

CLUSTER BASED INFERENCE FOR EXTREMES OF TIME SERIES

BY HOLGER DREES¹, ANJA JANSSEN², SEBASTIAN NEBLUNG^{1,*}

¹UNIVERSITY OF HAMBURG, DEPARTMENT OF MATHEMATICS, SPST, BUNDESSTR. 55, 20146 HAMBURG, GERMANY, DREES@MATH.UNI-HAMBURG.DE; * SEBASTIAN.NEBLUNG@UNI-HAMBURG.DE

²OTTO-VON-GUERICKE UNIVERSITY OF MAGDEBURG, FACULTY OF MATHEMATICS, IMST, UNIVERSITÄTSPLATZ 2, 39106 MAGDEBURG, GERMANY, ANJA.JANSSEN@OVGU.DE

Abstract We introduce a new type of estimator for the spectral tail process of a regularly varying time series. The approach is based on a characterizing invariance property of the spectral tail process, which is incorporated into the new estimator via a projection technique. We show uniform asymptotic normality of this estimator, both in the case of known and of unknown index of regular variation. In a simulation study the new procedure shows a more stable performance than previously proposed estimators.

1. Introduction Statistical methods for assessing environmental hazards, technical failures or financial risks often call for estimating rare events and make use of extreme value theory, see, e.g., [Castillo \(2012\)](#), [Coles \(2001\)](#), [Embrechts et al. \(2013\)](#) or [Finkenstädt & Rootzén \(2003\)](#). In many of these applications, the temporal dependence structure of sequential observations matters: extremal events may happen in clusters and the length and shape of a cluster are crucial for its impact. For example, heavy rainfall over several days has a different effect on water levels of adjacent rivers than a short cloudburst. Likewise, consecutive large daily losses of a stock will have a more severe impact on an investor's portfolio than a single drop in its value.

In order to have a theoretical framework for such extremal dependence structures, we will assume that our quantities of interest can be described by an $\mathbb{R}^d$ -valued stationary regularly varying time series $(X_t)_{t \in \mathbb{Z}}$. The distribution of its *tail process* $(Y_t)_{t \in \mathbb{Z}}$ is defined by

$$(1.1) \quad \lim_{u \rightarrow \infty} P \left(\frac{X_{-m}}{u} \leq x_{-m}, \dots, \frac{X_n}{u} \leq x_n \mid \|X_0\| > u \right) = P \{Y_{-m} \leq x_{-m}, \dots, Y_n \leq x_n\}$$

for all $m, n \in \mathbb{N}_0$ and continuity points $(x_{-m}, \dots, x_n) \in (\mathbb{R}^d)^{n+m+1}$ of $P(Y_t)_{-m \leq t \leq n}$, where $\|\cdot\|$ denotes an arbitrary but fixed norm on $\mathbb{R}^d$ and all inequalities between vectors are meant componentwise. As shown in [Basrak & Segers \(2009\)](#), Theorems 2.1 and 3.1, $\|Y_0\|$ is Pareto distributed with $P\{\|Y_0\| > x\} = x^{-\alpha}, x \geq 1$, for some $\alpha > 0$ (the so-called index of regular variation), and it is independent of the *spectral tail process* $\Theta = (\Theta_t)_{t \in \mathbb{Z}} := (Y_t / \|Y_0\|)_{t \in \mathbb{Z}}$, which thus carries all information about the dependence structure of the tail process. It is our aim to provide estimators for quantities derived from Θ , exemplarily done here for $P\{\Theta_i \in A\}$ for suitable Borel sets A in $\mathbb{R}^d$ and $i \in \mathbb{Z}$. If we observe a finite stretch from a stationary regularly varying time series $(X_t)_{t \in \mathbb{Z}}$, Equation (1.1) together with the definition of Θ leads to a straightforward estimator

AMS 2000 subject classifications: Primary 62G32; secondary 62M10, 62G05, 60F17.

Keywords and phrases: cluster of extremes, empirical processes, extreme value analysis, projection estimator, spectral tail process, time series, uniform central limit theorems.

of $P\{\Theta_i \in A\}$ given by

$$(1.2) \quad \hat{p}_{n,A}^f := \frac{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \mathbb{1}_A(X_{t+i}/\|X_t\|)}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}},$$

where u_n denotes a suitably high threshold, depending on n , the number of consecutively observed vectors $(X_t, \dots, X_{t+i})$. This estimator is called *forward estimator* and asymptotic normality as $n \rightarrow \infty$ has been shown under suitable assumptions in the univariate case in [Davis et al. \(2018\)](#), Theorem 3.1.

Now, the assumption of stationarity of $(X_t)_{t \in \mathbb{Z}}$ does not only allow for the construction of consistent estimators, it also has an effect on the structure of the spectral tail process, which can be shown to satisfy the so-called *time change formula*, see [Basrak & Segers \(2009\)](#), Theorem 3.1. Namely, for all bounded and measurable $f : (\mathbb{R}^d)^{n+m+1} \rightarrow \mathbb{R}$, $n, m \in \mathbb{N}_0$ satisfying $f(x_{-m}, \dots, x_0, \dots, x_n) = 0$ whenever $x_0 = 0$, we have

$$E[f((\Theta_{t+i})_{-m \leq t \leq n})] = E\left[f\left(\frac{(\Theta_t)_{-m \leq t \leq n}}{\|\Theta_{-i}\|}\right) \|\Theta_{-i}\|^\alpha\right]$$

for all $i \in \mathbb{Z}$.

For the univariate case and sets of the form $A = (x, \infty)$ or $A = (-\infty, -x)$, $x \geq 0$, this identity with $f(x_0) = \mathbb{1}_A(x_0)$ was used in [Drees et al. \(2015\)](#) for $i = 1$ and in [Davis et al. \(2018\)](#) for general i to design an alternative to $\hat{p}_{n,A}^f$, based on the same number n of consecutively observed vectors, given by

$$(1.3) \quad \hat{p}_{n,A}^b := \frac{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \mathbb{1}_A(X_t/\|X_{t-i}\|) \|X_{t-i}/X_t\|^{\hat{\alpha}_n}}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}},$$

where $\hat{\alpha}_n$ is the Hill estimator of α , given by

$$(1.4) \quad \hat{\alpha}_n := \frac{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}}{\sum_{t=1}^n \log^+(\|X_t\|/u_n)},$$

with $\log^+ x = \log(x) \vee 0$. Simulation studies have shown that this so-called *backward estimator* can have a smaller root mean squared error than the forward estimator in particular for larger values of x , see [Drees et al. \(2015\)](#) and [Davis et al. \(2018\)](#).

In this article, we take a more comprehensive approach for using the stationarity of $(X_t)_{t \in \mathbb{Z}}$ when constructing an estimator for the distribution of Θ . It has been shown in [Janßen \(2019\)](#) that the class of all possible spectral tail processes of underlying stationary processes can be characterized with the help of the so-called RS-transform. For a stochastic processes $W = (W_t)_{t \in \mathbb{Z}}$ we write P^W for its distribution and set

$$\|W\|_\alpha := \left(\sum_{t \in \mathbb{Z}} \|W_t\|^\alpha \right)^{1/\alpha} \in [0, \infty]$$

for some $\alpha > 0$. If $0 < \|W\|_\alpha < \infty$ a.s. then the *RS-transform* of P^W is given by

$$(P^W)^{RS}(B) = E \left(\sum_{j \in \mathbb{Z}} \frac{\|W_j\|^\alpha}{\|W\|_\alpha^\alpha} \mathbb{1}_B \left(\frac{(W_{t+j})_{t \in \mathbb{Z}}}{\|W_j\|} \right) \right)$$

for all cylinder sets B in $(\mathbb{R}^d)^\mathbb{Z}$. Note that this RS-transform depends on α , which we suppress in the notation. Theorem 2.4 in [Janßen \(2019\)](#) states that processes $\Theta = (\Theta_t)_{t \in \mathbb{Z}}$ with $0 < \|\Theta\|_\alpha < \infty$ a.s. are spectral tail processes of some underlying stationary time series with index α if and only if their distribution is invariant under the RS-transformation, i.e. $(P^\Theta)^{RS} = P^\Theta$.

Direct calculations furthermore show that for any P^W with $0 < \|W\|_\alpha < \infty$ a.s. we have

$$((P^W)^{RS})^{RS} = (P^W)^{RS},$$

i.e. for such P^W , its RS-transform $(P^W)^{RS}$ satisfies the stated invariance property and is thereby a valid distribution of a spectral tail process of a stationary time series. The RS-transformation can thus be seen as a projection of an initial distribution into the class of admissible limit objects for our estimation problem. Note that common projection estimators would typically transform an initial estimator into its best approximation from the space of admissible quantities by minimizing a suitably chosen norm, see for example [Fils-Villetard et al. \(2008\)](#). In contrast, our approach transforms an initial estimator for Θ onto the space of admissible processes by making sure that this transformation induces only a random shift in time and scale.

A straightforward initial empirical version of the distribution of Θ , trimmed to a finite length $2s_n + 1, s_n \geq 0$, can be derived from n consecutive (and overlapping) stretches of $(X_t)_{t \in \mathbb{Z}}$ of this length by

$$(1.5) \quad \widehat{P^\Theta} = \frac{1}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}} \delta_{(\dots, 0, (\frac{X_{t+j}}{\|X_t\|})_{|j| \leq s_n}, 0, \dots)},$$

where $\delta_{(\dots, 0, (\frac{X_{t+j}}{\|X_t\|})_{|j| \leq s_n}, 0, \dots)}$ is the Dirac measure placing mass 1 in the double sided sequence $(z_j)_{j \in \mathbb{Z}}$ with $z_j = X_{t+j}/\|X_t\|, |j| \leq s_n$, and $z_j = 0, |j| > s_n$. We then apply the RS-transformation to the realization of $\widehat{P^\Theta}$ which yields

$$(\widehat{P^\Theta})^{RS} = \frac{1}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \frac{\mathbf{1}_{\{\|X_t\| > u_n\}}}{\sum_{h=-s_n}^{s_n} \|X_{t+h}\|^\alpha} \sum_{h=-s_n}^{s_n} \|X_{t+h}\|^\alpha \delta_{(\dots, 0, (\frac{X_{t+h+j}}{\|X_{t+h}\|})_{|h+j| \leq s_n}, 0, \dots)}$$

as a projection based estimator of P^Θ . Thus we ensure that our final estimator is derived from a quantity which has all essential structural properties of a spectral tail process. This gives for $P\{\Theta_i \in A\}$ the *projection based estimator*

$$\begin{aligned} \hat{p}_{n,A} &:= (\widehat{P^\Theta})^{RS} \{(x_t)_{t \in \mathbb{Z}} \in (\mathbb{R}^d)^\mathbb{Z} \mid x_i \in A\} = \frac{1}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \\ &\times \sum_{t=1}^n \frac{\mathbf{1}_{\{\|X_t\| > u_n\}}}{\sum_{h=-s_n}^{s_n} \|X_{t+h}\|^\alpha} \left(\sum_{h \in H_n} \|X_{t+h}\|^\alpha \mathbf{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \sum_{h \in H_n^c} \|X_{t+h}\|^\alpha \mathbf{1}_A(0) \right), \end{aligned}$$

where $H_n := \{(-s_n - i) \vee (-s_n), \dots, (s_n - i) \wedge s_n\}$ and $H_n^c := \{-s_n, \dots, s_n\} \setminus H_n$ which equals $\{s_n - i + 1, \dots, s_n\}$ for $i > 0$, $\{-s_n, \dots, -s_n - i - 1\}$ for $i < 0$, and it is the empty set for $i = 0$. We denote the version with estimated α according to (1.4) by

$$\begin{aligned} \hat{\hat{p}}_{n,A} &:= \frac{1}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \frac{\mathbf{1}_{\{\|X_t\| > u_n\}}}{\sum_{h=-s_n}^{s_n} \|X_{t+h}\|^{\hat{\alpha}_n}} \times \\ &\left(\sum_{h \in H_n} \|X_{t+h}\|^{\hat{\alpha}_n} \mathbf{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \sum_{h \in H_n^c} \|X_{t+h}\|^{\hat{\alpha}_n} \mathbf{1}_A(0) \right). \end{aligned}$$

In Section 2 we state the uniform asymptotic normality of these projection based estimators. In Subsection 2.1 the conditions needed in these results are verified for solutions to stochastic recurrence equations. The simulations in Section 3 show that the finite sample performance of $\hat{\hat{p}}_{n,A}$ is favorable compared to that of $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ in several settings. The proofs of the results in Section 2 are given in the Appendix, with some auxiliary lemmas and calculations and further simulations deferred to the Supplement.

Throughout, for vectors $x, y \in \mathbb{R}^q$, we interpret $x \leq y$ componentwise, i.e. as $x_j \leq y_j$ for all $1 \leq j \leq q$, while $x \preceq y$ means $x \leq y$ and $x \neq y$. Where needed, 0 is also interpreted as a vector with all coordinates equal to 0, $[x, y] := \times_{j=1}^q [x_j, y_j]$ and $[x, y] := \times_{j=1}^q [x_j, y_j]$. We denote the positive and negative part of the logarithm by $\log^+ x := \log x \vee 0$ and $\log^- x := (-\log x) \vee 0$, respectively. The limit from the left of a function f at x is denoted by $f(x-)$.

2. Asymptotics In this section we establish convergence of the suitably standardized estimators $\hat{p}_{n,A}$ and $\hat{\hat{p}}_{n,A}$ uniformly for all sets A in some family $\mathcal{A} \subset \mathbb{B}^d$. To this end, note that the estimator with known index of regular variation can be written in the form $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$ where

$$T_{n,A} := \sum_{t=1}^n g_A(W_{n,t}) \quad \text{with} \quad W_{n,t} := \left(\frac{X_{t+h}}{u_n} \right)_{-s_n \leq h \leq s_n},$$

$$g_A((w_h)_{-s \leq h \leq s}) := \frac{\mathbb{1}_{\{\|w_0\| > 1\}}}{\sum_{h=-s}^s \|w_h\|^\alpha} \sum_{h=-s}^s \|w_h\|^\alpha \left(\mathbb{1}_A \left(\frac{w_{h+i}}{\|w_h\|} \right) \mathbb{1}_{\{|h+i| \leq s\}} + \mathbb{1}_A(0) \mathbb{1}_{\{|h+i| > s\}} \right)$$

for all $s \in \mathbb{N}$. We will use the theory on the asymptotic behavior of quite general statistics based on sliding blocks of observations, developed by [Drees & Neblung \(2021\)](#), to derive the asymptotics of $(T_{n,A})_{A \in \mathcal{A}}$ and thus of our estimators.

Throughout, we fix the sequences $s_n \in \mathbb{N}$ and $u_n > 0$, $n \in \mathbb{N}$, and the family $\mathcal{A}$ of Borel sets in $\mathbb{R}^d$. To ensure consistency of the estimators, u_n must tend to ∞ sufficiently slowly such that the expected number of exceedances among the observations tends to ∞ , that is $nv_n \rightarrow \infty$ with

$$v_n := P\{\|X_1\| > u_n\}.$$

We assume that the observed time series is stationary, regularly varying and absolutely regular (β -mixing) with β -mixing coefficients

$$\beta_k := E \left[\sup_{B \in \mathcal{B}_{k+1}^\infty} |P(B|\mathcal{B}_{-\infty}^0) - P(B)| \right]$$

where $\mathcal{B}_{-\infty}^0$ and $\mathcal{B}_{k+1}^\infty$ denote the σ -fields generated by $(X_l)_{l \leq 0}$ and $(X_l)_{l \geq k+1}$, respectively. More precisely, we assume

- (RV) $(X_t)_{t \in \mathbb{Z}}$ is an $\mathbb{R}^d$ -valued stationary time series which is regularly varying with index $\alpha > 0$, spectral tail process $\Theta = (\Theta_t)_{t \in \mathbb{Z}}$ and tail process $Y = (Y_t)_{t \in \mathbb{Z}}$.
- (S) The sequences $s_n \in \mathbb{N}$ and nv_n tend to ∞ as $n \rightarrow \infty$. There exist sequences $l_n, r_n \in \mathbb{N}$, $n \in \mathbb{N}$, such that $s_n \leq l_n = o(r_n)$, $r_n v_n \rightarrow 0$, $r_n = o(\sqrt{nv_n})$ and $\beta_{l_n} = o(r_n/n)$.

Note that if (S) holds for some sequence $(l_n)_{n \in \mathbb{N}}$, then it is fulfilled for any $\tilde{l}_n \geq l_n$ such that $\tilde{l}_n = o(r_n)$. Hence, w.l.o.g. we may assume that, for any fixed $k \in \mathbb{N}$, $\beta_{l_n - ks_n} = o(r_n/n)$ holds. While (S) ensures that in the definition of $\hat{p}_{n,A}$ summands whose indices t differ by at least l_n are almost independent, it does not suffice to control the size of clusters of extremes. To this end, we use the following two conditions:

- (BC) For each $c \in (0, 1]$, $n \in \mathbb{N}$ and $1 \leq k \leq r_n$ there exists

$$e_{n,c}(k) \geq P(\|X_k\| > cu_n \mid \|X_0\| > cu_n)$$

such that $e_{\infty,c}(k) = \lim_{n \rightarrow \infty} e_{n,c}(k)$ exists for all $k \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} e_{n,c}(k) = \sum_{k=1}^{\infty} e_{\infty,c}(k) < \infty$ holds.

$$(\mathbf{TC}) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} E \left(\frac{\sum_{m < |h| \leq s_n} \|X_{h+j}\|^\alpha \mathbf{1}_{\{\|X_{h+j}\| \leq cu_n\}}}{\sum_{|h| \leq s_n} \|X_{h+j}\|^\alpha} \mid \|X_0\| > u_n \right) = 0$$

for some $c \in (0, 1)$ and all $j \in \mathbb{Z}$.

