad \left. + \frac{\lambda_2 \sqrt{k_n}}{n - b_n + 1} \left\{ \log(Z_{ns}^{\text{sb}}) - E[\log(Z_{ns}^{\text{sb}})] \right\} \right] + o_{\mathbb{P}}(1), \end{aligned}$$

for some arbitrary  $\lambda_1, \lambda_2 \in \mathbb{R}$ , where the negligible term stems from omitting a negligible number of summands.

We are going to apply a big block-small block argument, based on a suitable ‘blocking of blocks’ to take care of the serial dependence introduced through the use of sliding blocks. For that purpose, let  $k_n^* < k_n$  be an integer sequence with  $k_n^* \rightarrow \infty$  and  $k_n^* = o(k_n^{\delta/(2(1+\delta))})$ , where  $\delta$  is from Condition 3.1(ii). For  $q_n^* = \lfloor k_n / (k_n^* + 2) \rfloor$  and  $j = 1, \dots, q_n^*$ , define

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

Thus we have  $q_n^*$  big blocks  $J_j^+$  of size  $k_n^* b_n$ , which are separated by a small block  $J_j^-$  of size  $2b_n$ , just as in the construction in the proof of Lemma 10.3 in [3]. Consequently, we have  $\lambda_1 e_n(x) + \lambda_2 B_n^{\text{sb}} = L_n^+ + L_n^- + o_{\mathbb{P}}(1)$ , where

$$L_n^\pm = \frac{1}{\sqrt{q_n^*}} \sum_{j=1}^{q_n^*} W_{nj}^\pm$$

with

$$\begin{aligned} W_{nj}^\pm &= \sqrt{\frac{q_n^*}{k_n}} \sum_{s \in J_j^\pm} \lambda_1 \left\{ \mathbb{1} \left( U_s > 1 - \frac{x}{b_n} \right) - \frac{x}{b_n} \right\} \\ &\quad + \frac{\lambda_2 n}{n - b_n + 1} \frac{1}{b_n} \left\{ \log(Z_{ns}^{\text{sb}}) - E[\log(Z_{ns}^{\text{sb}})] \right\} \end{aligned}$$

for  $j = 1, \dots, q_n^*$ . In the following, we show that, on the one hand,  $L_n^- \mathbb{1}_{A'_n} = o_{\mathbb{P}}(1)$  and that, on the other hand,  $L_n^+ \mathbb{1}_{A'_n}$  converges to the claimed normal distribution. First, we cover  $L_n^- \mathbb{1}_{A'_n}$ . As in the proof of Lemma B.1, we have

$$Z_{ns}^{\text{sb}} = b_n \left( 1 - \max_{t=s, \dots, s+b_n-1} U_t \right) = b_n \left( 1 - \max_{t=s, \dots, s+b_n-1} U_t^\varepsilon \right) =: Z_{ns}^{\varepsilon, \text{sb}}$$

on the event  $A'_n$ , where  $U_t^\varepsilon = U_t \mathbb{1}(U_t > 1 - \varepsilon)$ . Hence, we can write  $L_n^- \mathbb{1}_{A'_n} = \tilde{L}_n^- \mathbb{1}_{A'_n} + o_{\mathbb{P}}(1) = \tilde{L}_n^- + o_{\mathbb{P}}(1)$  with

$$\tilde{L}_n^- = \frac{1}{\sqrt{q_n^*}} \sum_{j=1}^{q_n^*} W_{nj}^{\varepsilon-},$$

where

$$W_{nj}^{\varepsilon-} = \sqrt{\frac{q_n^*}{k_n}} \sum_{s \in J_j^-} \lambda_1 \left\{ \mathbb{1}\left(U_s^\varepsilon > 1 - \frac{x}{b_n}\right) - \frac{x}{b_n} \right\} + \frac{\lambda_2 n}{n - b_n + 1} \frac{1}{b_n} \left\{ \log(Z_{ns}^{\varepsilon, \text{sb}}) - E[\log(Z_{ns}^{\varepsilon, \text{sb}})] \right\}.$$

We proceed by showing that  $\text{Var}[\tilde{L}_n^-] = o(1)$ . By stationarity, one has

$$\text{Var}[\tilde{L}_n^-] = \text{Var}[W_{n1}^{\varepsilon-}] + \frac{2}{q_n^*} \sum_{j=1}^{q_n^*} (q_n^* - j) \text{Cov}(W_{n1}^{\varepsilon-}, W_{n,j+1}^{\varepsilon-}),$$

which is bounded by  $3 \text{Var}[W_{n1}^{\varepsilon-}] + 2 \sum_{j=2}^{q_n^*} |\text{Cov}(W_{n1}^{\varepsilon-}, W_{n,j+1}^{\varepsilon-})|$  in absolute value. First, we show  $\text{Var}[W_{n1}^{\varepsilon-}] = o(1)$ , for which it suffices to show that  $\|W_{n1}^{\varepsilon-}\|_p = o(1)$  for some  $p \in (2, 2 + \delta)$ . By Minkowski's inequality, one has

$$\begin{aligned} \|W_{n1}^{\varepsilon-}\|_p &\leq 2 \sqrt{\frac{q_n^*}{k_n}} \left[ |\lambda_1| \|N_{b_n}^{(x)}(E)\|_p + |\lambda_2| \|\log(Z_{n1}^{\varepsilon, \text{sb}}) - E[\log(Z_{n1}^{\varepsilon, \text{sb}})]\|_p \right] \\ &= O(\sqrt{q_n^*/k_n}) = o(1) \end{aligned} \tag{B.17}$$

by Condition 3.1(ii) and 3.2(i). It remains to treat the sum over the covariances. Since  $W_{nj}^{\varepsilon-}$  is  $\mathcal{B}_{\{(j(k_n^*+2)-2)b_n+1\}:\{(j(k_n^*+2)b_n\}}^\varepsilon$ -measurable, we may apply Lemma 3.11 in [9] to obtain

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

By Condition 3.1(iii), the sum  $\sum_{j=2}^{q_n^*} \alpha_{c_2}(jk_n^*b_n)^{1-2/p}$  converges to zero, hence  $\|W_{n1}^{\varepsilon-}\|_p = o(1)$  as asserted.