By stationarity, Condition (BC) implies the analogous inequality for $k < 0$ with $e_{n,c}(k) := e_{n,c}(|k|)$, and it also implies a bound on the second moment of the cluster size. (TC) enables us to truncate clusters to finite lengths in our asymptotic analysis. Moreover, the following condition guarantees that $p_A := P\{\Theta_i \in A\}$ can be calculated as a limit using the definition of the spectral tail process:

$$(\mathbf{C}\Theta) \quad P\{\Theta_i \in \partial A\} = 0 \text{ for all } A \in \mathcal{A}.$$

Our first result states the joint asymptotic normality of finitely many estimators $\hat{p}_{n,A}$. Define the process of standardized estimators

$$Z_n^{(p)}(A) := \sqrt{nv_n}(\hat{p}_{n,A} - E(g_A(W_{n,0}) \mid \|X_0\| > u_n)), \quad A \in \mathcal{A}.$$

Since the functions g_A are bounded by 1, the weak convergence (1.1) implies

$$(2.1) \quad E(g_A(W_{n,0}) \mid \|X_0\| > u_n) \rightarrow E \left(\|\Theta\|_\alpha^{-\alpha} \sum_{h \in \mathbb{Z}} \|\Theta_h\|^\alpha \mathbf{1}_A(\Theta_{h+i}/\|\Theta_h\|) \right) = P\{\Theta_i \in A\},$$

where the last step holds by the invariance of the spectral tail process under the RS-transformation; see Supplement for details.

THEOREM 2.1. If the conditions (RV), (S), (BC), (TC) and (C Θ) are met, then the finite dimensional distributions (fidis) of $Z_n^{(p)}$ weakly converge to the fidis of $Z^{(p)}(A) := Z(A) - P\{\Theta_i \in A\}Z(\mathbb{R}^d)$ with Z denoting a centered Gaussian process with covariance function

$$(2.2) \quad c(A, B) := Cov(Z(A), Z(B)) \\ = \sum_{j \in \mathbb{Z}} E \left[(\|\Theta_j\|^\alpha \wedge 1) \left(\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right) \left(\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_B \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right) \right].$$

Hence, if, in addition, the bias condition

$$(2.3) \quad E(g_A(W_{n,0}) \mid \|X_0\| > u_n) - P\{\Theta_i \in A\} = o((nv_n)^{-1/2})$$

holds for all $A \in \mathcal{A}$, then

$$\sqrt{nv_n}(\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \tilde{\mathcal{A}}} \rightarrow (Z^{(p)}(A))_{A \in \tilde{\mathcal{A}}}$$

weakly for all finite subsets $\tilde{\mathcal{A}}$ of $\mathcal{A}$.

Note that (2.3) only imposes a rate, while convergence is ensured by (2.1). Straight-forward calculations show that the covariance function of the limit process $Z^{(p)}$ can be represented as follows:

$$Cov(Z^{(p)}(A), Z^{(p)}(B)) = E[\chi(\xi(A) - p_A)(\xi(B) - p_B)]$$

with

$$\chi := \sum_{j \in \mathbb{Z}} (\|\Theta_j\|^\alpha \wedge 1), \quad \xi := \sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \delta_{\Theta_{h+i}/\|\Theta_h\|}.$$

Note that ξ is a random probability measure with expectation $E[\xi(A)] = (P^\Theta)^{RS}\{x \in (\mathbb{R}^d)^\mathbb{Z} \mid x_i \in A\} = p_A$, whose realizations only depend on the shape of the paths of the spectral tail process, in the sense that they are invariant under a shift and re-scaling of the spectral tail process. In particular, ξ is deterministic if the spectral tail process is the RS-transform of some deterministic process. In that case, the asymptotic variances are all equal to 0 although the spectral tail process may be non-trivial. The expected value of the random weight χ equals the expected number of values in the tail process with norm larger than 1. It reflects the fact that longer clusters of large observations contribute more summands to the estimator $\hat{p}_{n,A}$ than shorter clusters do.

The above convergence will only hold uniformly over all $A \in \mathcal{A}$ if the family of sets is not too complex. Such a process convergence in the space of bounded functions indexed by $\mathcal{A}$ (equipped with the supremum norm) is relatively easy to establish using Vapnik-Chervonenkis theory if the sets under consideration are linearly ordered w.r.t. inclusion. However, this assumption is too restrictive if multivariate data is observed, i.e. if $d > 1$. Here we assume that $\mathcal{A}$ may be indexed by a unit cube of arbitrary dimension in a suitable way:

- (A) For some $q \in \mathbb{N}$, there exists a map $[0, 1]^q \rightarrow \mathcal{A}, t \mapsto A_t$ such that
- (i) $\mathcal{A} = \{A_t \mid t \in [0, 1]^q\}$, $A_{(1, \dots, 1)} = \mathbb{R}^d$ and $A_{(t_1, \dots, t_q)} = \emptyset$ if $t_j = 0$ for some $1 \leq j \leq q$;
 - (ii) for all $1 \leq j, k \leq q$, and all $s_j, t_l \in [0, 1]$, $(l \in \{1, \dots, q\} \setminus \{k\})$ the mapping $t_k \mapsto A_{(t_1, \dots, t_q)} \setminus A_{(t_1, \dots, t_{j-1}, s_j, t_{j+1}, \dots, t_q)}$ is non-decreasing on $[0, 1]$;
 - (iii) the processes $(\sum_{j=1}^n g_{A_t}(W_{n,j}))_{t \in [0, 1]^q}$ are separable;
 - (iv) $P\{\Theta_i \in \partial A_t^-\} = 0$ for all $t \in [0, 1 + \iota]^q$ for some $\iota > 0$ where $A_t^- := \bigcup_{s \in [0, t)} A_s$ and $A_t := A_{t \wedge 1}$ for $t \notin [0, 1]^q$;
 - (v) $P\{\Theta_i \in \bigcap_{s \in (t, 1]} A_{s^{(k)}} \setminus A_{t^{(k)}}\} = 0$ for all $t \in [0, 1)$ and $1 \leq k \leq q$ where $t^{(k)} := (1, \dots, 1, t, 1, \dots, 1)$ with t in the k -th coordinate;
 - (vi) $P\{\|X_0\| > 0, X_i/\|X_0\| \in \bigcap_{s \in (t, 1]} A_{s^{(k)}} \setminus A_{t^{(k)}}\} = 0$ for all $t \in [0, 1)$ and $1 \leq k \leq q$;
 - (vii) there exists $w \in [0, 1]^q$ such that $0 \in A_w \setminus \bigcup_{s \prec w} A_s$.

The archetypical example is $\mathcal{A} = \{(-\infty, t] \mid t \in \mathbb{R}^d\} \cup \{\mathbb{R}^d, \emptyset\}$ for which condition (A) is satisfied, provided the marginal distributions of Θ_i and $X_i/\|X_0\|$ are continuous. To see this, let $q = d$ and $A_{(t_1, \dots, t_d)} := \times_{i=1}^d (-\infty, \varpi(t_i)] \cap \mathbb{R}^d$ for some continuous, increasing mapping ϖ from $[0, 1]$ onto $[-\infty, \infty]$ and observe $\bigcap_{s \in (t, 1]} A_{s^{(k)}} \setminus A_{t^{(k)}} = \emptyset$ for all $t \in [0, 1)$ and $1 \leq k \leq d$. Indeed, the above conditions cover almost all natural finite-dimensional families of sets if the distributions of Θ_i and $X_i/\|X_0\|$ are sufficiently smooth.

THEOREM 2.2. If the conditions (RV), (S), (BC), (TC), (C Θ) and (A) are fulfilled, then $Z_n^{(p)}$ weakly converge to the Gaussian process $Z^{(p)}$ defined in Theorem 2.1. If, in addition,

$$(2.4) \quad \sup_{A \in \mathcal{A}} |E(g_A(W_{n,0}) \mid \|X_0\| > u_n) - P\{\Theta_i \in A\}| = o((nv_n)^{-1/2}),$$

then

$$\sqrt{nv_n}(\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}} \rightarrow (Z^{(p)}(A))_{A \in \mathcal{A}}.$$

REMARK 2.1. In the case $q = 1$, assumptions (A) (vi) and (vii) are not needed.

Next we investigate the uniform asymptotic normality of our estimator $\hat{p}_{n,A}$ in the case when the index of regular variation is unknown. To analyze the joint asymptotic behavior of the denominator of the Hill type estimator $\hat{\alpha}_n$ and the statistics $T_{n,A}$, we amend the cluster size condition (BC). Since the logarithm function occurring in the definition of the Hill estimator is unbounded, further moment conditions are needed. In addition, we require the bias of the Hill type estimator to be negligible.

(BC') For each $1 \leq k \leq r_n$ there exists

$$e'_n(k) \geq E\left(\max\left(\log \frac{\|X_0\|}{u_n}, 1\right) \max\left(\log \frac{\|X_k\|}{u_n}, \mathbf{1}_{\{\|X_k\| > u_n\}}\right) \mid \|X_0\| > u_n\right)$$

such that $e'_\infty(k) = \lim_{n \rightarrow \infty} e'_n(k)$ exists for all $k \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} e'_n(k) = \sum_{k=1}^\infty e'_\infty(k) < \infty$ holds.

(M) There exists $\delta > 0$ such that the following moment bounds hold:

$$(i) \sum_{k=1}^{r_n} \left[E\left(\left(\log^+ \left(\frac{\|X_0\|}{u_n} \right) \log^+ \left(\frac{\|X_k\|}{u_n} \right) \right)^{1+\delta} \mid \|X_0\| > u_n \right) \right]^{1/(1+\delta)} = O(1),$$

$$(ii) \limsup_{m \rightarrow \infty} E\left(\frac{\sum_{|h| \leq m} |\log \|\Theta_h\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha} \right) < \infty,$$

$$(iii) \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} E\left(\frac{\sum_{m < |h| \leq s_n} \log^-(\|X_h\|/u_n) \|X_h\|^\alpha}{\sum_{|k| \leq s_n} \|X_k\|^\alpha} \mid \|X_0\| > u_n \right) = 0.$$

$$(H) E\left(\log \left(\frac{\|X_0\|}{u_n} \right) \mid \|X_0\| > u_n \right) - \frac{1}{\alpha} = o((nv_n)^{-1/2})$$

THEOREM 2.3. If the conditions (RV), (S), (BC), (TC), (CΘ), (A), (BC'), (M), (H) and (2.4) are satisfied and $\log^4 n = o(nv_n)$, then

$$\sqrt{nv_n}(\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}} \rightarrow (Z(A) - (P\{\Theta_i \in A\} - \alpha d_A)Z(\mathbb{R}^d) - \alpha^2 d_A Z_\phi)_{A \in \mathcal{A}}$$

weakly. Here

$$d_A := - \sum_{k \in \mathbb{Z}} E[\|\Theta\|_\alpha^{-\alpha} \|\Theta_k\|^\alpha \log \|\Theta_k\| \mathbf{1}_A(\Theta_i)] \in \mathbb{R}$$

and $((Z(A))_{A \in \mathcal{A}}, Z_\phi)$ is a centered Gaussian process whose covariance function is given by (2.2) and

$$\begin{aligned} Cov(Z(A), Z_\phi) &= \sum_{h \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} E\left[\frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{\Theta_{h+i}}{\|\Theta_h\|}\right) (\|\Theta_k\|^\alpha \wedge 1) (\alpha^{-1} + \log^+ \|\Theta_k\|) \right], \\ Var(Z_\phi) &= \alpha^{-1} \sum_{k \in \mathbb{Z}} E\left[(\|\Theta_k\|^\alpha \wedge 1) (2\alpha^{-1} + |\log \|\Theta_k\||) \right]. \end{aligned}$$

REMARK 2.2. (a) Condition (M) (ii) is equivalent to

$$(2.5) \quad \limsup_{m \rightarrow \infty} E\left(\frac{\sum_{|h| \leq m} |\log \|Y_h\||^{1+\delta} \|Y_h\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha} \right) < \infty.$$

(b) Condition (M) (ii) is implied by the following condition, which may sometimes be easier to check:

(M') There exists $\delta' > \delta$ such that the following moment bounds hold:

$$(i) E\left(\sup_{|h| \leq s_n} \left(\log^+ \left(\frac{\|X_h\|}{u_n} \right) \right)^{1+\delta'} \mid \|X_0\| > u_n \right) = O(1),$$

$$(ii) \sum_{|h| \leq s_n} E\left(\left(\log^+ \left(\frac{\|X_h\|}{u_n} \right) \right)^{1+\delta'} \left(\frac{\|X_h\|}{u_n} \right)^\alpha \mid \|X_0\| > u_n \right) = O(1).$$

Moreover, Condition (TC) is implied by (M') (ii); see Supplement for the proofs.

In general, neither $\hat{p}_{n,A}$ nor $\hat{p}_{n,A}^b$ nor $\hat{p}_{n,A}^f$ has uniformly smallest asymptotic variance, as is illustrated in the next example.

EXAMPLE 2.4. Define a process $W = (W_t)_{t \in \mathbb{Z}}$ by

$$P\{W_0 = a^{-1}, W_1 = -1, W_t = 0, \forall t \notin \{0, 1\}\} = p,$$

$$P\{W_0 = 1, W_1 = b^{-1}, W_t = 0, \forall t \notin \{0, 1\}\} = 1 - p,$$

for some $a > 1$, $b > 1$ and $p = [0, 1]$. Fix the index of regular variation to be $\alpha = 1$, and define $\Theta := (\Theta_t)_{t \in \mathbb{Z}}$ such that $P^\Theta = (P^W)^{RS}$. Note that Θ has only four possible realizations and that it vanishes for lags larger than 1 in absolute value. We say W is the shape of Θ , since standardizing a realization of Θ to sup-norm 1 and shifting the first non-zero observation to time 0 gives W .

We want to estimate the probability $P\{\Theta_i \in A\}$ for lags $i \in \{-1, 0, 1\}$ and half lines $A = (x, \infty)$ such that $P\{\Theta_i = x\} = 0$. The asymptotic variance of $\hat{p}_{n,A}$ can be calculated using Theorem 2.3 and the asymptotic variance of $\hat{p}_{n,A}^b$ and $\hat{p}_{n,A}^f$ can be obtained using Theorem 3.1 of Davis et al. (2018); see Supplement for details.

As shown in Figure 1, where we fix $a = 10, b = 2$, each of the three estimators can have largest or smallest asymptotic variance depending on the model parameter p (and a, b) and the probability we want to estimate (i.e. on i and x). (Observe that for lag $i = 0$ the backward estimator equals the forward estimator.) For $p \rightarrow 0$ or $p \rightarrow 1$, the asymptotic variance of the estimator $\hat{p}_{n,A}$ with known α (not shown in the plots) tends to 0 since the shape of the spectral tail process becomes deterministic. However, the asymptotic variance of $\hat{p}_{n,A}$ need not vanish due to the remaining variability of the Hill estimator.

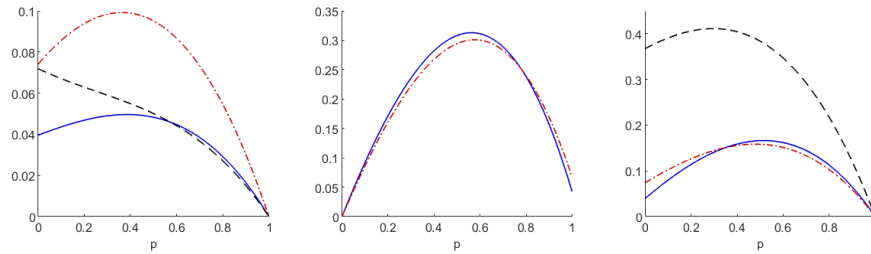


FIGURE 1. Variances of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) as a function of $p \in [0, 1]$, for lag $i = -1$ and $A = (3/4, \infty)$ (left), $i = 0$, $A = (1/20, \infty)$ (middle) and $i = 1$, $A = (1/20, \infty)$ (right).

2.1. Solutions to stochastic recurrence equations In this subsection we discuss solutions to stochastic recurrence equations, which can e.g. be used to analyze GARCH time series (cf. Basrak et al. (2002)). We derive conditions under which the previous theorems apply to this model. Consider a stationary solution to the stochastic recurrence equation

$$(2.6) \quad X_t = C_t X_{t-1} + D_t, \quad t \in \mathbb{Z},$$

where C_t are random $d \times d$ -matrices with non-negative entries and D_t are $[0, \infty)^d$ -valued random vectors such that (C_t, D_t) , $t \in \mathbb{Z}$, are i.i.d. In what follows, we denote both the Euclidean norm on $\mathbb{R}^d$ and the corresponding operator norm for $d \times d$ -matrices by $\|\cdot\|$, and (C, D) is a generic random variable with the same distribution as (C_1, D_1) .

It is well known (Buraczewski et al. (2016), Cor. 4.4.6) that under Condition (SRE) (see below) there exists an a.s. unique stationary solution to (2.6), namely

$$X_t = \sum_{k=0}^{\infty} \Pi_{t-k+1,t} D_{t-k}, \quad \Pi_{j,k} := C_k \cdots C_j$$

(with the convention $\Pi_{t+1,t}$ being equal to the identity matrix I_d) and this solution is regularly varying with index α .

- (SRE) (i) $\lim_{m \rightarrow \infty} m^{-1} E[\log \|\Pi_{1,m}\|] < 0$;
(ii) $\rho(s) := \lim_{m \rightarrow \infty} m^{-1} \log E[\|\Pi_{1,m}\|^s] \in (0, \infty)$ for some $s > 0$;
(iii) $E[\|C\|^\alpha \log^+ \|C\|] < \infty$, $E[\|D\|^\alpha] < \infty$ with α denoting the unique solution to $\rho(\alpha) = 0$;
(iv) $P\{D = 0\} < 1$, $P\{(C_{jk})_{1 \leq k \leq d} \neq 0\} = 1$, $\forall 1 \leq j \leq d$;
(v) the additive subgroup generated by the logarithms of the spectral radii of arbitrary finite products of matrices in the support of P^C is dense in $\mathbb{R}$.

Observe that under conditions (i) and (ii) the existence of a unique zero of the function ρ is guaranteed (Buraczewski et al. (2016), Lemma 4.4.2). According to Janssen & Segers (2014), Theorem 2.1 and Example 6.1, the forward spectral tail process has the representation $\Theta_t = \Pi_{1,t} \Theta_0$, for all $t \geq 0$, with Θ_0 independent of $(C_t)_{t \in \mathbb{Z}}$.

We now discuss conditions that ensure that the limit Theorems 2.1, 2.2 and 2.3 are applicable for the family of sets $\mathcal{A} = \{[0, x] \mid x \in [0, \infty)^d\} \cup \{[0, \infty)^d, \emptyset\}$, which is distribution determining for Θ_i , $i \geq 0$. We are going to apply results for general Markovian time series established by Kulik et al. (2019). To this end, we fix an arbitrary $q \in (0, \alpha)$ and verify Assumption 2.1 of this paper for $\mathbb{Y} := X$, the functions $g(x) = \|x\|$ and $V(x) = \|x\|^q + 1$ and $q_0 = q$. Note that, in view of the above discussion, parts (i), (ii) and (v) of Assumption 2.1 are obviously fulfilled.

If, in addition to (SRE), we assume the following set of conditions (SRE'), then, by Theorems 2.1 and 2.2 and Corollary 2.3 of Alsmeyer (2003) (see also Buraczewski et al. (2016), Prop. 4.2.1) $(X_t)_{t \in \mathbb{Z}}$ is an aperiodic, positive Harris recurrent P^{X_0} -irreducible Feller process:

- (SRE') (i) The interior of the support of P^{X_0} is not empty.
(ii) There is a σ -finite non-null measure ν on $\mathbb{R}^d$ and an open set $E_0 \in \mathbb{B}^d$ with $P\{X_0 \in E_0\} > 0$ such that P^{Cx+D} has a non-trivial component that is absolutely continuous w.r.t. ν for all $x \in E_0$.

Condition (ii) is e.g. fulfilled if (C, D) is absolutely continuous w.r.t. the Lebesgue measure; see Lemma 4.2.2 of Buraczewski et al. (2016) for further sufficient conditions. By Meyn & Tweedie (1992), Theorem 3.4, all compact subsets of $\mathbb{R}^d$ are petite sets, hence also small sets by Theorem 9.4.10 of Douc et al. (2018). Since $V^{-1}[0, u]$ is compact for all $u > 0$, part (iv) of Assumption 2.1 of Kulik et al. (2019) is fulfilled with $m = 1$. To verify part (iii), check that, for all $t \geq 0$,

$$(2.7) \quad X_t = \Pi_{1,t} X_0 + R_t$$

with $\Pi_{1,t}$ and X_0 independent, and $R_t := \sum_{j=1}^t \Pi_{j+1,t} D_j$. Moreover, $\rho(q) < 0$ for all $q \in (0, \alpha)$ by the Hölder inequality and $\rho(\alpha) = 0$, so that

$$(2.8) \quad \kappa := E(\|\Pi_{1,m}\|^q) < 1$$

for some sufficiently large m . Since $(a+b)^q \leq ((1+\eta)a)^q \mathbf{1}_{\{b \leq \eta a\}} + ((1+\eta)b/\eta)^q \mathbf{1}_{\{b > \eta a\}}$ for all $a, b \geq 0$ and $\eta > 0$, we may conclude

$$\begin{aligned} E(V(X_m) \mid X_0 = y) &= E[\|\Pi_{1,m}y + R_m\|^q] + 1 \\ &\leq E[(\|\Pi_{1,m}\| \|y\| + \|R_m\|)^q] + 1 \\ &\leq (1+\eta)^q \kappa (\|y\|^q + 1) + ((1+\eta)/\eta)^q E[\|R_m\|^q] + 1 \\ (2.9) \quad &=: \tilde{\beta} V(y) + \tilde{b} \end{aligned}$$

with $\tilde{\beta} < 1$ for η sufficiently small. By Douc et al. (2018), Proposition 9.2.13 and Theorem 9.3.11, for suitable $M > 0$, the set $V^{-1}[0, M]$ is accessible by the m -skeleton of the Markov chain. Thus Theorem 14.1.4 of Douc et al. (2018) shows that an analog to (2.9) holds if $\tilde{b}$ is replaced with $b\mathbb{1}_{V^{-1}[0, \tilde{M}]}$ for some $\tilde{M} > 0$ and possibly $\tilde{\beta}$ with some other constant $\beta \in (0, 1)$. Since, by the same reasoning as above, $E(\|X_1\|^q \mid X_0 = y) \leq (1 + \eta)^q E[\|C_1\|^q] \|y\|^q + ((1 + \eta)/\eta)^q E[\|R_1\|^q]$, Lemma 2.1 of Mikosch & Wintenberger (2014) yields part (iii) of Assumption 2.1 of Kulik et al. (2019).

By regular variation of $\|X_0\|$, the remaining part (vi) of Assumption 2.1 of Kulik et al. (2019) follows if we show $E[\|X_0\|^q \mathbb{1}_{\{\|X_0\| > u_n\}}] = O(u_n^q v_n)$. This, in turn, is an easy consequence of integration by parts and Karamata's theorem:

$$\begin{aligned} E[\|X_0\|^q \mathbb{1}_{\{\|X_0\| > u_n\}}] &= \int_{u_n}^{\infty} z^q P^{\|X_0\|}(dz) \\ &= -z^q P\{\|X_0\| > z\} \Big|_{z=u_n}^{\infty} + q \int_{u_n}^{\infty} z^{q-1} P\{\|X_0\| > z\} dz \\ &= u_n^q v_n + q \frac{u_n^q v_n}{\alpha - q} (1 + o(1)). \end{aligned}$$

The drift condition established above and irreducibility imply that the time series is geometrically β -mixing (see Kulik et al. (2019) or Mikosch & Wintenberger (2014), p. 161), i.e. there exist $\tau > 0$ and $\sigma \in (0, 1)$ such that $\beta_k \leq \tau \sigma^k$ for all $k \in \mathbb{N}$. It is easy to check that condition (S) is met, provided $v_n = o(1/\log n)$ and $(\log^2 n)/n = o(v_n)$ and the block size is chosen such that $s_n = o(\min(v_n^{-1}, (nv_n)^{1/2}))$ (for instance, with $l_n \geq \max(s_n, \log n / |2 \log \sigma|)$ and r_n of larger order than l_n , but of smaller order than $\min(v_n^{-1}, (nv_n)^{1/2})$).

Moreover, the following straightforward generalization of Lemma 4.3 of Kulik et al. (2019) holds: for all functions $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ that vanish on a neighborhood of 0 such that $|\psi(x)| \leq c\|x\|^{q/2}$ for some $c > 0$ and all $x \in \mathbb{R}^d$, one has

$$(2.10) \quad \lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \sum_{j=L+1}^{r_n} E[|\psi(X_0/u_n) \psi(X_j/u_n)|] = 0,$$

provided $r_n v_n \rightarrow 0$.

To verify Condition (BC), observe that the operator norm is sub-multiplicative and thus, with κ defined in (2.8),

$$\begin{aligned} E[\|\Pi_{1,k}\|^q] &\leq (E[\|\Pi_{1,m}\|^q])^{[k/m]} E[\|\Pi_{1,k-m[k/m]}\|^q] \\ &\leq (\kappa^{1/m})^k \max_{0 \leq j < m} \frac{E[\|\Pi_{1,j}\|^q]}{\kappa^j} =: \tilde{\kappa}^k c_m, \end{aligned}$$

for all $k \in \mathbb{N}$. Therefore, using (2.7), one may first deduce

$$(2.11) \quad P(\|X_h\| > cu_n \mid \|X_0\| > cu_n) \leq P\left\{\|X_0\| > \frac{cu_n}{2}\right\} + 2^q c_m \tilde{\kappa}^h E\left(\left(\frac{\|X_0\|}{cu_n}\right)^q \mid \|X_0\| > cu_n\right)$$

and then (BC) in the same way as condition (C) in Drees et al. (2015); see Supplement for details.

According to Remark 2.2 (b) it suffices to prove (M') (ii) instead of (TC). To this end we will need even stronger conditions on v_n and s_n :

(SRE'') $v_n = o((\log n)^{-(3+\zeta)})$ for some $\zeta > 0$, $(\log^2 n)/n = o(v_n)$, $s_n = o(\min(v_n^{-1/(3+\zeta)}, (nv_n)^{1/2}))$.

Choose some $q \in (\alpha/(1+\zeta/2), \alpha)$ and some $\tau \in (1/q, (1+\zeta/2)/\alpha)$, and define $\varepsilon_h := |h|^{-\tau}$ for $h \in \mathbb{Z}$, so that $(\varepsilon_h^q)_{h \in \mathbb{Z}}$ is summable. Then, with $v_{n,c} := P\{\|X_0\| > cu_n\}$,

$$\begin{aligned}
 & \sum_{h=-s_n}^{s_n} E\left(\left(\frac{\|X_h\|}{u_n}\right)^q \mathbf{1}_{\{\|X_h\| < u_n\}} \mid \|X_0\| > u_n\right) \\
 & \leq \sum_{h=-s_n}^{s_n} [P(\|X_h\| > \varepsilon_h u_n \mid \|X_0\| > u_n) + \varepsilon_h^q] \\
 (2.12) \quad & \leq 2 \sum_{h=1}^{s_n} P(\|X_h\| > \varepsilon_h u_n \mid \|X_0\| > \varepsilon_h u_n) \frac{v_{n,\varepsilon_h}}{v_n} + O(1).
 \end{aligned}$$

By (2.11), the h -th summand on the right hand side is bounded by

$$\left[v_{n,\varepsilon_h/2} + 2^q c_m \tilde{\kappa}^h E\left(\left(\frac{\|X_0\|}{\varepsilon_h u_n}\right)^q \mid \|X_0\| > \varepsilon_h u_n\right) \right] \frac{v_{n,\varepsilon_h}}{v_n}.$$

Check that, for all $1 \leq h \leq s_n$, $\varepsilon_h u_n \geq s_n^{-\tau} u_n$ is of larger order than $s_n^{-\tau} v_n^{-1/q} = (s_n v_n^{1/(\tau q)})^{-\tau}$ which tends to ∞ by condition (SRE'') because $\tau q < 1 + \zeta/2$. Therefore, the regular variation of $\|X_0\|$ combined with the Potter bounds shows that, for all $\alpha^* > \alpha$, (2.12) is bounded by

$$4 \sum_{h=1}^{s_n} \varepsilon_h^{-\alpha^*} \left[\left(\frac{\varepsilon_h}{2}\right)^{-\alpha^*} v_n + 2^q c_m \tilde{\kappa}^h E[\|Y_0\|^q] \right] + O(1) = O(s_n^{2\alpha^* \tau + 1} v_n) + O(1).$$

In view of (SRE''), this in turn is bounded if α^* is chosen such that $\alpha^* \tau \leq 1 + \zeta/2$. Now Condition (M') (ii) is obvious, since $|\log x|^{1+\delta} x^\alpha$ can be bounded by a multiple of x^q for all $x \in (0, 1)$.

So far we have verified all assumptions of Theorem 2.1 except for the bias condition (2.3), which is always fulfilled if u_n is chosen sufficiently large, and Condition (C Θ), which is obviously satisfied if Θ_i has continuous marginal distributions. In particular, if $i > 0$ and C is absolutely continuous, then this is also true for Θ_i and hence for its marginal distributions. More generally, our results apply to subsets of the family $\mathcal{A}$ where all sets $(-\infty, y]$ are omitted for which y_j belongs to a neighborhood of some jump point of the j -th marginal distribution for some $1 \leq j \leq d$.

For process convergence established in Theorem 2.2, in addition we need Condition (A). Parts (i) and (ii) are obvious with $A_t := \times_{i=1}^d [0, \varpi(t_i)] \cap \mathbb{R}^d$ for some continuous, increasing mapping ϖ from $[0, 1]$ onto $[-\infty, \infty]$. Then $\bigcap_{s \in (t, 1]} A_{s^{(k)}} \setminus A_{t^{(k)}} = \emptyset$, so that (A) (v), (vi) and (vii) are trivial, and (iv) is equivalent to (C Θ). Finally, part (iii) is fulfilled, because the processes are continuous from the right in each coordinate.

To apply Theorem 2.3 on the asymptotic normality of the estimator when α is unknown, we have to check the conditions (BC') and (M), while the bias condition (H) is again satisfied for u_n sufficiently large. The former can be proved in the same way as condition (2.9) of Drees & Knezevic (2020) using (2.10). Likewise, Condition (M) (i) corresponds to (2.11) of Drees & Knezevic (2020) and it can be established by completely analogous arguments, provided $r_n^{1+\zeta} v_n \rightarrow 0$ for some $\zeta > 0$, which holds by (SRE''); see Supplement for details.

Since (M') (ii) has already been verified, in view of Remark 2.2 (b), (M) (ii) follow if we establish (M') (i). Since, for all $\eta > 0$, $\psi(x) := \max((\log^+ x)^\eta, \mathbf{1}_{[1, \infty)}(x))$ can be bounded by a multiple of $x^{q/2}$ and the time series is stationary,

$$E\left(\sup_{L < |h| \leq s_n} \left(\log^+ \frac{\|X_h\|}{u_n}\right)^\eta \mid \|X_0\| > u_n\right) \leq \frac{1}{v_n} E\left[\sup_{L < |h| \leq s_n} \psi\left(\frac{\|X_0\|}{u_n}\right) \psi\left(\frac{\|X_h\|}{u_n}\right)\right]$$

$$\leq \frac{2}{v_n} \sum_{h=L+1}^{s_n} E[\psi(\|X_0\|/u_n)\psi(\|X_h\|/u_n)]$$

$$\leq 1$$

eventually if L is chosen sufficiently large; cf. (2.10). Because $E((\log^+(\|X_h\|/u_n))^\eta \mid \|X_0\| > u_n)$ is bounded for all fixed h due to the regular variation of the time series, the condition (M') (i) is obvious.

Finally, (M) (iii) can be established by the same arguments as (M') (ii). Since the denominator is at least 1 and $\log^-(x)x^\alpha \leq \tilde{c}_q x^q$ for $q < \alpha$ and some $\tilde{c}_q > 0$, the boundedness of (2.12) implies Condition (M) (iii).

To sum up, all limit theorems apply if the conditions (SRE), (SRE'), (SRE'') and the bias conditions (2.4) and (H) are met, and Θ_i have continuous marginal distributions.

3. Simulations In this section we present the results of a Monte Carlo simulation study to evaluate the finite sample performance of the projection based estimator $\hat{p}_{n,A}$ with estimated α . As competitors we consider the forward estimator $\hat{p}_{n,A}^f$, see (1.2), and the backward estimator $\hat{p}_{n,A}^b$, see (1.3). We study their performance in two well-known real valued time series models:

- **GARCHt:** Firstly, we consider a GARCH(1,1) time series as a typical model for financial data: $X_t = \sigma_t \epsilon_t$ with $\sigma_t^2 = 0.1 + 0.14X_{t-1}^2 + 0.84\sigma_{t-1}^2$ and independent innovations ϵ_t with Student's t_4 -distribution, standardized to unit variance. These parameters ensure that a stationary, regularly varying solution with index $\alpha = 2.6$ exists, see Mikosch & Stărică (2000), Section 2.2, and Davis et al. (2018), Section 4. Its forward spectral tail process $(\Theta_t)_{t \geq 0}$ can be represented as follows:

$$\Theta_t = \frac{\tilde{\epsilon}_t}{|\tilde{\epsilon}_0|} \prod_{i=1}^t (0.14\tilde{\epsilon}_{t-i}^2 + 0.84)^{1/2}, \quad t \geq 0,$$

where $\tilde{\epsilon}_h$, $h \geq 1$, are i.i.d. with a Student's t_4 -distribution, standardized to unit variance; we denote the corresponding density by f_ϵ . The random variable $\tilde{\epsilon}_0$ is independent of $(\tilde{\epsilon}_t)_{t \geq 1}$ with density $x \mapsto f_\epsilon(x)|x|^{2.6}/E[|\epsilon_1|^{2.6}]$, see Proposition 6.2 of Ehlert et al. (2015).

- **SRE:** The second model is defined as a stationary solution to the univariate stochastic recurrence equation $X_t = C_t X_{t-1} + D_t$, $t \in \mathbb{Z}$, with i.i.d. $\mathbb{R}^2$ -valued (C_t, D_t) . We choose C_t and D_t to be independent with $C_t \sim \mathcal{N}(1/3, 8/9)$ and $D_t \sim \mathcal{N}(-10, 1)$; see Drees et al. (2015), Sections 5.2 and 6 for a discussion of this model. (We have chosen this model, which does not exactly fit in the setting of Section 2.1, to enable a comparison with this reference.) The unique stationary solution is regularly varying with index $\alpha = 2$. Moreover, the forward spectral tail processes is of the form

$$\Theta_t = \Theta_0 \prod_{h=1}^t C_h, \quad t \geq 0,$$

where Θ_0 is uniformly distributed on $\{-1, 1\}$ and independent of $(C_h)_{h \geq 1}$; see Janssen & Segers (2014) and Goldie (1991), Theorem 4.1.