Let us now treat the term  $L_n^+ \mathbb{1}_{A_n'}$  and show weak convergence to the asserted normal distribution. One can write

$$L_n^+ \mathbb{1}_{A_n'} = \frac{1}{\sqrt{q_n^*}} \sum_{j=1}^{q_n^*} \tilde{W}_{nj}^+ + o_{\mathbb{P}}(1), \quad \tilde{W}_{nj}^+ = W_{nj}^+ \mathbb{1}\left(\max_{t \in J_j^+} Z_{nt}^{\text{sb}} < \varepsilon b_n\right).$$

A standard argument based on characteristic functions shows that the weak limit of  $q_n^{*-1/2} \sum_{j=1}^{q_n^*} \tilde{W}_{nj}^+$  is the same as if the summands were independent. By arguments as before, we may also pass back to an independent sample  $W_{nj}^+$ ,  $j = 1, \dots, q_n^*$ . The assertion then follows from Ljapunov's central limit theorem, once we have shown the Ljapunov condition.

For that purpose, note that  $\|W_{nj}^+\|_{2+\delta} = O(\sqrt{q_n^*k_n}) = O(\sqrt{k_n^*})$  by similar arguments as in (B.17) such that  $E[|W_{nj}^+|^{2+\delta}] = O(k_n^{*(2+\delta)/2})$ . As a consequence,

$$\frac{\sum_{j=1}^{q_n^*} E[|W_{nj}^+|^{2+\delta}]}{\left[\sum_{j=1}^{q_n^*} E[|W_{nj}^+|^2]\right]^{\frac{2+\delta}{2}}} = q_n^{*-\frac{\delta}{2}} \frac{E[|W_{n1}^+|^{2+\delta}]}{E[|W_{n1}^+|^2]^{\frac{2+\delta}{2}}} = O(k_n^{-\delta/2} k_n^{*1+\delta}) = o(1),$$

since  $k_n^* = o(k_n^{\delta/(2(1+\delta))})$  by construction and provided that the limit of  $E[|W_{n1}^+|^2]$  exists. If it does, we can conclude that  $L_n^+ \xrightarrow{d} \mathcal{N}(0, \lim_{n \rightarrow \infty} E[|W_{n1}^+|^2])$ . and it suffices to show that

$$\lim_{n \rightarrow \infty} E[|W_{n1}^+|^2] = \lambda_1^2 r(x, x) + 2\lambda_1 \lambda_2 f(x) + \lambda_2^2 \{8 \log(2) - 4\}.$$

To this, note that  $W_{n1}^+ = \lambda_1 e_{n^*}(x) + \lambda_2 B_{n^*}^{\text{sb}} + o_{\mathbb{P}}(1)$ , where  $e_{n^*}$  and  $B_{n^*}^{\text{sb}}$  are defined as  $e_n$  and  $B_n^{\text{sb}}$  with  $n$  replaced by  $n^* = k_n^* b_n$  and  $k_n$  by  $k_n^*$ ; and our general conditions still hold with this replacement. The result follows from Lemma B.13 and Lemma B.14 and the proof of Theorem 4.1 in [27].  $\square$

*Proof of Lemma B.8.* Up to notation, the proof is exactly the same as the one of Lemma B.2 in the disjoint blocks case.  $\square$

*Proof of Lemma B.9.* The result follows immediately from the argument in the proof of Lemma B.3 and the proof of Lemma 10.2 in [3].  $\square$

*Proof of Lemma B.10.* Up to notation, the proof is exactly the same as the one of Lemma B.4 in the disjoint blocks case.  $\square$

*Proof of Lemma B.11.* By the definition of  $\tau_m^2$  and  $\tau_{\text{sb},m}^2$  in Lemma B.4 and B.10, we have

$$\tau_{\text{sb},m}^2 = \tau_m^2 - \pi^2/6 + 8 \log(2) - 4.$$

Hence, by the proof of Lemma B.5 and the definition of  $\sigma_{\text{sb},C}^2$  in Theorem 4.1,

$$\lim_{m \rightarrow \infty} \tau_{\text{sb},m}^2 = \sigma_{\text{db},C}^2/\theta^2 - \pi^2/6 + 8 \log 2 - 4 = \sigma_{\text{sb},C}^2/\theta^2. \quad \square$$

*Proof of Lemma B.12.* The proof is similar to the one of Lemma B.6, which is why we keep it short. Write  $|E_{n,m}^{\text{sb}} - E_n^{\text{sb}}| \leq |V_{n1}| + |V_{n2}|$  with

$$\begin{aligned} V_{n1} &= \int_0^\infty \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \sqrt{k_n} \mathbb{1}_{(0,1/m]}(x) \, d\hat{H}_n^{\text{sb}}(x), \\ V_{n2} &= \int_0^\infty \log \left( 1 + \frac{\tilde{e}_n(x)}{x\sqrt{k_n}} \right) \sqrt{k_n} \mathbb{1}_{[m,\infty)}(x) \, d\hat{H}_n^{\text{sb}}(x), \end{aligned}$$

where  $\tilde{e}_n(x) = e_n(x) + k_n^{-1/2}$ . For some  $\varepsilon > 0$  define the event

$$B_n = \left\{ \max_{Z_{ni}^{\text{sb}} \geq m} \left| \frac{\tilde{e}_n(Z_{ni}^{\text{sb}})}{Z_{ni}^{\text{sb}} \sqrt{k_n}} \right| \leq \varepsilon \right\},$$

such that  $\mathbb{P}(B_n) \rightarrow 1$  by Condition 3.4(iii). As in the proof of Lemma B.6, with  $f$  defined in (B.11), we can write

$$V_{n2} \mathbb{1}_{B_n} = k_n^{-3/2} \sum_{i=1}^{k_n-1} \sum_{w \in I_i} \frac{1}{Z_{nw}^{\text{sb}}} \mathbb{1}(Z_{nw}^{\text{sb}} \geq m) \int_0^1 \frac{1}{1 + s \frac{\tilde{e}_n(Z_{nw}^{\text{sb}})}{Z_{nw}^{\text{sb}} \sqrt{k_n}}} \, ds$$

$$\times b_n^{-1} \left\{ \sum_{j=1}^{k_n} \sum_{t \in I_j} f(U_t, Z_{nw}^{\text{sb}}) + 1 \right\} \mathbb{1}_{B_n} + o_{\mathbb{P}}(1).$$

By Condition 3.1(v),  $\mathbb{P}(C_n) \rightarrow 1$  where  $C_n = \{ \min_{i=1, \dots, n-b_n+1} N_{ni}^{\text{sb}} > 1 - \varepsilon/2 \}$ . Hence,  $V_{n2} \mathbb{1}_{B_n} = \bar{V}_{n2} \mathbb{1}_{B_n} \mathbb{1}_{C_n} + o_{\mathbb{P}}(1)$ , where

$$\begin{aligned} \bar{V}_{n2} &= k_n^{-3/2} \sum_{i=1}^{k_n-1} \sum_{w \in I_i} \frac{1}{Z_{nw}^{\text{sb}}} \mathbb{1}(\varepsilon b_n/2 > Z_{nw}^{\text{sb}} \geq m) \int_0^1 \frac{1}{1+s} \frac{1}{\frac{\bar{\varepsilon}_n(Z_{nw}^{\text{sb}})}{Z_{nw}^{\text{sb}} \sqrt{k_n}}} ds \\ &\quad \times b_n^{-1} \left\{ \sum_{j=1}^{k_n} \sum_{t \in I_j} f(U_t, Z_{nw}^{\text{sb}}) + 1 \right\}, \end{aligned}$$

such that  $\bar{V}_{n2}$  can be bounded as in the proof of Lemma B.6 as follows

$$\begin{aligned} |\bar{V}_{n2} \mathbb{1}_{B_n}| &\leq \frac{1}{m} \frac{1}{1-\varepsilon} k_n^{-3/2} \sum_{i=1}^{k_n-1} \sum_{w \in I_i} \mathbb{1}(\varepsilon b_n/2 > Z_{nw}^{\text{sb}} \geq m) \\ &\quad \times b_n^{-1} \left| \sum_{j=1}^{k_n} \sum_{t \in I_j} f(U_t, Z_{nw}^{\text{sb}}) \right| + o_{\mathbb{P}}(1). \end{aligned}$$

This expression can be handled as in the proof of Lemma 10.1 in [3], such that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(|\bar{V}_{n2} \mathbb{1}_{B_n} \mathbb{1}_{C_n}| > \delta) = 0.$$

The remaining term  $|V_{n1}|$  can be treated analogously to the eponymous term in the proof of Lemma B.6.  $\square$

**Lemma B.13.** (a) For  $x \geq 0$ , as  $n \rightarrow \infty$ ,

$$\text{Cov}(e_n(x), B_n^{\text{sb}}) \rightarrow 2 \int_0^1 h_{\text{sb},x}(\xi) d\xi - 2x\varphi_{(C)}(\theta),$$

where

$$\begin{aligned} h_{\text{sb},x}(\xi) &= \sum_{i=1}^{\infty} i \int_0^{\infty} \mathbb{1}(y \leq \log(x)) \sum_{l=0}^i p^{(\xi x)}(l) p_2^{((1-\xi)x, (1-\xi)e^y)}(i-l, 0) e^{-\theta \xi e^y} \\ &\quad + \mathbb{1}(y > \log(x)) p^{(\xi x)}(i) e^{-\theta e^y} dy \\ &\quad - \sum_{i=1}^{\infty} i \int_{-\infty}^0 p^{(x)}(i) - \mathbb{1}(y \leq \log(x)) \sum_{l=0}^i p^{(\xi x)}(l) p_2^{((1-\xi)x, (1-\xi)e^y)}(i-l, 0) \\ &\quad \quad \quad \times e^{-\theta \xi e^y} \\ &\quad - \mathbb{1}(y > \log(x)) p^{(\xi x)}(i) e^{-\theta e^y} dy. \end{aligned}$$

(b) We have

$$2 \int_0^1 h_{\text{sb},x}(\xi) \, d\xi = h(x) + x\varphi_{(C)}(\theta),$$

where  $h$  is defined in Lemma B.1.

*Proof.* (a) We assume that both  $U_s$  and  $Z_{nt}^{\text{sb}}$  are measurable with respect to the appropriate  $\mathcal{B}_{\cdot}^{\varepsilon}$  sigma-algebra; the general case can be treated by multiplying with suitable indicator functions as in the proof of Lemma B.7. Let  $A_j = \sum_{s \in I_j} \mathbb{1}(U_s > 1 - \frac{x}{b_n})$  and  $D_j = \sum_{s \in I_j} \log(Z_{ns}^{\text{sb}})$ . Then

$$\begin{aligned} \text{Cov}(e_n(x), B_n^{\text{sb}}) &= \frac{1}{n - b_n + 1} \sum_{i=1}^{k_n} \sum_{j=1}^{k_n-1} \text{Cov}(A_i, D_j) \\ &\quad + \frac{1}{n - b_n + 1} \sum_{i=1}^{k_n} \text{Cov}(A_i, \log(Z_{n,n-b_n+1}^{\text{sb}})). \end{aligned}$$