We consider the family of sets $A = (-\infty, x]$, $x \in \mathbb{R}$; the corresponding probabilities describe the cdf of Θ_i . For all lags $i \in \{1, \dots, 10\}$, the cdf of Θ_i is approximately calculated on $[-2, 2]$ (in steps of 0.01) via 10^7 Monte Carlo repetitions. For 1000 time series of length 2000 we calculate the three estimates of $P\{\Theta_i \leq x\}$, $x \in [-2, 2]$. In each

simulation, we choose the empirical 95% quantile of the absolute values as threshold u_n .

In addition, for the projection based estimator $\hat{p}_{n,A}$ one has to choose the block length, determined by s_n . If s_n is chosen too small, then the estimated distribution of Θ_i has substantial mass at 0, an artefact caused by the cut off of the time series at lags larger than s_n in absolute value in (1.5). In the results presented below, we have chosen $s_n = 30$, but in most cases the performance of the projection based estimator turned out to be quite stable for a wide range of values of s_n including the interval $\{20, \dots, 50\}$; see Supplement for details.

First we examine the GARCHt model for lag $i = 10$, where the Hill type estimator of α exhibits a bias of about 0.014 and a standard deviation of about 0.461. The top row of Figure 2 shows mean and standard deviation of all three estimators. In addition, the relative efficiencies of the forward and backward estimator w.r.t. the newly proposed estimator, i.e. the ratio between the root mean squared error (RMSE) of the respective estimator and the RMSE of $\hat{p}_{n,A}$, is shown in the right column. By construction, both the projection based estimator and the backward estimator of the distribution of Θ_i typically have mass at 0, leading to a bias of the estimated cdf at small values of $|x|$ and thus to an underperformance w.r.t. the forward estimator for $|x| < 0.2$. In contrast, for larger values of $|x|$ the relative efficiency of the forward estimator w.r.t. $\hat{p}_{n,A}$ is less than 1/2, and the backward estimator is outperformed by the newly proposed estimator for all x .

The bottom row of Figure 2 shows the results for the SRE model and lag $i = 1$, when the estimator of α has bias ≈ 0.20 and standard deviation ≈ 0.385 . For all three estimators we observe a strong bias caused by D_t , which adds a negative drift to the time series on finite levels, while it vanishes in the limit. The standard deviation of $\hat{p}_{n,A}$ is smaller than or equal to that of $\hat{p}_{n,A}^f$, but as the RMSE is dominated by the bias (which is somewhat smaller for $\hat{p}_{n,A}^f$), the relative efficiency of the forward estimator is larger than 1 for most values of x . In contrast, the backward estimator, with its huge standard deviation

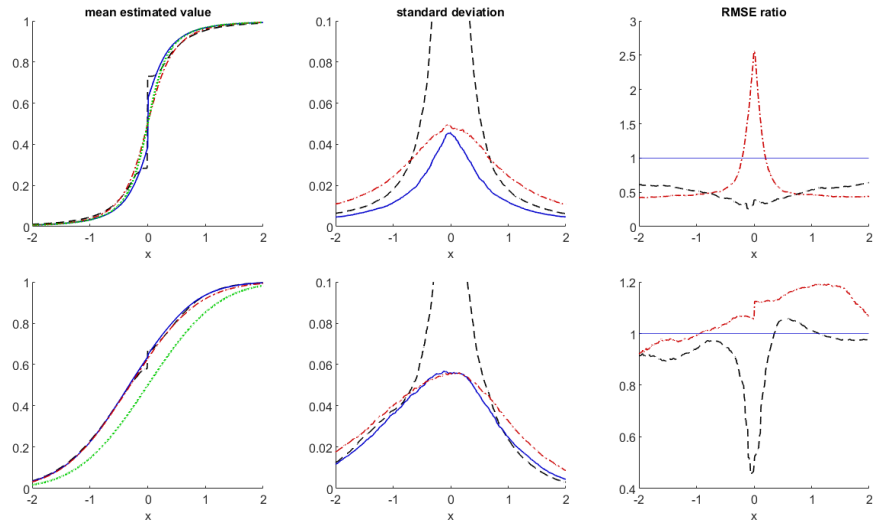


FIGURE 2. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right), of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) for GARCHt model and lag $i = 10$ (top) and SRE model and lag $i = 1$ (bottom); the true cdf is indicated by the green dotted line.

near the origin, is less efficient than the projection based estimator for most x . If one evaluates the performance of $\hat{p}_{n,A}$, $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ as estimators of the “pre-asymptotic” probabilities $P(X_i/|X_0| \leq x \mid |X_0| > u_n)$ (as it has been done in [Davis et al. \(2018\)](#)), then the bias almost vanishes and for most values of x the projection based estimator outperforms the forward estimator; see Supplement for details.

Further simulation results for other lags, thresholds and models can be found in the Supplement. Overall, the projection based estimator tends to have the smallest variance of the three estimators, in particular for higher lags i and larger values of $|x|$. Usually it shows the most robust behavior which combines the good performance of the forward estimators for small values of $|x|$ with the improvements of the backward estimator for large $|x|$.

APPENDIX A: PROOFS

A.1. Proof of Theorem 2.1. First we prove weak convergence of the empirical processes $(Z_n^{(T)}(A))_{A \in \mathcal{A} \cup \{\mathbb{R}^d\}}$ associated with the sums $T_{n,A}$ of sliding blocks statistics defined by

$$Z_n^{(T)}(A) = \frac{1}{\sqrt{nv_n}} (T_{n,A} - E[T_{n,A}]).$$

PROPOSITION A.1. If the conditions (RV), (S), (BC), (TC) and (C Θ) are satisfied, then the fidis of $Z_n^{(T)}$ weakly converge to the fidis of the centered Gaussian process Z defined in Theorem 2.1.

To prove this proposition, we want to apply Theorem 2.1 of [Drees & Neblung \(2021\)](#) (abbreviated to D&N in what follows) to empirical processes based on the transformed observations $X'_{n,t} := X_{t-s_n}/u_n$, $1 \leq t \leq n' := n + 2s_n$, and block lengths $s'_n := 2s_n + 1$ and $l'_n := 2s'_n - 1$. We will consider the family $\mathcal{G} = \{g_A \mid A \in \mathcal{A} \cup \{\mathbb{R}^d\}\}$ of bounded functions g_A taking values in $[0, 1]$, and use the normalization $b_n(g) = (nv_n/p_n)^{1/2}$ for all $g \in \mathcal{G}$ with $p_n = P\{\max_{1 \leq t \leq r_n} \|X_t\| > u_n\}$.

We start with a verification of the convergence of the normalized covariances, i.e. condition (C) of D&N. To this end, we need some technical lemmas. The proof of the first lemma is given in the Supplement.

LEMMA A.2. If (RV) and (BC) holds, then the so-called anti clustering condition

$$(A.1) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\max_{m \leq |t| \leq r_n} \|X_t\| > cu_n \mid \|X_0\| > cu_n\right) = 0$$

is satisfied for all $c \in (0, \infty)$.

LEMMA A.3. If (RV) holds, then condition (C Θ) is satisfied if and only if

$$P\left\{\exists t \in \mathbb{Z} : \frac{Y_{t+i}}{\|Y_t\|} \in \partial A, \|Y_t\| > 0\right\} = 0, \quad \forall A \in \mathcal{A}.$$

PROOF. Since $P^\Theta = (P^\Theta)^{RS}$ and $\Theta_t = Y_t/\|Y_0\|$, $P\{\Theta_i \in \partial A\} = 0$ holds if and only if

$$0 = P\left\{\sum_{t \in \mathbb{Z}} \frac{\|\Theta_t\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_{\partial A}\left(\frac{\Theta_{t+i}}{\|\Theta_t\|}\right) > 0\right\} = P\left\{\exists t \in \mathbb{Z} : \frac{Y_{t+i}}{\|Y_t\|} \in \partial A, \|Y_t\| > 0\right\}. \quad \square$$

In the following proofs we consider functions on the sequence space $l_\alpha = \{(x_t)_{t \in \mathbb{Z}} \mid \sum_{t \in \mathbb{Z}} \|x_t\|^\alpha < \infty\}$ (equipped with the supremum norm). For arbitrary $r \in \mathbb{N}$, $\mathbb{R}^{2r+1}$ is embedded in l_α by the mapping $(z_i)_{|i| \leq r} \mapsto (z_t)_{t \in \mathbb{Z}}$ with $z_t := 0$ for $|t| > r$. Note that (A.1) ensures that the realizations of the spectral tail process a.s. belong to l_α (cf. Remark 2.3 of JanBen (2019)). The following result establishes condition (C) of D&N.

LEMMA A.4. If the conditions (RV), (S), (BC), (TC) and (C Θ) are satisfied, then

$$\frac{1}{r_n v_n} \text{Cov} \left(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t}) \right) \rightarrow c(A, B), \quad \forall A, B \in \mathcal{A} \cup \{\mathbb{R}^d\},$$

with c defined in Theorem 2.1.

PROOF. First note that

$$\begin{aligned} & \text{Cov} \left(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t}) \right) \\ &= E \left[\sum_{t=1}^{r_n} \sum_{s=1}^{r_n} g_A(W_{n,t}) g_B(W_{n,s}) \right] - E \left[\sum_{t=1}^{r_n} g_A(W_{n,t}) \right] E \left[\sum_{t=1}^{r_n} g_B(W_{n,t}) \right] =: I + II, \end{aligned}$$

where

$$(A.2) \quad \frac{|II|}{r_n v_n} \leq \frac{1}{r_n v_n} \left(E \left[\sum_{t=1}^{r_n} \mathbf{1}_{\{\|X_t\| > u_n\}} \right] \right)^2 = r_n v_n \rightarrow 0.$$

By stationarity

$$\begin{aligned} I &= \sum_{s=1}^{r_n} \sum_{t=1}^{r_n} E[g_A(W_{n,s}) g_B(W_{n,t})] \\ &= \sum_{j=-r_n}^{r_n} (r_n - |j|) E[g_A(W_{n,j}) g_B(W_{n,0})] \\ &= r_n v_n \sum_{j=-r_n}^{r_n} \left(1 - \frac{|j|}{r_n} \right) E \left(\mathbf{1}_{\{\|X_j\| > u_n\}} \right. \\ &\quad \times \left(\sum_{h \in H_n} \frac{\|X_{j+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{j+k}\|^\alpha} \mathbf{1}_A \left(\frac{X_{j+h+i}}{\|X_{j+h}\|} \right) + \sum_{h \in H_n^c} \frac{\|X_{j+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{j+k}\|^\alpha} \mathbf{1}_A(0) \right) \\ &\quad \times \left(\sum_{l \in H_n} \frac{\|X_l\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_k\|^\alpha} \mathbf{1}_B \left(\frac{X_{l+i}}{\|X_l\|} \right) + \sum_{l \in H_n^c} \frac{\|X_l\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_k\|^\alpha} \mathbf{1}_B(0) \right) \Big| \|X_0\| > u_n \Big). \end{aligned} \quad (A.3)$$

Define a function $f = f_{A,B} : l_\alpha \times l_\alpha \rightarrow [0, 1]$ by

$$\begin{aligned} f((y_t)_{t \in \mathbb{Z}}, (z_t)_{t \in \mathbb{Z}}) &:= \mathbf{1}_{\{\|y_0\| > 1\}} \mathbf{1}_{\{\|z_0\| > 1\}} \\ &\quad \times \left(\sum_{h \in \mathbb{Z}} \frac{\|z_h\|^\alpha}{\sum_{k \in \mathbb{Z}} \|z_k\|^\alpha} \mathbf{1}_A \left(\frac{z_{h+i}}{\|z_h\|} \right) \right) \left(\sum_{l \in \mathbb{Z}} \frac{\|y_l\|^\alpha}{\sum_{k \in \mathbb{Z}} \|y_k\|^\alpha} \mathbf{1}_B \left(\frac{y_{l+i}}{\|y_l\|} \right) \right), \end{aligned}$$

with the convention $0/0 := 0$. Using Condition (TC), one may replace sums over $\{-s_n, \dots, s_n\}$ with sums over $\{-m, \dots, m\}$ to conclude from (1.1) and (C Θ) that

$$(A.4) \quad E(f((X_t/u_n)_{-s_n \leq t \leq s_n}, (X_{t+j}/u_n)_{-s_n \leq t \leq s_n}) \mid \|X_0\| > u_n) \rightarrow E[f((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}})];$$

see Supplement for details. Moreover, condition (BC) implies

$$\begin{aligned} E(f((X_t/u_n)_{-s_n \leq t \leq s_n}, (X_{t+j}/u_n)_{-s_n \leq t \leq s_n}) \mid \|X_0\| > u_n) &\leq E(\mathbf{1}_{\{\|X_j\| > u_n\}} \mid \|X_0\| > u_n) \\ &\leq e_{n,1}(|j|). \end{aligned}$$

Hence, using Pratt's lemma (Pratt 1960), we may conclude from (A.3) and (A.4) that

$$\begin{aligned} \frac{1}{r_n v_n} I &= \sum_{j=-r_n}^{r_n} \left(1 - \frac{|j|}{r_n}\right) E(f((X_t/u_n)_{-s_n \leq t \leq s_n}, (X_{t+j}/u_n)_{-s_n \leq t \leq s_n}) \mid \|X_0\| > u_n) \\ &\rightarrow \sum_{j \in \mathbb{Z}} E[f((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}})]. \end{aligned}$$

To sum up, we have shown that

$$\begin{aligned} &\frac{1}{r_n v_n} \text{Cov}\left(\sum_{t=1}^{r_n} g_A(W_{n,t}), \sum_{t=1}^{r_n} g_B(W_{n,t})\right) \\ &\rightarrow \sum_{j \in \mathbb{Z}} E\left[\mathbf{1}_{\{\|Y_j\| > 1\}} \left(\sum_{h \in \mathbb{Z}} \frac{\|Y_{j+h}\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{Y_{j+h+i}}{\|Y_{j+h}\|}\right)\right) \left(\sum_{l \in \mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right] \\ (A.5) \quad &= \sum_{j \in \mathbb{Z}} E\left[\mathbf{1}_{\{\|Y_j\| > 1\}} \left(\sum_{k \in \mathbb{Z}} \frac{\|Y_k\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{Y_{k+i}}{\|Y_k\|}\right)\right) \left(\sum_{l \in \mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_B\left(\frac{Y_{l+i}}{\|Y_l\|}\right)\right)\right]. \end{aligned}$$

Since $Y_t = \Theta_t \|Y_0\|$ with $(\Theta_t)_{t \in \mathbb{Z}}$ and $\|Y_0\|$ independent and $P\{\|Y_0\| > x\} = x^{-\alpha} \wedge 1$, the limit can be rewritten as

$$\begin{aligned} &\sum_{j \in \mathbb{Z}} E\left[\mathbf{1}_{\{\|\Theta_j\| \|Y_0\| > 1\}} \left(\sum_{k \in \mathbb{Z}} \frac{\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_B\left(\frac{\Theta_{k+i}}{\|\Theta_k\|}\right)\right) \left(\sum_{l \in \mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right] \\ &= \sum_{j \in \mathbb{Z}} \int_{l_\alpha} P\{\|Y_0\| > \|\theta_j\|^{-1}\} \left(\sum_{k \in \mathbb{Z}} \frac{\|\theta_k\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbf{1}_B\left(\frac{\theta_{k+i}}{\|\theta_k\|}\right)\right) \left(\sum_{l \in \mathbb{Z}} \frac{\|\theta_l\|^\alpha}{\|\theta\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{\theta_{l+i}}{\|\theta_l\|}\right)\right) P^\Theta(d\theta) \\ &= \sum_{j \in \mathbb{Z}} E\left[(\|\Theta_j\|^\alpha \wedge 1) \left(\sum_{k \in \mathbb{Z}} \frac{\|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_B\left(\frac{\Theta_{k+i}}{\|\Theta_k\|}\right)\right) \left(\sum_{l \in \mathbb{Z}} \frac{\|\Theta_l\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A\left(\frac{\Theta_{l+i}}{\|\Theta_l\|}\right)\right)\right]. \quad \square \end{aligned}$$

PROOF OF PROPOSITION A.1. As mentioned above, the assertion follows readily if we verify the conditions of Theorem 2.1 of D&N. Note that p_n is of the same order as $r_n v_n$ according to Proposition 4.2 of Basrak & Segers (2009), since (BC) implies the anti clustering condition (A.1) (which in fact just controls the cluster size). Hence the conditions (A1), (A2) and (MX) of D&N immediately follow from our condition (S) and stationarity. Moreover, condition (D0) of D&N is trivially fulfilled for finite families of sets $A \in \mathcal{A}$, which suffices here.

As $\mathbf{1}_{\{g(W_{n,j}) \neq 0\}} \leq \mathbf{1}_{\{\|X_j\| > u_n\}} = \mathbf{1}_{\{g_{\mathbb{R}^d}(W_{n,j}) \neq 0\}}$ for all $j = 1, \dots, r_n$ and all $g \in \mathcal{G}$, condition (2.6) of D&N is an obvious consequence of Lemma A.4 which yields

$$(A.6) \quad N_n := \frac{1}{r_n v_n} E\left[\left(\sum_{j=1}^{r_n} \mathbf{1}_{\{\|X_j\| > u_n\}}\right)^2\right] = O(1).$$

Since the remaining condition (C) of D&N has been proved in Lemma A.4, the assertion follows. $\square$

PROOF OF THEOREM 2.1. Since $\hat{p}_{n,A} = T_{n,A}/T_{n,\mathbb{R}^d}$ and $E[T_{n,A}] = nv_n E(g_A(W_{n,0}) \mid \|X_0\| > u_n)$, direct calculations show that

$$\begin{aligned} & \sqrt{nv_n}(\hat{p}_{n,A} - E(g_A(W_{n,0}) \mid \|X_0\| > u_n)) \\ &= \sqrt{nv_n} \left(\frac{\sqrt{nv_n} Z_n^{(T)}(A) + nv_n E(g_A(W_{n,0}) \mid \|X_0\| > u_n)}{\sqrt{nv_n} Z_n^{(T)}(\mathbb{R}^d) + nv_n} - E(g_A(W_{n,0}) \mid \|X_0\| > u_n) \right) \\ &= \frac{Z_n^{(T)}(A) - E(g_A(W_{n,0}) \mid \|X_0\| > u_n) Z_n^{(T)}(\mathbb{R}^d)}{1 + (nv_n)^{-1/2} Z_n^{(T)}(\mathbb{R}^d)} \\ &\rightarrow Z(A) - P\{\Theta_i \in A\} Z(\mathbb{R}^d) = Z^{(p)}(A) \end{aligned}$$

uniformly over any finite family of sets $A \in \mathcal{A}$, where in the last step we have used Proposition A.1, (2.1) and $nv_n \rightarrow \infty$. $\square$

A.2. Proof of Theorem 2.2. The assertion can be concluded from the following proposition by the continuous mapping argument used in the proof of Theorem 2.1.

PROPOSITION A.5. If the conditions (RV), (S), (BC), (TC), (C Θ) and (A) are fulfilled, then $Z_n^{(T)}$ weakly converge to the Gaussian process Z defined in Theorem 2.1.

PROOF. We are going to apply Theorem 2.3 (i) of Drees & Neblung (2021) (D&N). Since condition (D0) of D&N follows immediately from condition (A) (iii), and the other conditions used in Theorem 2.1 of D&N have already been verified in the proof of Proposition A.1, only the conditions (D1) and (D2) of D&N remain to be shown for some semi-metric on $\mathcal{G}$ such that $\mathcal{G}$ is totally bounded.

First we verify that $\mathcal{G}$ is totally bounded w.r.t. $\tilde{\rho}(g_A, g_B) := \rho(A, B)$, where the semi-metric ρ on $\mathcal{A}$ is given by

$$\rho(A, B) := \sum_{j \in \mathbb{Z}} E \left[\mathbb{1}_{\{\|Y_j\| > 1\}} \sum_{h \in \mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|^\alpha} \left| \mathbb{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) - \mathbb{1}_B \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \right| \right].$$

Note that $\rho(A, B) = \rho(A \setminus B, \emptyset)$ for $B \subset A$ and $\rho(\bigcup_{n \in \mathbb{N}} B_n, \emptyset) = \sum_{n \in \mathbb{N}} \rho(B_n, \emptyset)$ for disjoint sets $B_n \in \mathbb{B}^d$, $n \in \mathbb{N}$.

Fix $\delta > 0$ and, for the time being, $k \in \{1, \dots, q\}$. Recall that, for $t \in [0, 1 + \iota]$, $t^{(k)} \in [0, 1]^q$ denotes the vector with k -th coordinate equal to $t \wedge 1$ and all other coordinates equal to 1. By condition (A) (ii) the mapping $H_k : t \mapsto \rho(A_{t^{(k)}}, \emptyset)$ is non-decreasing. Observe that the arguments used in the proof of Lemma A.3 applied to condition (A) (v) show that $P\{\exists h \in \mathbb{Z} : Y_{h+i}/\|Y_h\| \in \bigcap_{s \in (t, 1]} A_{s^{(k)}} \setminus A_{t^{(k)}}\} = 0$ for all $t \in [0, 1)$. By the monotone convergence theorem, this implies that H_k is right-continuous and thus it is the measure generating function of a measure on $[0, 1]$, which is finite because $\rho(A_1, \emptyset) = c(\mathbb{R}^d, \mathbb{R}^d) < \infty$ by Lemma A.4. Thus, there exist $J_k < \infty$ and a partition $\mathcal{T}_k := \{[s_{k,j-1}, s_{k,j}] \mid 1 \leq j \leq J_k\} \cup \{\{1\}\}$ of $[0, 1]$ (with $s_{k,0} = 0$ and $s_{k,J_k} = 1$) such that $H_k(s_{k,j}) - H_k(s_{k,j-1}) \leq \delta$ for all $1 \leq j \leq J_k$. (For instance, one may define the interval boundaries iteratively by $s_{k,j} = \inf \{t \in (s_{k,j-1}, 1] \mid \rho(A_{t^{(k)}}, A_{s_{k,j-1}^{(k)}}) > \delta\}$. Observe that although it is possible that the measure pertaining to H_k has mass greater than δ at some of the boundary points $s_{k,j}$, such jumps of H_k do not play any role in the following calculations.)

We now consider the finite cover of $[0, 1]^q$ by sets in

$$\mathcal{T}^{(\delta)} := \{\times_{k=1}^q T_k \mid T_k \in \mathcal{T}_k, \forall 1 \leq k \leq q\}.$$

For any set $T = \times_{k=1}^q T_k \in \mathcal{T}^{(\delta)}$ define $K_T := \{1 \leq k \leq q \mid T_k \neq \{1\}\}$,

$$(A.7) \quad \underline{A}_T := \bigcap_{s \in T} A_s = A_{(\min T_1, \dots, \min T_q)} \in \mathcal{A} \quad \text{and} \quad \bar{A}_T := \bigcup_{s \in T} A_s \in \tilde{\mathcal{A}} := \{A_t^- \mid t \in [0, 1 + \iota]^q\}.$$

Check that $\bar{A}_T \setminus \underline{A}_T \subset \bigcup_{k \in K_T} (\bigcup_{s \in T_k} A_{s^{(k)}} \setminus A_{(\min T_k)^{(k)}})$. Hence, by construction of $\mathcal{T}^{(\delta)}$, for all $t \in T$,

$$(A.8) \quad \rho(A_t, \underline{A}_T) \leq \rho(\bar{A}_T \setminus \underline{A}_T, \emptyset) \leq \sum_{k \in K_T} H_k(\sup T_k -) - H_k(\min T_k) \leq q\delta.$$

Therefore, all sets of the form $\{A_t \mid t \in T\}$ have a radius of at most $q\delta$ w.r.t. ρ , which shows that $\mathcal{A}$ is totally bounded w.r.t. ρ and thus $\mathcal{G}$ is totally bounded w.r.t. $\tilde{\rho}$.

Next we turn to the continuity condition (D1) of D&N. Let $V_n(g_A) := \sum_{t=1}^{r_n} g_A(W_{n,t})$.

According to (A.2) and (A.5), one has

$$\begin{aligned} & \frac{1}{r_n v_n} E \left[(V_n(g_A) - V_n(g_B))^2 \right] \\ & \rightarrow \sum_{j \in \mathbb{Z}} E \left[\mathbb{1}_{\{\|Y_j\| > 1\}} \left(\sum_{l \in \mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \left(\mathbb{1}_A \left(\frac{Y_{l+i}}{\|Y_l\|} \right) - \mathbb{1}_B \left(\frac{Y_{l+i}}{\|Y_l\|} \right) \right) \right)^2 \right] \\ & \leq \sum_{j \in \mathbb{Z}} E \left[\mathbb{1}_{\{\|Y_j\| > 1\}} \sum_{l \in \mathbb{Z}} \frac{\|Y_l\|^\alpha}{\|Y\|_\alpha^\alpha} \left| \mathbb{1}_A \left(\frac{Y_{l+i}}{\|Y_l\|} \right) - \mathbb{1}_B \left(\frac{Y_{l+i}}{\|Y_l\|} \right) \right| \right] = \rho(A, B) \end{aligned}$$

for all $A, B \in \mathcal{A}$. In view of condition (A) (iv), one can prove by the same arguments that this convergence even holds for all $A, B \in \mathcal{A} \cup \tilde{\mathcal{A}}$. In particular, for all fixed $\delta, \varepsilon > 0$ one has eventually

$$(A.9) \quad \frac{1}{r_n v_n} E \left[(V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S}))^2 \right] \leq \rho(\bar{A}_T, \underline{A}_S) + \frac{\varepsilon}{2}$$

uniformly for all sets $S, T \in \mathcal{T}^{(\delta)}$.

Now consider two sets $A_s, A_t \in \mathcal{A}$ with $\rho(A_s, A_t) \leq \delta$. Because $\mathcal{T}^{(\delta)}$ defines a partition of $[0, 1]^q$, there exist unique sets $S, T \in \mathcal{T}^{(\delta)}$ such that $s \in S$ and $t \in T$. Then, obviously $\underline{A}_T \subset A_t \subset \bar{A}_T$ and $\underline{A}_S \subset A_s \subset \bar{A}_S$. Hence, we may conclude from (A.9) and (A.8) that eventually

$$\begin{aligned} & \frac{1}{r_n v_n} E \left[(V_n(g_{A_t}) - V_n(g_{A_s}))^2 \right] \\ & \leq \frac{1}{r_n v_n} E \left[\max \left((V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S}))^2, (V_n(g_{\underline{A}_T}) - V_n(g_{\bar{A}_S}))^2 \right) \right] \\ & \leq \frac{1}{r_n v_n} E \left[(V_n(g_{\bar{A}_T}) - V_n(g_{\underline{A}_S}))^2 \right] + \frac{1}{r_n v_n} E \left[(V_n(g_{\underline{A}_T}) - V_n(g_{\bar{A}_S}))^2 \right] \\ & \leq \rho(\bar{A}_T, \underline{A}_S) + \rho(\underline{A}_T, \bar{A}_S) + \varepsilon \\ & \leq \rho(\bar{A}_T, A_t) + \rho(A_t, A_s) + \rho(A_s, \underline{A}_S) + \rho(\underline{A}_T, A_t) + \rho(A_t, A_s) + \rho(A_s, \bar{A}_S) + \varepsilon \\ & \leq 2(\rho(\bar{A}_T, \underline{A}_T) + \rho(\bar{A}_S, \underline{A}_S) + \rho(A_t, A_s)) + \varepsilon \\ & \leq (4q + 2)\delta + \varepsilon. \end{aligned}$$

Therefore,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{s, t \in [0, 1]^q : \rho(A_s, A_t) \leq \delta} \frac{1}{r_n v_n} E \left[(V_n(g_{A_s}) - V_n(g_{A_t}))^2 \right] = 0$$

and condition (D1) is satisfied.

Finally we turn to condition (D2) of D&N. By condition (A) (vii) there exists $w = (w_k)_{1 \leq k \leq q} \in [0, 1]^q$ such that $0 \in A_w$, but $0 \notin A_s$ for all $s \prec w$. By the monotonicity assumption (A) (ii) it follows that $0 \in A_s$ if $s \geq w$. Next we show that here even equivalence holds. To this end, suppose that $0 \in A_s$ for some s for which $s \geq w$ does not hold. Then on the one hand $0 \in A_s \setminus A_{s \wedge w}$, but on the other hand $0 \notin A_{s \vee w} \setminus A_w$. However, this contradicts condition (A) (ii), which implies that $A_s \setminus A_{s \wedge w} \subset A_{s \vee w} \setminus A_w$, as can be easily seen by successively replacing each $s_k < w_k$ with w_k .

Fix some $n \in \mathbb{N}$ and let $\varepsilon > 0$ be arbitrary. In the same way as above, one may conclude from conditions (A) (ii) and (vi) that, for all $1 \leq k \leq q$, the function $F_k : [0, 1] \rightarrow \mathbb{R}$,

$$t \mapsto \frac{1}{r_n v_n} E \left[\left(\sum_{l=1}^{r_n} \mathbb{1}_{\{\|X_l\| > u_n\}} \sum_{h \in H_n} \frac{\|X_{l+h}\|^\alpha}{\sum_{j=-s_n}^{s_n} \|X_{l+j}\|^\alpha} \mathbb{1}_{A_{t(k)}} \left(\frac{X_{l+h+i}}{\|X_{l+h}\|} \right) \right)^2 \right]$$

is the measure generating function of a measure on $[0, 1]$ with mass $F_k(1) \leq N_n$ (cf. (A.6)). Hence, one can choose $t_{k,j} \in [0, 1]$, $0 \leq j \leq J_k^* \leq \lceil N_n(q/\varepsilon)^2 \rceil + 1$, such that $t_{k,0} := 0$, $t_{k,J_k^*} = 1$, $w_k \in \{t_{k,j} | 1 \leq j \leq J_k^*\}$, $t_{k,j-1} < t_{k,j}$ and $F_k(t_{k,j}) - F_k(t_{k,j-1}) \leq (\varepsilon/q)^2$ for all $1 \leq j \leq J_k^*$. These points define a partition $\mathcal{T}_k^* := \{[t_{k,j-1}, t_{k,j}) | 1 \leq j \leq J_k^*\} \cup \{\{1\}\}$ of $[0, 1]$ and hence $\mathcal{T}^* := \{\times_{k=1}^q T_k | T_k \in \mathcal{T}_k^*, \forall 1 \leq k \leq q\}$ is a partition of $[0, 1]^q$.

Now we define brackets

$$\mathcal{A}_T^{\varepsilon,n} := \{A_t | t \in T\}$$

for all $T \in \mathcal{T}^*$, so that the set $\mathcal{A}^{\varepsilon,n} = \{\mathcal{A}_T^{\varepsilon,n} | T \in \mathcal{T}^*\}$ is a partition of $\mathcal{A}$. Since $0 \in A_s$ if and only if $s \geq w$, the above construction ensures that, for all $\bar{A} \in \mathcal{A}^{\varepsilon,n}$, either all or none of the sets in $\bar{A}$ contain 0. Consequently, $\mathbb{1}_{\{0 \in A\}} = \mathbb{1}_{\{0 \in B\}}$ for all $A, B \in \bar{A}$.

Using the Cauchy-Schwarz inequality for sums and $(a-b)^2 \leq a^2 - b^2$ for all $a \geq b \geq 0$, we conclude similarly as above that, for all $T = \times_{k=1}^q T_k \in \mathcal{T}^*$ and $A_{T_k}^+ := \bigcup_{s \in T_k} A_s(k)$,

$$\begin{aligned} & \frac{1}{r_n v_n} E \left[\sup_{A, B \in \mathcal{A}_T^{\varepsilon,n}} \left(\sum_{l=1}^{r_n} (g_A(W_{n,l}) - g_B(W_{n,l})) \right)^2 \right] \\ & \leq \frac{1}{r_n v_n} E \left[\left(\sum_{l=1}^{r_n} (g_{\bar{A}_T}(W_{n,l}) - g_{\underline{A}_T}(W_{n,l})) \right)^2 \right] \\ & \leq \frac{1}{r_n v_n} E \left[\left(\sum_{k \in K_T} \sum_{l=1}^{r_n} \mathbb{1}_{\{\|X_l\| > u_n\}} \sum_{h \in H_n} \frac{\|X_{l+h}\|^\alpha}{\sum_{j=-s_n}^{s_n} \|X_{l+j}\|^\alpha} \mathbb{1}_{A_{T_k}^+ \setminus A_{(\min T_k)(k)}} \left(\frac{X_{l+h+i}}{\|X_{l+h}\|} \right) \right)^2 \right] \\ & \leq \frac{|K_T|}{r_n v_n} \sum_{k \in K_T} E \left[\left(\sum_{l=1}^{r_n} \mathbb{1}_{\{\|X_l\| > u_n\}} \sum_{h \in H_n} \frac{\|X_{l+h}\|^\alpha}{\sum_{j=-s_n}^{s_n} \|X_{l+j}\|^\alpha} \mathbb{1}_{A_{T_k}^+ \setminus A_{(\min T_k)(k)}} \left(\frac{X_{l+h+i}}{\|X_{l+h}\|} \right) \right)^2 \right] \\ & \leq |K_T| \sum_{k \in K_T} \frac{1}{r_n v_n} E \left[\left(\sum_{l=1}^{r_n} \mathbb{1}_{\{\|X_l\| > u_n\}} \sum_{h \in H_n} \frac{\|X_{l+h}\|^\alpha}{\sum_{j=-s_n}^{s_n} \|X_{l+j}\|^\alpha} \mathbb{1}_{A_{T_k}^+} \left(\frac{X_{l+h+i}}{\|X_{l+h}\|} \right) \right)^2 \right] \\ & \quad - E \left[\left(\sum_{l=1}^{r_n} \mathbb{1}_{\{\|X_l\| > u_n\}} \sum_{h \in H_n} \frac{\|X_{l+h}\|^\alpha}{\sum_{j=-s_n}^{s_n} \|X_{l+j}\|^\alpha} \mathbb{1}_{A_{(\min T_k)(k)}} \left(\frac{X_{l+h+i}}{\|X_{l+h}\|} \right) \right)^2 \right] \\ & = |K_T| \sum_{k \in K_T} (F_k(\sup T_k) - F_k(\min T_k)) \leq |K_T| \sum_{k \in K_T} \varepsilon^2 q^{-2} \leq \varepsilon^2. \end{aligned}$$

Hence, $\mathcal{A}_T^{\varepsilon,n}$ is indeed an ε -bracket wr.t. the L_2^n -metric considered in condition (D2) of D&N. Since there are $\prod_{k=1}^q (J_k^* + 1)$ brackets, the bracketing number $N_{[]}(\varepsilon, \mathcal{A}, L_2^n)$

can be bounded by $\prod_{k=1}^q (J_k^* + 1) \leq (N_n q^2 / \varepsilon^2 + 3)^q$. Now recall from (A.6) that N_n is stochastically bounded. Thus, for sufficient large n and a suitable constant $c_1 > 0$, it follows

$$\int_0^{\tau_n} \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{A}, L_2^n)} d\varepsilon \leq \int_0^{\tau_n} \sqrt{c_1 + \log N_n - 2q \log \varepsilon} d\varepsilon$$

which tends to 0 in probability as $n \rightarrow \infty$ for all sequences $\tau_n \rightarrow 0$. Hence, (D2) is satisfied.