The second sum is asymptotically negligible, since  $\|A_j\|_2 = \|N_{b_n}^{(x)}(E)\|_2 = O(1)$  and  $\|\log(Z_{n,n-b_n+1}^{\text{sb}})\|_2 = O(1)$  by Condition 3.1(ii) and 3.2(i). Next, following the argument in the proof of Lemma B.1 in [3], we may write

$$\begin{aligned} \text{Cov}(e_n(x), B_n^{\text{sb}}) &= \frac{1}{b_n} \text{Cov}(A_2, D_1 + D_2) + o(1) \\ &= \frac{1}{b_n} \sum_{t=1}^{2b_n} \text{Cov} \left\{ \sum_{s \in I_2} \mathbb{1}(U_s > 1 - \frac{x}{b_n}), \log(Z_{nt}^{\text{sb}}) \right\} + o(1) \\ &= \int_0^1 f_n(\xi) + g_n(\xi) \, d\xi - 2x \, \text{E} [\log(Z_{n1}^{\text{sb}})] + o(1), \end{aligned}$$

where

$$\begin{aligned} f_n(\xi) &= \sum_{t=1}^{b_n} \text{E} \left[ \sum_{s \in I_2} \mathbb{1}(U_s > 1 - \frac{x}{b_n}) \log(Z_{nt}^{\text{sb}}) \right] \mathbb{1} \left\{ \xi \in [\frac{t-1}{b_n}, \frac{t}{b_n}] \right\}, \\ g_n(\xi) &= \sum_{t=b_n+1}^{2b_n} \text{E} \left[ \sum_{s \in I_2} \mathbb{1}(U_s > 1 - \frac{x}{b_n}) \log(Z_{nt}^{\text{sb}}) \right] \mathbb{1} \left\{ \xi \in [\frac{t-b_n-1}{b_n}, \frac{t-b_n}{b_n}] \right\}. \end{aligned}$$

Note that  $\lim_{n \rightarrow \infty} \text{E}[\log(Z_{n1}^{\text{sb}})] = \varphi_{(C)}(\theta)$  by uniform integrability of  $\log(Z_{1:n})$ , and that  $\sup_{n \in \mathbb{N}} \|f_n\|_{\infty} + \|g_n\|_{\infty} < \infty$  as a consequence of  $\|\sum_{s \in I_1} \mathbb{1}(U_s > 1 - \frac{x}{b_n})\|_2 \times \|\log(Z_{n1}^{\text{sb}})\|_2 < \infty$  by Condition 3.1(ii) and 3.2(i). Hence, the lemma is proven if we show that, for any  $\xi \in (0, 1)$ ,

$$\lim_{n \rightarrow \infty} f_n(1 - \xi) = \lim_{n \rightarrow \infty} g_n(\xi) = h_{\text{sb},x}(\xi).$$

Since the proof for  $f_n(1 - \xi)$  is similar, we only treat  $g_n(\xi)$ , which can be written as

$$g_n(\xi) = \text{E} \left[ \sum_{s \in I_2} \mathbb{1}(U_s > 1 - \frac{x}{b_n}) \log(Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}}) \right].$$

Let us proceed by showing joint weak convergence of  $\sum_{s \in I_2} \mathbf{1}(U_s > 1 - \frac{x}{b_n})$  and  $\log(Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}})$ . For that purpose, note that

$$\begin{aligned} G_n(i, y) &:= \mathbb{P}\left(\sum_{s \in I_2} \mathbf{1}(U_s > 1 - \frac{x}{b_n}) = i, \log(Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}}) \geq y\right) \\ &= \mathbb{P}\left(\sum_{s \in I_2} \mathbf{1}(U_s > 1 - \frac{x}{b_n}) = i, Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}} \geq e^y\right), \end{aligned}$$

coincides with  $F_n(i, e^y)$  in the proof of Lemma B.1 in [3]. Hence, by that proof, we have

$$\lim_{n \rightarrow \infty} G_n(i, y) = \sum_{l=0}^i p^{(\xi x)}(l) p_2^{((1-\xi)x, (1-\xi)e^y)}(i-l, 0) e^{-\theta \xi e^y}$$

for  $y \leq \log x$  and

$$\lim_{n \rightarrow \infty} G_n(i, y) = p^{(\xi x)}(i) e^{-\theta e^y}$$

for  $y > \log x$ . Further, note that

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_{b_n}^{(x)}(E) = i) = p^{(x)}(i).$$

As a consequence of the previous three displays, and since weak convergence and uniform integrability imply convergence of moments, we have

$$\begin{aligned} g_n(\xi) &= \sum_{i=1}^{\infty} i \int_0^{\infty} \mathbb{P}\left(\sum_{s=b_n+1}^{2b_n} \mathbf{1}(U_s > 1 - \frac{x}{b_n}) = i, \log(Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}}) \geq y\right) dy \\ &\quad - i \int_{-\infty}^0 \mathbb{P}\left(\sum_{s=b_n+1}^{2b_n} \mathbf{1}(U_s > 1 - \frac{x}{b_n}) = i, \log(Z_{n, \lfloor (1+\xi)b_n \rfloor + 1}^{\text{sb}}) \leq y\right) dy \\ &= \sum_{i=1}^{\infty} i \int_0^{\infty} G_n(i, y) dy - i \int_{-\infty}^0 \mathbb{P}(N_{b_n}^{(x)}(E) = i) - G_n(i, y) dy \\ &\rightarrow h_{\text{sb}, x}(\xi) \end{aligned}$$

as asserted, which implies part (a) of the lemma.