To sum up, we have shown that all conditions of Theorem 2.3 (i) of D&N are fulfilled, which yields the assertion. $\square$

PROOF OF REMARK 2.1. The remark follows by completely analogous arguments as Theorem 2.2, if we apply part (ii) of Theorem 2.3 of D&N (instead of part (i)). Then condition (D2) of D&N is not needed any more, but instead we have to verify (D3) of D&N. Note that, by (A) (ii), the sets in $\mathcal{A}$ are linearly ordered in the case $q = 1$. The same then holds true for the functions in $\mathcal{G}$, which is thus a $VC(2)$ -class. This in turn implies (D3) of D&N (cf. Remark 2.11 of Drees & Rootzén (2010)).

Note that the conditions (A) (vi) and (vii) were only used for the proof of (D2) and are hence not needed in the case $q = 1$. $\square$

A.3. Proof of Theorem 2.3. Define the function ϕ on l_α by $\phi((x_h)_{h \in \mathbb{Z}}) := \log^+ \|x_0\|$ and let $b_n(\phi) := (nv_n/p_n)^{1/2}$.

The weak convergence of the first summand in the representation

$$\begin{aligned} & \sqrt{nv_n}(\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}} \\ (A.10) \quad &= \sqrt{nv_n}(\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}} + \sqrt{nv_n}(\hat{p}_{n,A} - \hat{p}_{n,A})_{A \in \mathcal{A}} \end{aligned}$$

has already been shown in Theorem 2.2.

To deal with the second term, we first analyze the asymptotic behavior of the estimator $\hat{\alpha}_n = T_{n, \mathbb{R}^d} / \tilde{T}_{n, \phi}$ with $\tilde{T}_{n, \phi} := \sum_{t=1}^n \phi(W_{n,t}) = \sum_{t=1}^n \log^+(\|X_t\|/u_n)$. Let $Z_{n, \phi} := (nv_n)^{-1/2}(\tilde{T}_{n, \phi} - E[\tilde{T}_{n, \phi}])$. With similar arguments as in the proof of Proposition A.5, we will show weak convergence of $(Z_n^{(T)}, Z_{n, \phi})$ to the centered Gaussian process (Z, Z_ϕ) defined in Theorem 2.3. Note that to this end it suffices to prove convergence of the fidis, because asymptotic tightness follows immediately from Proposition A.5. Since ϕ is unbounded, here we apply a modification of Theorem 2.4 of Drees & Neblung (2021) (D&N) (see Theorem 7.1 in the Supplement) instead of Theorem 2.3. To ease the formulas, in what follows we use the notation

$$X_{n,t} := \frac{X_t}{u_n}.$$

Check that $E((\log^+ \|X_{n,0}\|)^\eta \mid \|X_0\| > u_n) = O(1)$ for $\eta \geq 0$ due to regular variation of $\|X_0\|$. By condition (BC') and stationarity, we obtain

$$\begin{aligned} & \frac{1}{r_n v_n} E\left(\left(\sum_{t=1}^{r_n} \log^+ \|X_{n,t}\|\right)^2\right) = \sum_{s=1}^{r_n} \sum_{t=1}^{r_n} \frac{1}{r_n v_n} E(\log^+ \|X_{n,s}\| \cdot \log^+ \|X_{n,t}\|) \\ & \leq 2 \sum_{k=0}^{r_n} E(\log^+ \|X_{n,0}\| \cdot \log^+ \|X_{n,k}\| \mid \|X_0\| > u_n) = 2 \sum_{k=1}^{r_n} e'_n(k) + O(1) = O(1), \end{aligned}$$

that is, condition (7.2) of Theorem 7.1 in the Supplement is satisfied. Markov's inequality yields

$$P\left\{\sum_{t=1}^{r_n} \log^+ \|X_{n,t}\| > \epsilon \sqrt{nv_n}\right\} \leq \frac{1}{\epsilon^2 nv_n} E\left(\left(\sum_{t=1}^{r_n} \log^+ \|X_{n,t}\|\right)^2\right) = O\left(\frac{r_n}{n}\right)$$

for all $\epsilon > 0$. Applying Hölder's inequality, stationarity and condition (M) (i), we conclude for $\delta > 0$

$$\begin{aligned} & E\left(\left(\sum_{t=1}^{r_n} \log^+ \|X_{n,t}\|\right)^2 \mathbf{1}_{\{\sum_{j=1}^{r_n} \log^+ \|X_{n,j}\| > \epsilon \sqrt{nv_n}\}}\right) \\ & \leq \sum_{s=1}^{r_n} \sum_{t=1}^{r_n} E((\log^+ \|X_{n,s}\| \cdot \log^+ \|X_{n,t}\|)^{1+\delta})^{1/(1+\delta)} P\left\{\sum_{j=1}^{r_n} \log^+ \|X_{n,j}\| > \epsilon \sqrt{nv_n}\right\}^{\delta/(1+\delta)} \\ & \leq 2r_n v_n^{1/(1+\delta)} \sum_{k=0}^{r_n} E((\log^+ \|X_{n,0}\| \cdot \log^+ \|X_{n,k}\|)^{1+\delta} | \|X_0\| > u_n)^{1/(1+\delta)} \cdot O\left(\left(\frac{r_n}{n}\right)^{\delta/(1+\delta)}\right) \\ & = O\left(r_n v_n^{1/(1+\delta)} \left(\frac{r_n}{n}\right)^{\delta/(1+\delta)}\right) = O\left(r_n v_n \left(\frac{r_n}{nv_n}\right)^{\delta/(1+\delta)}\right) = o(r_n v_n). \end{aligned}$$

Thus condition (L2) of D&N is satisfied for ϕ , which implies condition (L) of D&N (cf. Supplement of that paper). All remaining conditions of Theorem 7.1 (i) in the Supplement have been verified in the proof of Proposition A.1, except for the convergence of the standardized variance of $Z_{n,\phi}$ and the covariance of $Z_{n,\phi}$ and $Z_n^{(T)}(A)$ for all $A \in \mathcal{A}$, which is proved in Lemma A.12. Hence we may conclude that

$$((Z_n^{(T)}(A))_{A \in \mathcal{A}}, Z_{n,\phi}) \xrightarrow{w} ((Z(A))_{A \in \mathcal{A}}, Z_\phi).$$

By similar arguments as in the proof of Theorem 2.1 using the bias conditions (2.4) and (H), we conclude

$$(A.11) \quad \sqrt{nv_n}((\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}}, \hat{\alpha}_n - \alpha) \xrightarrow{w} ((Z^{(p)}(A))_{A \in \mathcal{A}}, Z_\alpha),$$

with $Z_\alpha := \alpha Z(\mathbb{R}^d) - \alpha^2 Z_\phi$ (cf. Drees et al. (2015), Lemma 4.4).

For $A \in \mathcal{A}$, define a function $p_{n,A} : (0, \infty) \rightarrow \mathbb{R}$ by

$$\begin{aligned} p_{n,A}(a) &:= \frac{1}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \frac{\|X_{n,t+h}\|^a}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^a} \\ &\quad \times \left(\mathbf{1}_{\{h \in H_n\}} \mathbf{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \mathbf{1}_{\{h \in H_n^c\}} \mathbf{1}_A(0) \right) \end{aligned}$$

so that $\hat{p}_{n,A} = p_{n,A}(\alpha)$ and $\hat{\hat{p}}_{n,A} = p_{n,A}(\hat{\alpha}_n)$. A Taylor expansion yields

$$\begin{aligned} & \hat{\hat{p}}_{n,A} - \hat{p}_{n,A} \\ &= \frac{\hat{\alpha}_n - \alpha}{\sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \mathbf{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \left(\frac{\log(\|X_{n,t+h}\|) \|X_{n,t+h}\|^\alpha}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha} \right. \\ &\quad \left. - \frac{\|X_{n,t+h}\|^\alpha \sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|) \|X_{n,t+k}\|^\alpha}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^\alpha)^2} \right) \\ &\quad \times \left(\mathbf{1}_{\{h \in H_n\}} \mathbf{1}_A\left(\frac{X_{t+h+i}}{\|X_{t+h}\|}\right) + \mathbf{1}_{\{h \in H_n^c\}} \mathbf{1}_A(0) \right) \end{aligned}$$

$$\begin{aligned}
& + \frac{(\hat{\alpha}_n - \alpha)^2}{2 \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}} \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \left(\frac{\log^2(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right. \\
& \quad - 2 \frac{\log(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}} \sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|) \|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
& \quad - \frac{\|X_{n,t+h}\|^{\bar{\alpha}} \sum_{k=-s_n}^{s_n} \log^2(\|X_{n,t+k}\|) \|X_{n,t+k}\|^{\bar{\alpha}}}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^2} \\
& \quad \left. + 2 \frac{\|X_{n,t+h}\|^{\bar{\alpha}} (\sum_{k=-s_n}^{s_n} \log(\|X_{n,t+k}\|) \|X_{n,t+k}\|^{\bar{\alpha}})^2}{(\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}})^3} \right) \\
& \quad \times \left(\mathbb{1}_{\{h \in H_n\}} \mathbb{1}_A \left(\frac{X_{t+h+i}}{\|X_{t+h}\|} \right) + \mathbb{1}_{\{h \in H_n^c\}} \mathbb{1}_A(0) \right) \\
& =: I(A) + II(A),
\end{aligned}
\tag{A.12}$$

with $\bar{\alpha} = \bar{\alpha}_n = \lambda_n \alpha + (1 - \lambda_n) \hat{\alpha}_n$ for some (random) $\lambda_n \in (0, 1)$. Define functions $f_A^{(m)}$ on $l_\alpha \rightarrow \mathbb{R}$ by

$$\begin{aligned}
f_A^{(m)}((w_h)_{h \in \mathbb{Z}}) &:= \mathbb{1}_{\{\|w_0\| > 1\}} \sum_{|h| \leq m} \left(\frac{\log(\|w_h\|) \|w_h\|^\alpha}{\sum_{|k| \leq m} \|w_k\|^\alpha} \right. \\
& \quad \left. - \frac{\|w_h\|^\alpha \sum_{|k| \leq m} \log(\|w_k\|) \|w_k\|^\alpha}{(\sum_{|k| \leq m} \|w_k\|^\alpha)^2} \right) \mathbb{1}_A \left(\frac{w_{h+i}}{\|w_h\|} \right)
\end{aligned}
\tag{A.13}$$

and $f_A := f_A^{(\infty)}$ with the convention $\sum_{|h| \leq \infty} = \sum_{h \in \mathbb{Z}}$. In the Supplement it is shown that the constant d_A defined in Theorem 2.3 equals $E[f_A(Y)]$ (with $(Y = (Y_h)_{h \in \mathbb{Z}})$). Let

$$d_n(A) := \frac{1}{nv_n} \sum_{t=1}^n f_A(W_{n,t}).$$

A combination of Lemma A.8 and Lemma A.9 (given below) shows that $\sup_{A \in \mathcal{A}} |d_n(A) - d_A| = o_P(1)$. Because $nv_n / \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \rightarrow 1$ in probability (which is immediate from Proposition A.1), it follows from convergence (A.11) that

$$\sqrt{nv_n} I(A) = \sqrt{nv_n} (\hat{\alpha}_n - \alpha) \frac{nv_n}{\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}} d_n(A) \xrightarrow{w} d_A Z_\alpha$$

uniformly for all $A \in \mathcal{A}$. By Lemma A.11, the term $II(A)$ is asymptotically negligible. Hence, we may conclude from (A.12) that

$$\sqrt{nv_n} (\hat{p}_{n,A} - \hat{p}_{n,A})_{A \in \mathcal{A}} \rightarrow (d_A Z_\alpha)_{A \in \mathcal{A}}.$$

Indeed, the arguments show that this convergence holds jointly with the first component in (A.11). Hence, by (A.10) and the continuous mapping theorem

$$\begin{aligned}
& \sqrt{nv_n} (\hat{p}_{n,A} - P\{\Theta_i \in A\})_{A \in \mathcal{A}} \xrightarrow{w} (Z^{(p)}(A) + d_A Z_\alpha)_{A \in \mathcal{A}} \\
& = (Z(A) - (P\{\Theta_i \in A\} - \alpha d_A) Z(\mathbb{R}^d) - \alpha^2 d_A Z_\phi)_{A \in \mathcal{A}}. \quad \square
\end{aligned}$$

The Lemmas A.6–A.9 establish uniform convergence of $d_n(A) - d_A$ to 0 in probability.

LEMMA A.6. Under Conditions (RV), (S) and (C Θ), we have

$$\lim_{n \rightarrow \infty} \frac{1}{v_n} E[f_A^{(m)}(W_{n,0})] = E[f_A^{(m)}(Y)] =: d_A^{(m)} \in \mathbb{R}$$

for all $A \in \mathcal{A}$ and all $m \in \mathbb{N}$, with $f_A^{(m)}$ defined by (A.13).

PROOF. Since $P\{Y_{h+i}/\|Y_h\| \in \partial A, \|Y_h\| > 0\} = 0$ for all $h \in \mathbb{Z}$ and all $A \in \mathcal{A}$ by Lemma A.3, the function $f_A^{(m)}$ is P^Y -a.s. continuous as finite sum of continuous functions. Because $\mathcal{L}((X_h/u_n)_{|h| \leq m} \mid \|X_0\| > u_n) \xrightarrow{w} \mathcal{L}((Y_h)_{|h| \leq m})$ by the definition of the tail process, the assertion $v_n^{-1} E[f_A^{(m)}(W_{n,0})] = E(f_A^{(m)}(W_{n,0}) \mid \|X_0\| > u_n) \rightarrow d_A \in \mathbb{R}$ follows, provided the random variables $v_n^{-1} f_A^{(m)}(W_{n,0}), n \in \mathbb{N}$, are uniformly integrable. This, in turn, is implied by the uniform moment bound

$$\sup_{n \in \mathbb{N}} E(|f_A^{(m)}(W_{n,0})|^{1+\eta} \mid \|X_0\| > u_n) < \infty$$

for some $\eta > 0$. Due to the Minkowski inequality, we can verify this bound separately for both summands in the definition of $f_A^{(m)}$, whose moduli are maximal for $A = \mathbb{R}^d$. Since $s_n > m$ for n large enough, it thus suffices to show that

$$(A.14) \quad E\left(\left|\frac{\sum_{h=-m}^m \log(\|X_{n,h}\|) \|X_{n,h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,k}\|^\alpha}\right|^{1+\eta} \mid \|X_0\| > u_n\right)$$

is bounded. Using Jensen's inequality and the fact that the denominator is at least 1 if $\|X_0\| > u_n$, this expectation can be bounded by

$$\begin{aligned} & E\left(\frac{\sum_{h=-m}^m |\log \|X_{n,h}\||^{1+\eta} \|X_{n,h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n\right) \\ & \leq \sum_{h=-m}^m \left[E((\log^+ \|X_{n,h}\|)^{1+\eta} \mid \|X_0\| > u_n) + E((\log^- \|X_{n,h}\|)^{1+\eta} \|X_{n,h}\|^\alpha \mid \|X_0\| > u_n) \right]. \end{aligned}$$

The first expectation is bounded because the time series is regularly varying and $(\log^+ x)^{1+\eta} = o(x^{\alpha-\varepsilon})$ for some $\varepsilon > 0$ (cf. Kulik & Soulier (2020), Section 2.3.3), while the second is bounded, because $(\log^- x)^{1+\eta} x^\alpha$ is a bounded function. Since m is fixed, (A.14) is bounded, which concludes the proof. $\square$

LEMMA A.7. If the conditions (RV), (S), (C Θ) are fulfilled, then

$$d_n^{(m)}(A) := \frac{1}{nv_n} \sum_{t=1}^n f_A^{(m)}(W_{n,t}) \rightarrow d_A^{(m)}$$

in probability for all $A \in \mathcal{A}$ and all $m \in \mathbb{N}$.