(b) In the proof of Lemma B.3 in [3] it is shown that, for  $y \leq \log(x)$ ,

$$\begin{aligned} S(x, y, \xi) &= e^{-\theta \xi e^y} \sum_{i=1}^{\infty} i \sum_{l=0}^i p^{(\xi x)}(l) p_2^{((1-\xi)x, (1-\xi)e^y)}(i-l, 0) \\ &= \xi x e^{-\theta e^y} + \mathbb{E} \left[ \xi_{11}^{(e^y/x)} \mathbf{1}(\xi_{12}^{(e^y/x)} = 0) \right] \theta (1-\xi) x e^{-\theta e^y}, \end{aligned}$$

where  $(\xi_{11}^{(y/x)}, \xi_{12}^{(y/x)}) \sim \pi_2^{(y/x)}$ . Equation (B.8) then allows to rewrite

$$S(x, y, \xi) = \xi x e^{-\theta e^y} + (1-\xi) \sum_{i=1}^{\infty} i p_2^{(x, e^y)}(i, 0) \equiv \xi x e^{-\theta e^y} + (1-\xi) T(x, y).$$

As a consequence, further noting that  $\sum_{i=1}^{\infty} i p^{(\xi x)}(i) = \xi x$ , we obtain

$$h_{\text{sb},x}(\xi) = \int_0^\infty \xi x e^{-\theta e^y} + \mathbf{1}(y \leq \log(x))(1 - \xi)T(x, y) \, dy - \int_{-\infty}^0 x - \xi x e^{-\theta e^y} - \mathbf{1}(y \leq \log(x))(1 - \xi)T(x, y) \, dy.$$

Then, by Fubini's theorem,

$$2 \int_0^1 h_{\text{sb},x}(\xi) \, d\xi = \int_0^\infty x e^{-\theta e^y} + \mathbf{1}(y \leq \log x)T(x, y) \, dy - \int_{-\infty}^0 x(1 - e^{-\theta e^y}) + x - \mathbf{1}(y \leq \log(x))T(x, y) \, dy.$$

The assertion now follows from the fact that

$$\int_0^\infty e^{-\theta e^y} \, dy = \int_\theta^\infty \frac{e^{-z}}{z} \, dz = -\text{Ei}(-\theta)$$

and

$$\begin{aligned} \int_{-\infty}^0 1 - e^{-\theta e^y} \, dy &= \int_0^\theta \frac{1 - e^{-z}}{z} \, dz = (1 - e^{-z}) \log(z) \Big|_0^\theta - \int_0^\theta e^{-z} \log(z) \, dz \\ &= \log(\theta) - e^{-\theta} \log(\theta) - \left\{ \gamma - \int_\theta^\infty e^{-z} \log(z) \, dz \right\} \\ &= \log(\theta) - e^{-\theta} \log(\theta) - \gamma + \left\{ -e^{-z} \log(z) \Big|_\theta^\infty + \int_\theta^\infty \frac{e^{-z}}{z} \, dz \right\} \\ &= \log(\theta) + \gamma - \text{Ei}(-\theta) = -\varphi_{(C)}(\theta) - \text{Ei}(-\theta) \end{aligned}$$

after assembling terms, where  $\text{Ei}(x) = -\int_{-x}^\infty e^{-t}/t \, dt$  for  $x > 0$  is the exponential integral. □

**Lemma B.14.** *One has*

$$\lim_{n \rightarrow \infty} \text{Var}(B_n^{\text{sb}}) = 8 \log(2) - 4 \approx 1.545.$$

*Proof.* As in the proof of Lemma B.13, we assume that the  $Z_{nt}^{\text{sb}}$  are measurable with respect to the appropriate  $\mathcal{B}_{\cdot}^{\varepsilon}$  sigma-algebra. We may then argue as in that proof to obtain

$$\begin{aligned} \text{Var}(B_n^{\text{sb}}) &= \frac{2}{b_n} \sum_{t=1}^{b_n} \text{E} [\log(Z_{n1}^{\text{sb}}) \log(Z_{n,1+t}^{\text{sb}})] - 2 \text{E}[\log(Z_{n1}^{\text{sb}})]^2 + o(1) \\ &= 2 \int_0^1 f_n(\xi) \, d\xi - 2 \text{E}[\log(Z_{n1}^{\text{sb}})]^2 + o(1), \end{aligned} \tag{B.18}$$

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

$$\begin{aligned} f_n(\xi) &= \sum_{t=1}^{b_n} \mathbb{E}[\log(Z_{n1}^{\text{sb}}) \log(Z_{n,1+t}^{\text{sb}})] \mathbf{1}(\xi \in [\frac{t-1}{b_n}, \frac{t}{b_n})) \\ &= \mathbb{E}[\log(Z_{n1}^{\text{sb}}) \log(Z_{n, \lfloor b_n \xi \rfloor + 1}^{\text{sb}})]. \end{aligned}$$

By Condition 3.2(i), we have  $\mathbb{E}[\log(Z_{n1}^{\text{sb}})] \rightarrow \varphi_{(C)}(\theta)$ . Further,

$$\sup_{n \in \mathbb{N}} \|f_n\|_{\infty} \leq \sup_{n \in \mathbb{N}} \mathbb{E}[\log(Z_{n1}^{\text{sb}})^2] < \infty,$$

whence convergence of the integral over  $f_n$  in (B.18) may be concluded from the dominated convergence theorem, once we have shown pointwise convergence of  $f_n$ . To this end we show that, for any fixed  $\xi \in (0, 1)$ ,

$$(\log(Z_{n1}^{\text{sb}}), \log(Z_{n, \lfloor b_n \xi \rfloor + 1}^{\text{sb}})) \xrightarrow{d} (X^{(\xi)}, Y^{(\xi)}) \quad (\text{B.19})$$

for some random vector  $(X^{(\xi)}, Y^{(\xi)})$ . This in turn will imply

$$\lim_{n \rightarrow \infty} f_n(\xi) = \lim_{n \rightarrow \infty} \mathbb{E}[\log(Z_{n1}^{\text{sb}}) \log(Z_{n, \lfloor b_n \xi \rfloor + 1}^{\text{sb}})] = \mathbb{E}[X^{(\xi)} Y^{(\xi)}]$$

by Condition 3.2(i) and therefore

$$\lim_{n \rightarrow \infty} \text{Var}(B_n^{\text{sb}}) = 2 \int_0^1 \mathbb{E}[X^{(\xi)} Y^{(\xi)}] d\xi - 2\varphi_{(C)}(\theta)^2 = 2 \int_0^1 \text{Cov}(X^{(\xi)}, Y^{(\xi)}) d\xi. \quad (\text{B.20})$$