PROOF. Rewrite $d_n^{(m)}(A)$ as

$$(A.15) \quad d_n^{(m)}(A) = \frac{r_n}{n} \sum_{l=1}^2 \sum_{j=1}^{m_{n,l}} F_{A,n}^{(m)}(l, j) + \frac{1}{nv_n} \sum_{t=m_n r_n + 1}^n f_A^{(m)}(W_{n,t}),$$

with $m_n := \lfloor n/r_n \rfloor$, $m_{n,l} := \max\{j \in \mathbb{N}_0 \mid 2(j-1) + l \leq n/r_n\} \sim n/(2r_n)$ and

$$F_{n,A}^{(m)}(l, j) := \frac{1}{r_n v_n} \sum_{t=1}^{r_n} f_A^{(m)}(W_{n, (2(j-1)+l-1)r_n + t}).$$

We first prove

$$(A.16) \quad E[(F_{n,A}^{(m)}(1,1))^2] = o(n/r_n).$$

Check that

$$\begin{aligned} & E \left[\left(\sum_{t=1}^{r_n} |f_A^{(m)}(W_{n,t})| \right)^2 \right] \\ & \leq 4E \left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-m}^m \frac{|\log \|X_{n,t+h}\|| \|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,t+k}\|^\alpha} \right)^2 \right] \\ & \leq 8 \left(E \left[\left(\sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-m}^m \frac{|\log \|X_{n,t+h}\|| \|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,t+k}\|^\alpha} \mathbb{1}_{\{\|X_{t+h}\| < u_n\}} \right)^2 \right] \right. \\ & \quad \left. + r_n^2 E \left[\left(\frac{1}{r_n} \sum_{t=1}^{r_n} \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-m}^m \frac{\log(\|X_{n,t+h}\|) \|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,t+k}\|^\alpha} \mathbb{1}_{\{\|X_{t+h}\| \geq u_n\}} \right)^2 \right] \right) \\ (A.17) \quad & =: 8(E_1 + E_2). \end{aligned}$$

According to (A.6), E_1 is at most of the order $r_n v_n = o(nr_n v_n^2)$, since $\log^-(x)x^\alpha$ is bounded and the denominator is at least 1. Using Jensen's inequality twice and stationarity, we may conclude for the second term

$$\begin{aligned} E_2 & \leq r_n^2 E \left[\frac{1}{r_n} \sum_{t=1}^{r_n} \left(\mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-m}^m \frac{\log^+(\|X_{n,t+h}\|) \|X_{n,t+h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,t+k}\|^\alpha} \right)^2 \right] \\ & = r_n^2 E \left[\mathbb{1}_{\{\|X_0\| > u_n\}} \left(\sum_{h=-m}^m \frac{\log^+(\|X_{n,h}\|) \|X_{n,h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,k}\|^\alpha} \right)^2 \right] \\ & \leq r_n^2 v_n E \left(\sum_{h=-m}^m \frac{(\log^+(\|X_{n,h}\|))^2 \|X_{n,h}\|^\alpha}{\sum_{k=-m}^m \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\ & \leq r_n^2 v_n \sum_{h=-m}^m E \left((\log^+ \|X_{n,h}\|)^2 \mid \|X_0\| > u_n \right) = O(r_n^2 v_n) = o(nr_n v_n^2), \end{aligned}$$

where the last two steps follow from regular variation as before (cf. Kulik & Soulier (2020), Section 2.3.3) and $r_n = o(\sqrt{nv_n})$.

Combine this result with the bound on E_1 and (A.17) to conclude (A.16).

These calculations also show that the last summand in (A.15) vanishes in L_2 -norm and thus in probability:

$$E \left[\left(\frac{1}{nv_n} \sum_{t=m_n r_n + 1}^n f_A^{(m)}(W_{n,t}) \right)^2 \right] \leq \frac{1}{(nv_n)^2} E \left[\left(\sum_{t=1}^{r_n} |f_A^{(m)}(W_{n,t})| \right)^2 \right] = o(r_n/n) = o(1).$$

Denote independent copies of $F_{n,A}^{(m)}(l, j)$ by $F_{n,A}^{(m)*}(l, j)$. Since, for fixed $l \in \{1, 2\}$ and different $j \in \{1, \dots, m_{n,l}\}$, the random variables $F_{n,A}^{(m)}(l, j)$ are measurable functions of random variables X_t which are separated in time by at least $r_n - 2s_n - 1$ observations, Eberlein's inequality (Eberlein (1984)) combined with condition (S) shows that the total variation distance between the joint distribution of the $F_{n,A}^{(m)}(l, j)$ and the joint

distribution of the $F_{n,A}^{(m)*}(l, j)$ tends to 0:

$$\|P^{(F_{n,A}^{(m)*}(l, j))_{1 \leq j \leq m_{n,l}}} - P^{(F_{n,A}^{(m)}(l, j))_{1 \leq j \leq m_{n,l}}}\|_{TV} \leq m_{n,l} \beta_{r_n - 2s_n - 1} = O(n \beta_{l_n} / r_n) \rightarrow 0$$

for $l \in \{1, 2\}$. Thus, in view of (A.15), $d_n^{(m)}(A)$ converges to $d_A^{(m)}$ in probability if

$$\frac{r_n}{n} \sum_{j=1}^{m_{n,l}} F_{n,A}^{(m)*}(l, j) \rightarrow \frac{d_A^{(m)}}{2}$$

in probability for $l \in \{1, 2\}$. This, however, follows by Chebyshev's inequality from (A.16) and Lemma A.6, which shows that

$$\frac{r_n}{n} \sum_{j=1}^{m_{n,l}} E[F_{n,A}^{(m)*}(l, j)] = \frac{r_n m_{n,l}}{n} \frac{E[f_A^{(m)}(W_{n,0})]}{v_n} \rightarrow \frac{d_A^{(m)}}{2}. \quad \square$$

LEMMA A.8. Suppose (RV), (S), (C Θ), and (A) are satisfied. Then $\sup_{A \in \mathcal{A}} |d_A^{(m)}| < \infty$ and $\sup_{A \in \mathcal{A}} |d_n^{(m)}(A) - d_A^{(m)}| = o_P(1)$ for all $m \in \mathbb{N}$.

PROOF. Fix $m \in \mathbb{N}$ and let

$$f_A^{(I, \pm)}(w) = \mathbb{1}_{\{\|w_0\| > 1\}} \sum_{|h| \leq m} \frac{\log^\pm(\|w_h\|) \|w_h\|^\alpha}{\sum_{|k| \leq m} \|w_k\|^\alpha} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right),$$

$$f_A^{(II, \pm)}(w) = \mathbb{1}_{\{\|w_0\| > 1\}} \sum_{|h| \leq m} \frac{\|w_h\|^\alpha \sum_{|k| \leq m} \log^\pm(\|w_k\|) \|w_k\|^\alpha}{(\sum_{|k| \leq m} \|w_k\|^\alpha)^2} \mathbb{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right).$$

(To improve readability, here we suppress the dependence on m in the notation.) In the proof of Lemma A.6, it has been shown that the families of random variables $v_n^{-1} f_A^{(\sharp, \pm)}(W_{n,0})$, $n \in \mathbb{N}$, are uniformly integrable for $\sharp \in \{I, II\}$ (cf. (A.14)). Therefore, by the same arguments as used in the proof of Lemma A.7, it follows that

$$d_n^{(\sharp, \pm)}(A) := \frac{1}{n v_n} \sum_{t=1}^n f_A^{(\sharp, \pm)}(W_{n,t}) \xrightarrow{P} E[f_A^{(\sharp, \pm)}((Y_t)_{t \in \mathbb{Z}})] =: d_A^{(\sharp, \pm)} < \infty$$

for all $A \in \mathcal{A}$ and $\sharp \in \{I, II\}$. Indeed, this even holds for all $A \in \mathcal{A} \cup \tilde{\mathcal{A}} = \{A_t, A_t^- \mid t \in [0, 1 + \iota]^q\}$ as assumption (A) (iv) ensures the necessary continuity property for sets A_t^- . Since $f_A^{(m)} = f_A^{(I, +)} - f_A^{(I, -)} - f_A^{(II, +)} + f_A^{(II, -)}$ and $d_A^{(\sharp, \pm)} \leq d_{\mathbb{R}^d}^{(\sharp, \pm)} < \infty$, it suffices to prove $\sup_{A \in \mathcal{A}} |d_n^{(\sharp, \pm)}(A) - d_A^{(\sharp, \pm)}| = o_P(1)$ for all $\sharp \in \{I, II\}$.

Fix some $\varepsilon > 0$ and, for the time being, some $k \in \{1, \dots, q\}$. By assumptions (A) (ii) and (v), $[0, 1] \ni t \mapsto E[f_{A_{t(k)}}^{(\sharp, \pm)}(Y)]$ (with $Y = (Y_h)_{h \in \mathbb{Z}}$) is non-decreasing and right-continuous. Thus, by the same arguments as used in the proof of Proposition A.5 for the function H_k , there exist $J_k \in \mathbb{N}$ and $0 =: t_{k,0} < t_{k,1} < \dots < t_{k,J_k} := 1$ such that

$$E\left[f_{\cup_{s < t_{k,j}} A_{s(k)}}^{(\sharp, \pm)}(Y) - f_{A_{t_{k,j-1}(k)}}^{(\sharp, \pm)}(Y)\right] < \varepsilon$$

for all $1 \leq j \leq J_k$. Let $\mathcal{T}_k := \{[t_{k,j-1}, t_{k,j}] \mid 1 \leq j \leq J_k\} \cup \{\{1\}\}$.

We now define brackets for $(f_A^{(\sharp, \pm)})_{A \in \mathcal{A}}$ by

$$[T_1, \dots, T_q] := \{f_{A_{(s_1, \dots, s_l)}}^{(\sharp, \pm)} \mid s_k \in T_k, \forall 1 \leq k \leq q\}$$

for all $T_k \in \mathcal{T}_k$, $1 \leq k \leq q$. Recall definition (A.7) of $\bar{A}_T$ and $\underline{A}_T$. By construction, for all $s, t \in T := \times_{k=1}^q T_k$,

$$\begin{aligned} E[|f_{A_s}^{(\sharp, \pm)}(Y) - f_{A_t}^{(\sharp, \pm)}(Y)|] &\leq E[f_{\bar{A}_T}^{(\sharp, \pm)}(Y) - f_{\underline{A}_T}^{(\sharp, \pm)}(Y)] \\ &\leq \sum_{k=1}^q E[f_{\cup_{r \in T_k} A_{r(k)}}^{(\sharp, \pm)}(Y) - f_{A_{(\min T_k)(k)}}^{(\sharp, \pm)}(Y)] < q\varepsilon. \end{aligned}$$

This proves that, for arbitrary $\varepsilon > 0$, $(f_A^{(\sharp, \pm)})_{A \in \mathcal{A}}$ is covered by finitely many $q\varepsilon$ -brackets w.r.t. the $L_1(P^Y)$ -norm. Since $d_n^{(\sharp, \pm)}(A)$ converges pointwise for all $A \in \mathcal{A} \cup \tilde{\mathcal{A}}$, the proof of Theorem 2.4.1 of [Van der Vaart & Wellner \(1996\)](#) shows that $\sup_{A \in \mathcal{A}} |d_n^{(\sharp, \pm)}(A) - d_A^{(\sharp, \pm)}| = o_P(1)$ for all $\sharp \in \{I, II\}$. $\square$

LEMMA A.9. (i) Suppose the Conditions (RV), (S), (BC), (BC'), (TC), and (M) (ii), (iii) are satisfied. Then

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{A \in \mathcal{A}} E|d_n^{(m)}(A) - d_n(A)| = 0.$$

(ii) Under Conditions (RV), (S) and (M) (ii)

$$\lim_{m \rightarrow \infty} \sup_{A \in \mathcal{A}} |d_A^{(m)} - d_A| = 0.$$

PROOF. By stationarity one has for sufficiently large n (such that $\{-m, \dots, m\} \subset H_n$)

$$\begin{aligned} E(|d_n^{(m)}(A) - d_n(A)|) &\leq E(|f_A^{(m)}(W_{n,0}) - f_A(W_{n,0})| \mid \|X_0\| > u_n) \\ &\leq E\left(\left| \frac{\sum_{|h| \leq m} \log(\|X_{n,h}\|) \|X_{n,h}\|^\alpha \mathbf{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right)}{\sum_{|k| \leq m} \|X_{n,k}\|^\alpha} \right. \right. \\ &\quad \left. \left. - \frac{\sum_{|h| \leq s_n} \log(\|X_{n,h}\|) \|X_{n,h}\|^\alpha \left(\mathbf{1}_{\{h \in H_n\}} \mathbf{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) + \mathbf{1}_{\{h \in H_n^c\}} \mathbf{1}_A(0) \right)}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \right| \mid \|X_0\| > u_n\right) \\ &\quad + E\left(\left| \frac{\sum_{|h| \leq m} \|X_{n,h}\|^\alpha \mathbf{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) \sum_{|k| \leq m} \log(\|X_{n,k}\|) \|X_{n,k}\|^\alpha}{(\sum_{|k| \leq m} \|X_{n,k}\|^\alpha)^2} \right. \right. \\ &\quad \left. \left. - \frac{\sum_{|h| \leq s_n} \|X_{n,h}\|^\alpha \left(\mathbf{1}_{\{h \in H_n\}} \mathbf{1}_A\left(\frac{X_{h+i}}{\|X_h\|}\right) + \mathbf{1}_{\{h \in H_n^c\}} \mathbf{1}_A(0) \right) \sum_{|k| \leq s_n} \log(\|X_{n,k}\|) \|X_{n,k}\|^\alpha}{(\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha)^2} \right| \right. \\ &\quad \left. \left. \mid \|X_0\| > u_n \right) \right) \end{aligned}$$

$$=: T_1 + T_2.$$

In the following calculations we will use identities of the type

$$(A.18) \quad \sum_{|h| \leq s_n} a_h \sum_{|k| \leq m} b_k - \sum_{|h| \leq m} a_h \sum_{|k| \leq s_n} b_k = \sum_{m < |h| \leq s_n} a_h \sum_{|k| \leq m} b_k - \sum_{|h| \leq m} a_h \sum_{m < |k| \leq s_n} b_k$$

for $a_h, b_k \in \mathbb{R}$. Reducing the fractions to a common denominator, using (A.18) and the triangle inequality and bounding all indicators by 1, we obtain

$$T_1 \leq E\left(\left| \frac{\sum_{|h| \leq m} |\log \|X_{n,h}\|| \|X_{n,h}\|^\alpha \sum_{m < |k| \leq s_n} \|X_{n,k}\|^\alpha}{\sum_{|k| \leq m} \|X_{n,k}\|^\alpha \sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \right| \mid \|X_0\| > u_n\right)$$

$$\begin{aligned}
& + E \left(\frac{\sum_{m < |h| \leq s_n} |\log \|X_{n,h}\|| \|X_{n,h}\|^\alpha}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& =: T_{1,1} + T_{1,2}.
\end{aligned}$$

Applying the Hölder inequality for expectations and for sums, for $\delta > 0$, we can bound $T_{1,1}$ by

$$\begin{aligned}
T_{1,1} & \leq \left(E \left(\frac{\sum_{|h| \leq m} |\log \|X_{n,h}\||^{1+\delta} \|X_{n,h}\|^\alpha}{\sum_{|k| \leq m} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \right)^{1/(1+\delta)} \\
& \quad \times \left(E \left(\left(\frac{\sum_{m < |k| \leq s_n} \|X_{n,k}\|^\alpha}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \right)^{(1+\delta)/\delta} \mid \|X_0\| > u_n \right) \right)^{\delta/(1+\delta)}.
\end{aligned}$$

Because of uniform integrability (cf. (A.14)), the first expectation converges to the expectation that is bounded in (2.5). The second expectation is bounded by

$$\begin{aligned}
& E \left(\frac{\sum_{m < |k| \leq s_n} \|X_{n,k}\|^\alpha}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& \leq E \left(\frac{\sum_{m < |k| \leq s_n} \|X_{n,k}\|^\alpha \mathbf{1}_{\{\|X_k\| \leq cu_n\}}}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& \quad + \sum_{m < |k| \leq s_n} P(\|X_k\| > cu_n \mid \|X_0\| > u_n) \frac{P(\|X_0\| > cu_n)}{P(\|X_0\| > u_n)} \\
& \leq E \left(\frac{\sum_{m < |k| \leq s_n} \|X_{n,k}\|^\alpha \mathbf{1}_{\{\|X_k\| \leq cu_n\}}}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) + 4c^{-\alpha} \sum_{h=m}^{s_n} e_{n,c}(h),
\end{aligned}$$

for sufficiently large n and the constant $c \in (0, 1)$ from Condition (TC), and hence $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} T_{1,1} = 0$ by condition (M) (ii), (BC) and (TC). Condition (BC') implies

$$\begin{aligned}
& E \left(\frac{\sum_{m < |h| \leq s_n} \log^+(\|X_{n,h}\|) \|X_{n,h}\|^\alpha}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& \leq \sum_{m < |h| \leq s_n} E(\log^+ \|X_{n,h}\| \mid \|X_0\| > u_n) \leq 2 \sum_{m < h \leq s_n} e'_n(h),
\end{aligned}$$

and thus $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} T_{1,2} = 0$ by condition (BC') and (M) (iii). Similarly, one can show that

$$\begin{aligned}
T_2 & \leq 2E \left(\frac{\sum_{|k| \leq m} |\log \|X_{n,k}\|| \|X_{n,k}\|^\alpha \sum_{m < |h| \leq s_n} \|X_{n,h}\|^\alpha}{\sum_{|k| \leq m} \|X_{n,k}\|^\alpha \sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& \quad + E \left(\frac{\sum_{m < |k| \leq s_n} |\log \|X_{n,k}\|| \|X_{n,k}\|^\alpha}{\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha} \mid \|X_0\| > u_n \right) \\
& \quad + E \left(\frac{\sum_{m < |h| \leq s_n} \|X_{n,h}\|^\alpha \sum_{|k| \leq m} |\log \|X_{n,k}\|| \|X_{n,k}\|^\alpha}{(\sum_{|k| \leq s_n} \|X_{n,k}\|^\alpha)^2} \mid \|X_0\| > u_n \right) \\
& \leq 3T_{1,1} + T_{1,2},
\end{aligned}$$

so that $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} T_2 = 0$. Since the bounds on T_1 and T_2 do not depend on A , all convergences hold uniformly in A , which proves assertion (i). Assertion (ii) follows by similar arguments, which are given in detail in the Supplement. $\square$

The next two lemmas establish the asymptotic negligibility of $II(A)$. The proof of the first lemma is given in the Supplement.