For the proof of (B.19), define, for  $x, y \in \mathbb{R}$ ,

$$\begin{aligned} G_{n,\xi}(x, y) &= \mathbb{P}(\log(Z_{n1}^{\text{sb}}) > x, \log(Z_{n, \lfloor b_n \xi \rfloor + 1}^{\text{sb}}) > y) \\ &= \mathbb{P}(Z_{n1}^{\text{sb}} > e^x, Z_{n, \lfloor b_n \xi \rfloor + 1}^{\text{sb}} > e^y), \end{aligned}$$

which converges to

$$G_{\xi}(x, y) = \exp(-\theta[\xi(e^x \wedge e^y) + (e^x \vee e^y)])$$

by the proof of Lemma B.2 in [3]. Hence, (B.19), where the random vector  $(X^{(\xi)}, Y^{(\xi)})$  has joint c.d.f.

$$\begin{aligned} F_{\xi}(x, y) &= \mathbb{P}(X^{(\xi)} \leq x, Y^{(\xi)} \leq y) \\ &= 1 - \mathbb{P}(X^{(\xi)} > x) - \mathbb{P}(Y^{(\xi)} > y) + G_{\xi}(x, y), \\ &= 1 - \exp(-\theta e^x) - \exp(-\theta e^y) + G_{\xi}(x, y). \end{aligned}$$

We are left with calculating the right-hand side of (B.20). By Lemma B.16, we have



$$\begin{aligned}
V &\equiv \int_0^1 \text{Cov}(X^{(\xi)}, Y^{(\xi)}) \, d\xi \\
&= \int_0^1 \int_0^\infty \int_0^\infty G_\xi(x, y) - e^{-\theta e^x} e^{-\theta e^y} \, dx \, dy \, d\xi \\
&\quad + \int_0^1 \int_{-\infty}^0 \int_{-\infty}^0 F_\xi(x, y) - (1 - e^{-\theta e^x})(1 - e^{-\theta e^y}) \, dx \, dy \, d\xi \\
&\quad - 2 \int_0^1 \int_{-\infty}^0 \int_0^\infty \mathbb{P}(X^{(\xi)} > x, Y^{(\xi)} \leq y) - e^{-\theta e^x} (1 - e^{-\theta e^y}) \, dx \, dy \, d\xi, \\
&\equiv A + B - 2 \cdot C. \tag{B.21}
\end{aligned}$$

We start with the first summand  $A$ . Recall the exponential integral  $\text{Ei}(x) = -\int_{-x}^\infty e^{-t}/t \, dt$  for  $x > 0$ , and note that  $\int_y^\infty e^{-\theta e^x} \, dx = -\text{Ei}(-\theta e^y)$  for  $y \in \mathbb{R}$  and  $\int_0^1 e^{-a\xi} \, d\xi = (1 - e^{-a})/a$  for  $a > 0$ . Fubini's theorem allows to rewrite  $A$  as

$$\begin{aligned}
&\int_0^\infty \int_0^y e^{-\theta e^y} \left\{ \int_0^1 e^{-\theta \xi e^x} \, d\xi - e^{-\theta e^x} \right\} \, dx \\
&\quad + \int_y^\infty e^{-\theta e^x} \left\{ \int_0^1 e^{-\theta \xi e^y} \, d\xi - e^{-\theta e^y} \right\} \, dx \, dy \\
&= \int_0^\infty e^{-\theta e^y} \int_0^y \frac{1 - e^{-\theta e^x}}{\theta e^x} - e^{-\theta e^x} \, dx \\
&\quad + \int_y^\infty e^{-\theta e^x} \, dx \left\{ \frac{1 - e^{-\theta e^y}}{\theta e^y} - e^{-\theta e^y} \right\} \, dy \\
&= \int_0^\infty e^{-\theta e^y} \left\{ \frac{e^{-\theta e^y} - 1}{\theta e^y} - \frac{e^{-\theta} - 1}{\theta} \right\} + \{-\text{Ei}(-\theta e^y)\} \left\{ \frac{1 - e^{-\theta e^y}}{\theta e^y} - e^{-\theta e^y} \right\} \, dy.
\end{aligned}$$

Next, invoke the substitution  $z = \theta e^y$  to obtain that

$$A = \int_\theta^\infty \left\{ \frac{e^{-z}}{z} - \frac{1}{z} + \frac{1 - e^{-\theta}}{\theta} \right\} \frac{e^{-z}}{z} - \text{Ei}(-z) \left\{ \frac{1}{z} - \frac{e^{-z}}{z} - e^{-z} \right\} \frac{1}{z} \, dz. \tag{B.22}$$

A similar calculation allows to rewrite

$$\begin{aligned}
B &= \int_0^1 \int_{-\infty}^0 \int_{-\infty}^0 G_\xi(x, y) - e^{-\theta e^x} e^{-\theta e^y} \, dx \, dy \, d\xi \\
&= \int_{-\infty}^0 \int_{-\infty}^y e^{-\theta e^y} \left\{ \int_0^1 e^{-\theta \xi e^x} \, d\xi - e^{-\theta e^x} \right\} \, dx \\
&\quad + \int_y^0 e^{-\theta e^x} \left\{ \int_0^1 e^{-\theta \xi e^y} \, d\xi - e^{-\theta e^y} \right\} \, dx \, dy \\
&= \int_{-\infty}^0 e^{-\theta e^y} \int_{-\infty}^y \frac{1 - e^{-\theta e^x}}{\theta e^x} - e^{-\theta e^x} \, dx \\
&\quad + \int_y^0 e^{-\theta e^x} \, dx \left\{ \frac{1 - e^{-\theta e^y}}{\theta e^y} - e^{-\theta e^y} \right\} \, dy
\end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^0 e^{-\theta e^y} \left\{ \frac{e^{-\theta e^y} - 1}{\theta e^y} + 1 \right\} \\
&\quad + \left\{ \text{Ei}(-\theta) - \text{Ei}(-\theta e^y) \right\} \left\{ \frac{1 - e^{-\theta e^y}}{\theta e^y} - e^{-\theta e^y} \right\} dy,
\end{aligned}$$

and the substitution  $z = \theta e^y$  yields

$$B = \int_0^\theta \left\{ \frac{e^{-z}}{z} - \frac{1}{z} + 1 \right\} \frac{e^{-z}}{z} + \left\{ \text{Ei}(-\theta) - \text{Ei}(-z) \right\} \left\{ \frac{1}{z} - \frac{e^{-z}}{z} - e^{-z} \right\} \frac{1}{z} dz. \tag{B.23}$$

Finally, regarding the term  $C$ , we have

$$\begin{aligned}
C &= \int_0^1 \int_{-\infty}^0 \int_0^\infty e^{-\theta e^x} e^{-\theta e^y} - G_\xi(x, y) dx dy d\xi \\
&= \int_{-\infty}^0 \int_0^\infty e^{-\theta e^x} \left\{ e^{-\theta e^y} - \int_0^1 e^{-\theta \xi e^y} d\xi \right\} dx dy \\
&= \{-\text{Ei}(-\theta)\} \int_{-\infty}^0 e^{-\theta e^y} - \frac{1 - e^{-\theta e^y}}{\theta e^y} dy \\
&= \text{Ei}(-\theta) \int_0^\theta \left\{ \frac{1}{z} - \frac{e^{-z}}{z} - e^{-z} \right\} \frac{1}{z} dz. \tag{B.24}
\end{aligned}$$

Next, the expressions in (B.22), (B.23) and (B.24) may be plugged-into (B.21). Using the notations

$$g(z) = \left\{ \frac{1}{z} - \frac{e^{-z}}{z} - e^{-z} \right\} \frac{1}{z}, \quad h(z) = \left\{ \frac{e^{-z}}{z} - \frac{1}{z} + 1 \right\} \frac{e^{-z}}{z},$$

we obtain that

$$\begin{aligned}
V &= \int_\theta^\infty \left\{ \frac{1 - e^{-\theta}}{\theta} - 1 \right\} \frac{e^{-z}}{z} + h(z) + \{-\text{Ei}(-z)\} g(z) dz \\
&\quad + \int_0^\theta \left\{ \text{Ei}(-\theta) - \text{Ei}(-z) \right\} g(z) + h(z) - 2\text{Ei}(-\theta)g(z) dz \\
&= \int_0^\infty h(z) + \{-\text{Ei}(-z)\} g(z) dz + \frac{1 - e^{-\theta} - \theta}{\theta} \{-\text{Ei}(-\theta)\} \\
&\quad - \text{Ei}(-\theta) \int_0^\theta g(z) dz
\end{aligned}$$

The first integral is independent of  $\theta$ , and can be seen to be equal to  $4 \log 2 - 2$ . Further,  $\int_0^\theta g(z) dz = (e^{-\theta} - 1 + \theta)/\theta$ , whence the last two summands cancel out. This proves the lemma.  $\square$

### B.3. Further auxiliary lemmas

**Lemma B.15.** *Let  $A$  be a continuous function on  $[0, 1]$  with  $\lim_{x \rightarrow 0} A(x)/x^\eta = 0$  for some  $\eta \in (0, 1/2)$ . Further, let  $H_n$  and  $H$  be monotone and non-negative*

functions on  $[0, 1]$  with

$$\limsup_{n \rightarrow \infty} \int_{[0,1]} \frac{1}{x^{1-\eta}} dH_n(x) < \infty \quad \text{and} \quad \int_{[0,1]} \frac{1}{x^{1-\eta}} dH(x) < \infty.$$

If  $\lim_{n \rightarrow \infty} \sup_{x \in [0,1]} |B_n(x)| = 0$ , where  $B_n := H_n - H$ , and if there is a sequence of measurable functions  $A_n$  such that

$$\lim_{n \rightarrow \infty} \sup_{x \in [0,1]} \left| \frac{A_n(x) - A(x)}{x^\eta} \right| = 0,$$

then we have

$$\lim_{n \rightarrow \infty} \int_{[0,1]} \frac{A_n(x)}{x} dB_n(x) = 0.$$

*Proof.* For  $r \in \mathbb{N}$  define the piecewise constant function

$$\tilde{A}_r(x) := \sum_{k=1}^r \mathbf{1}_{(\frac{k-1}{r}, \frac{k}{r}]}(x) \frac{A(k/r)}{k/r}$$

as an approximation of  $A(x)/x$ . We write  $\int_{[0,1]} A_n(x)/x dB_n(x) = I_{n1} + I_{n2} + I_{n3}$ , where

$$\begin{aligned} I_{n1} &= \int_{[0,1]} \frac{A_n(x) - A(x)}{x} dB_n(x), & I_{n2} &= \int_{[0,1]} \frac{A(x)}{x} - \tilde{A}_r(x) dB_n(x), \\ I_{n3} &= \int_{[0,1]} \tilde{A}_r(x) dB_n(x). \end{aligned}$$

The first integral is bounded by

$$\begin{aligned} &\int_{[0,1]} \left| \frac{A_n(x) - A(x)}{x} \right| d(H_n + H)(x) \\ &\leq \sup_{x \in [0,1]} \left| \frac{A_n(x) - A(x)}{x^\eta} \right| \int_{[0,1]} \frac{1}{x^{1-\eta}} d(H_n + H)(x), \end{aligned}$$

which converges to zero by assumption. Regarding  $I_{n2}$ , we obtain

$$\begin{aligned} |I_{n2}| &= \left| \int_{[0,1]} \frac{A(x) - \tilde{A}_r(x)x}{x^\eta} \frac{1}{x^{1-\eta}} dB_n(x) \right| \\ &\leq \sup_{x \in [0,1]} \left| \frac{A(x) - \tilde{A}_r(x)x}{x^\eta} \right| \int_{[0,1]} \frac{1}{x^{1-\eta}} d(H_n + H)(x). \end{aligned} \tag{B.25}$$

By uniform continuity of  $x \mapsto A(x)/x^\eta$  on  $[0, 1]$ , we have

$$\sup_{x \in [0,1]} \left| \frac{A(x) - \tilde{A}_r(x)x}{x^\eta} \right| \rightarrow 0 \quad \text{for } r \rightarrow \infty.$$

Thus, the limes superior (for  $n \rightarrow \infty$ ) of the expression on the right-hand side of (B.25) can be made arbitrarily small by increasing  $r$ . Finally, we can bound  $|I_{n3}|$  as follows

$$\begin{aligned} |I_{n3}| &\leq \sum_{k=1}^r \frac{|A(k/r)|}{k/r} \left| \int_{[0,1]} \mathbb{1}_{\left(\frac{k-1}{r}, \frac{k}{r}\right]}(x) \, dB_n(x) \right| \\ &= \sum_{k=1}^r \frac{|A(k/r)|}{k/r} \left| B_n\left(\frac{k}{r}\right) - B_n\left(\frac{k-1}{r}\right) \right| \\ &\leq 2r^2 \sup_{x \in [0,1]} |A(x)| \sup_{x \in [0,1]} |B_n(x)|, \end{aligned}$$

which converges to zero by assumption.  $\square$

**Lemma B.16.** *Let  $X$  and  $Y$  be real-valued random variables such that  $XY$  is integrable. Then,*

$$\begin{aligned} \mathbb{E}[XY] &= \int_0^\infty \int_0^\infty \mathbb{P}(X > x, Y > y) \, dx \, dy + \int_{-\infty}^0 \int_{-\infty}^0 \mathbb{P}(X \leq x, Y \leq y) \, dx \, dy \\ &\quad - \int_{-\infty}^0 \int_0^\infty \mathbb{P}(X > x, Y \leq y) \, dx \, dy - \int_0^\infty \int_{-\infty}^0 \mathbb{P}(X \leq x, Y > y) \, dx \, dy. \end{aligned}$$

*Proof.* This is a standard calculation based on Fubini's theorem.  $\square$

## Acknowledgements

This work 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. The authors are grateful to two referees and an associate editor for their constructive comments on an earlier version of this article which lead to a substantial improvements. The authors are also grateful to Shi-Yuan Zhou for her support with the simulation study.

## References

- [1] 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. [MR2108013](#)
- [2] Berghaus, B. and A. Bücher (2017, 004). Goodness-of-fit tests for multivariate copula-based time series models. *Econometric Theory* 33(2), 292–330. [MR3600045](#)
- [3] Berghaus, B. and A. Bücher (2018, 10). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* 46(5), 2307–2335. [MR3845019](#)

- [4] Billingsley, P. (1979). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. New York-Chichester-Brisbane: John Wiley & Sons. [MR0534323](#)
- [5] Cai, J. J. (2019). A nonparametric estimator of the extremal index. *ArXiv e-prints*.
- [6] Capéraà, P., A.-L. Fougères, and C. Genest (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* *84*(3), 567–577. [MR1603985](#)
- [7] Darsow, W. F., B. Nguyen, and E. T. Olsen (1992). Copulas and Markov processes. *Illinois J. Math.* *36*(4), 600–642. [MR1215798](#)
- [8] 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. [MR1014450](#)
- [9] 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. [MR1958777](#)
- [10] Drees, H. (2000). Weighted approximations of tail processes for  $\beta$ -mixing random variables. *Ann. Appl. Probab.* *10*(4), 1274–1301. [MR1810875](#)
- [11] Ferreira, H. and M. Ferreira (2018). Estimating the extremal index through local dependence. *Annales de l'Institut Henri Poincaré, Probabilités et Statistique* *54*(2), 587–605. [MR3795060](#)
- [12] Ferreira, M. (2018). Analysis of estimation methods for the extremal index. *Electronic Journal of Applied Statistical Analysis* *11*(1), 296–306. [MR3807534](#)
- [13] 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. [MR1983763](#)
- [14] Genest, C. and J. Segers (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* *37*(5B), 2990–3022. [MR2541453](#)
- [15] 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. [MR2634031](#)
- [16] Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* *21*(4), 2043–2071. [MR1245780](#)
- [17] 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. [MR0940870](#)
- [18] Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* *65*(2), 291–306. [MR0722133](#)
- [19] Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. New York-Berlin: Springer-Verlag. [MR0691492](#)
- [20] Leadbetter, M. R. and H. Rootzén (1988). Extremal theory for stochastic processes. *Ann. Probab.* *16*(2), 431–478. [MR0929071](#)
- [21] Naveau, P., A. Guillo, D. Cooley, and J. Diebolt (2009). Modelling pairwise dependence of maxima in space. *Biometrika* *96*(1), 1–17. [MR2482131](#)
- [22] Northrop, P. J. (2015). An efficient semiparametric maxima estimator of

- the extremal index. *Extremes* 18(4), 585–603. [MR3418769](#)
- [23] O’Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.* 15(1), 281–291. [MR0877604](#)
- [24] Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab.* 4(2), 529–548. [MR1272738](#)
- [25] Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, Volume 49, pp. 859–878, 894–902. With a discussion. [MR0820979](#)
- [26] Rémillard, B., N. Papageorgiou, and F. Soustra (2012). Copula-based semi-parametric models for multivariate time series. *J. Multivariate Anal.* 110, 30–42. [MR2927508](#)
- [27] Robert, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.* 37(1), 271–310. [MR2488352](#)
- [28] 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. [MR2540849](#)
- [29] Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.* 119(2), 468–490. [MR2494000](#)
- [30] Shao, Q.-M. and H. Yu (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* 24(4), 2098–2127. [MR1415243](#)
- [31] Smith, R. L. and I. Weissman (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* 56(3), 515–528. [MR1278224](#)
- [32] Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes* 10(1-2), 41–55. [MR2407640](#)
- [33] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 3. Cambridge: Cambridge University Press. [MR1652247](#)
- [34] van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer. [MR1385671](#)