LEMMA A.10. For $m \in \mathbb{N}$ and $a = (a_1, \dots, a_m) \in [0, 1]^m$ let

$$M(a) := \frac{\sum_{k=1}^m a_k \log^2 a_k}{1 + \sum_{k=1}^m a_k}, \quad \tilde{M}(a) := \frac{\sum_{k=1}^m a_k |\log a_k|}{1 + \sum_{k=1}^m a_k}$$

with $0 \log^i(0) := 0$ for $i \in \{1, 2\}$. Then $\sup_{a \in [0, 1]^m} M(a) = O(\log^2 m)$ and $\sup_{a \in [0, 1]^m} \tilde{M}(a) = O(\log m)$ as $m \rightarrow \infty$.

LEMMA A.11. If the conditions (RV), (S), (BC), (TC), (CΘ), (BC'), (M) (i), (H) and $\log^4 n = o(nv_n)$ are fulfilled, then the term $II(A)$ defined in (A.12) is uniformly negligible:

$$\sup_{A \in \mathcal{A}} |II(A)| = o_P((nv_n)^{-1/2}).$$

PROOF. Because $(\hat{\alpha}_n - \alpha)^2 = O_P((nv_n)^{-1})$ by (A.11) and $nv_n/(\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}) \rightarrow 1$ in probability, due to the weak convergence in Proposition A.1 for $\mathbb{R}^d$, it suffices to show that

$$\begin{aligned} \frac{2 \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}}}{(\hat{\alpha}_n - \alpha)^2} |II(A)| &\leq 2 \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\ &\quad + 4 \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \left(\frac{\sum_{h=-s_n}^{s_n} |\log(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\ &=: 2T_1 + 4T_2 = o_P((nv_n)^{3/2}). \end{aligned}$$

For this purpose we distinguish whether $\|X_{t+h}\|$ exceeds u_n or not. Define the set $B_n := \{\max_{1 \leq t \leq n} \|X_{n,t}\| \leq n^{2/\alpha}\}$. The regular variation of $\|X_0\|$ implies, for all $\epsilon \in (0, 1/2)$ and sufficient large n ,

$$P(B_n^c) \leq nP\{\|X_{n,0}\| > n^{2/\alpha}\} \leq nP\{\|X_0\| > n^{2/\alpha}\} = o(n(n^{2/\alpha})^{-\alpha(1-\epsilon)}) = o(1).$$

Since on the set B_n

$$0 \leq \log(\|X_{n,t}\|) \mathbb{1}_{\{\|X_t\| \geq u_n\}} \leq \frac{2}{\alpha} \log n$$

for all $1 \leq t \leq n$, we may conclude that, with probability tending to 1,

$$\begin{aligned} &\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \frac{(\log^2(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\| > u_n\}})}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\ (A.19) \quad &\leq \frac{4}{\alpha^2} \log^2 n \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} = o((nv_n)^{3/2}), \end{aligned}$$

since $\log^2 n = o(\sqrt{nv_n})$ by assumption.

Next, observe that by the first assertion of Lemma A.10

$$\begin{aligned} &\sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \sum_{h=-s_n}^{s_n} \frac{\log^2(\|X_{n,t+h}\|) \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\| \leq u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \\ &\leq \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \frac{1}{\bar{\alpha}^2} \sum_{h=-s_n, h \neq 0}^{s_n} \frac{\log^2(\|X_{n,t+h}\|^{\bar{\alpha}}) \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{n,t+h}\| \leq 1\}}}{1 + \sum_{k=-s_n, k \neq 0}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \end{aligned}$$

$$(A.20) \quad = O(\log^2 s_n) \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} = o((nv_n)^{3/2}).$$

Combine (A.19) and (A.20) to see that T_1 is stochastically of smaller order than $(nv_n)^{3/2}$.

Similarly as in (A.19), on the set B_n we obtain

$$(A.21) \quad \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \left(\frac{\sum_{h=-s_n}^{s_n} |\log(\|X_{n,t+h}\|)| \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\| > u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\ \leq \frac{4}{\alpha^2} \log^2 n \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} = o((nv_n)^{3/2}).$$

Moreover, by the second assertion of Lemma A.10

$$(A.22) \quad \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} \left(\frac{\sum_{h=-s_n}^{s_n} |\log \|X_{n,t+h}\|| \|X_{n,t+h}\|^{\bar{\alpha}} \mathbb{1}_{\{\|X_{t+h}\| \leq u_n\}}}{\sum_{k=-s_n}^{s_n} \|X_{n,t+k}\|^{\bar{\alpha}}} \right)^2 \\ = O(\log^2 s_n) \sum_{t=1}^n \mathbb{1}_{\{\|X_t\| > u_n\}} = o((nv_n)^{3/2}).$$

A combination of both these results show that T_2 , which can be bounded by double the sum of the left-hand sides of (A.21) and (A.22), is stochastically of smaller order than $(nv_n)^{3/2}$. This concludes the proof. $\square$

In the last lemma, the proof of which is given in the Supplement, the limit behavior of the covariance between $(Z(A))_{A \in \mathcal{A}}$ and Z_ϕ and of the variance of Z_ϕ is analyzed.

LEMMA A.12. Suppose the conditions (RV), (S), (BC), (TC), (C Θ), and (BC') are satisfied. Then

(i)

$$\frac{1}{r_n v_n} \text{Var} \left(\sum_{t=1}^{r_n} \phi(W_{n,t}) \right) = \alpha^{-1} \sum_{k \in \mathbb{Z}} E \left[(1 \wedge \|\Theta_k\|^\alpha) (|\log(\|\Theta_k\|)| + 2\alpha^{-1}) \right],$$

(ii) For all $A \in \mathcal{A}$,

$$\frac{1}{r_n v_n} \text{Cov} \left(\sum_{j=1}^{r_n} \phi(W_{n,j}), \sum_{t=1}^{r_n} g_A(W_{n,t}) \right) \\ = \sum_{k \in \mathbb{Z}} E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbb{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) (1 \wedge \|\Theta_k\|^\alpha) (\log^+ \|\Theta_k\| + \alpha^{-1}) \right].$$

REFERENCES

- Alsmeyer, G. (2003), 'On the Harris recurrence of iterated random Lipschitz functions and related convergence rate results', *Journal of Theoretical Probability* **16**(1), 217–247.
- Basrak, B., Davis, R. A. & Mikosch, T. (2002), 'Regular variation of GARCH processes', *Stochastic Processes and Their Applications* **99**(1), 95–115.
- Basrak, B. & Segers, J. (2009), 'Regularly varying multivariate time series', *Stochastic Processes and Their Applications* **119**(4), 1055–1080.
- Buraczewski, D., Damek, E. & Mikosch, T. (2016), *Stochastic Models with Power-Law Tails*, Springer.
- Castillo, E. (2012), *Extreme Value Theory in Engineering*, Elsevier.
- Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer.

- Davis, R. A., Drees, H., Segers, J. & Warchoř, M. (2018), ‘Inference on the tail process with application to financial time series modeling’, *Journal of Econometrics* **205**(2), 508–525.
- Douc, R., Moulines, E., Priouret, P. & Soulier, P. (2018), *Markov Chains*, Springer.
- Drees, H. & Knezevic, M. (2020), ‘Peak-over-threshold estimators for spectral tail processes: random vs deterministic thresholds’, *Extremes* **23**(3), 465–491.
- Drees, H. & Neblung, S. (2021), ‘Asymptotics for sliding blocks estimators of rare events’, *Bernoulli* . to appear.
- Drees, H. & Rootzén, H. (2010), ‘Limit theorems for empirical processes of cluster functionals’, *The Annals of Statistics* **38**(4), 2145–2186.
- Drees, H., Segers, J. & Warchoř, M. (2015), ‘Statistics for tail processes of Markov chains’, *Extremes* **18**(3), 369–402.
- Eberlein, E. (1984), ‘Weak convergence of partial sums of absolutely regular sequences’, *Statistics & Probability Letters* **2**(5), 291–293.
- Ehlert, A., Fiebig, U.-R., Janßen, A. & Schlather, M. (2015), ‘Joint extremal behavior of hidden and observable time series with applications to GARCH processes’, *Extremes* **18**(1), 109–140.
- Embrechts, P., Klüppelberg, C. & Mikosch, T. (2013), *Modelling Extremal Events for Insurance and Finance*, Vol. 33, Springer Science & Business Media.
- Fils-Villetard, A., Guillou, A. & Segers, J. (2008), ‘Projection estimators of pickands dependence functions’, *Canadian Journal of Statistics* **36**(3), 369–382.
- Finkenstädt, B. & Rootzén, H. (2003), *Extreme Values in Finance, Telecommunications, and the Environment*, CRC Press.
- Goldie, C. M. (1991), ‘Implicit renewal theory and tails of solutions of random equations’, *The Annals of Applied Probability* **1**(1), 126–166.
- Janßen, A. (2019), ‘Spectral tail processes and max-stable approximations of multivariate regularly varying time series’, *Stochastic Processes and Their Applications* **129**(6), 1993–2009.
- Janssen, A. & Segers, J. (2014), ‘Markov tail chains’, *Journal of Applied Probability* **51**(4), 1133–1153.
- Kulik, R. & Soulier, P. (2020), *Heavy-Tailed Time Series*, Springer.
- Kulik, R., Soulier, P. & Wintenberger, O. (2019), ‘The tail empirical process of regularly varying functions of geometrically ergodic Markov chains’, *Stochastic Processes and their Applications* **129**(11), 4209–4238.
- Meyn, S. P. & Tweedie, R. L. (1992), ‘Stability of Markovian processes I: Criteria for discrete-time chains’, *Journal of Theoretical Probability* **24**, 542–574.
- Mikosch, T. & Stărică, C. (2000), ‘Limit theory for the sample autocorrelations and extremes of a GARCH (1, 1) process’, *The Annals of Statistics* **28**(5), 1427–1451.
- Mikosch, T. & Wintenberger, O. (2014), ‘The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains’, *Probability Theory and Related Fields* **159**(1-2), 157–196.
- Pratt, J. W. (1960), ‘On interchanging limits and integrals’, *The Annals of Mathematical Statistics* **31**(1), 74–77.
- Van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

SUPPLEMENT TO THE PAPER CLUSTER BASED INFERENCE FOR EXTREMES OF TIME SERIES

BY HOLGER DREES¹, ANJA JANSSEN², SEBASTIAN NEBLUNG^{1,*}

¹UNIVERSITY OF HAMBURG, DEPARTMENT OF MATHEMATICS, SPST, BUNDESSTR. 55, 20146 HAMBURG, GERMANY, DREES@MATH.UNI-HAMBURG.DE; * SEBASTIAN.NEBLUNG@UNI-HAMBURG.DE

²OTTO-VON-GUERICKE UNIVERSITY OF MAGDEBURG, FACULTY OF MATHEMATICS, IMST, UNIVERSITÄTSPLATZ 2, 39106 MAGDEBURG, GERMANY, ANJA.JANSSEN@OVGU.DE

Abstract We continue the simulation study from section 3 of the main paper and provide additional simulation results in another model. Furthermore, we specify the variance example at the end of Section 2 by giving the concrete variance formulas. We verify three conditions of Theorem 2.3 for solutions to stochastic recurrence equations. Moreover, we state a modified version of the sliding block limit theorem from [Drees & Neblung \(2021\)](#) which is used for the proof of Theorem 2.3 of the main paper. Finally, we provide the proofs for equation (2.1), equation (A.4), Remark 2.2, Lemma A.2, Lemma A.9 (ii), Lemma A.10 and Lemma A.12 of the main paper.

To avoid confusion, we continue the section and figure numbering from the main article.

4. Further simulations. In this section, we present additional simulation results about the finite sample performance of the projection based estimator $\hat{p}_{n,A}$ and its competitors $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ of the cdf of Θ_i . We keep the Monte Carlo setting (sample size, threshold, block length, number of Monte Carlo simulations) used in the paper. In Section 3, among other things, the estimators are discussed in the GARCHt model for lag $i = 10$, where $\hat{p}_{n,A}$ performs best. Figure 3 shows the corresponding results for smaller lags, namely $i = 1$ and $i = 5$, respectively. Because $\hat{p}_{n,A}$ and $\hat{p}_{n,A}^b$ have point mass at 0, $\hat{p}_{n,A}^f$ typically has smallest bias among all estimators for small $|x|$, which is also reflected in the RMSE. In contrast, $\hat{p}_{n,A}$ has smallest variance for most values of x and this effect becomes more pronounced as the lag increases. Consequently, the projection based estimator outperforms the backward estimator, in particular as i increases, and it has smallest RMSE for $|x| > 0.3$ and all lags under consideration.

While in the (asymmetric) SRE model (see Figure 2 in the paper for lag 1 and Figure 4 for lags $i \in \{5, 10\}$) the variances of the three estimators behave similarly, the large bias, which is due to the influence of D_t , often dominates the RMSE, leading to an underperformance of $\hat{p}_{n,A}$. This is particularly true for larger lags and negative x , while for $x > 0$ and lag 10, $\hat{p}_{n,A}$ has minimal RMSE.

Note that here the difference between the conditional distribution of the exceedances over the threshold $F_{|X|}^{\leftarrow}(0.95)$ and its limit distribution obscures almost all real differences between the estimators. It is thus instructive to compare, in addition, the performance of $\hat{p}_{n,A}$, $\hat{p}_{n,A}^f$ and $\hat{p}_{n,A}^b$ interpreted as estimators of the pre-asymptotic cdf $P(X_i/|X_0| \leq x \mid |X_0| > F_{|X|}^{\leftarrow}(0.95))$ (rather than its limit $P\{\Theta_i \leq x\}$), like it was done

AMS 2000 subject classifications: Primary 62G32; secondary 62M10, 62G05, 60F17.

Keywords and phrases: cluster of extremes, empirical processes, extreme value analysis, time series, uniform central limit theorems.

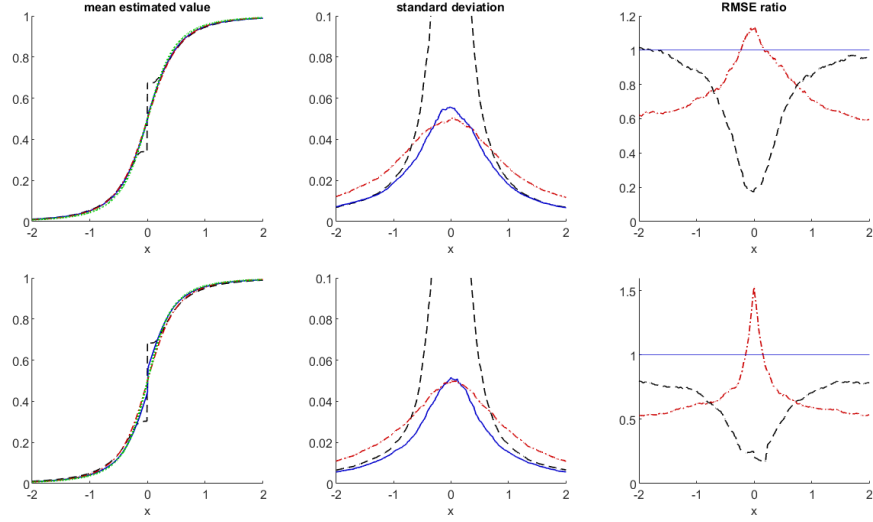


FIGURE 3. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right), of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) in the GARCHt model for lags $i = 1$ (top) and $i = 5$ (bottom); the true cdf is indicated by the green dotted line.

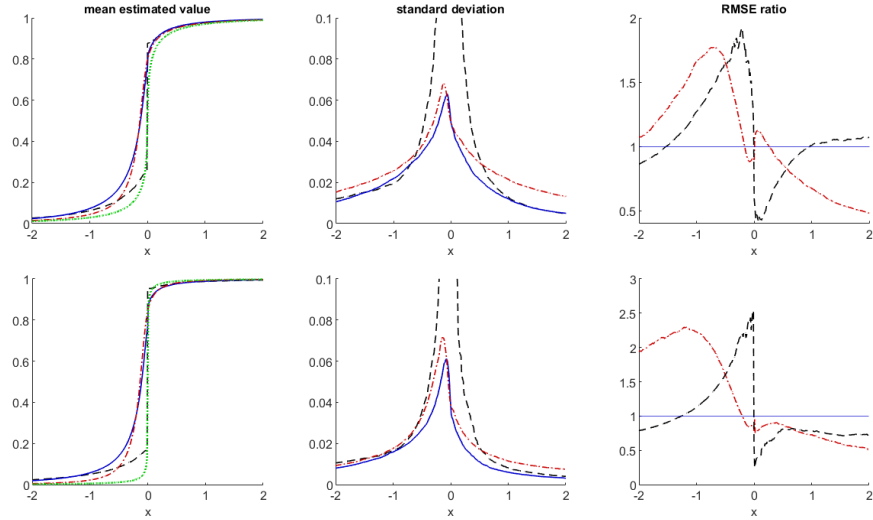


FIGURE 4. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right), of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) in the SRE model for lags $i = 5$ (top) and $i = 10$ (bottom); the true cdf is indicated by the green dotted line.

in Davis et al. (2018) for the backward estimator. The corresponding plots for lag $i = 1$ are shown in Figure 5. Of course, in the left plot only the true cdf has changed and the middle plot is exactly the same as in the bottom line of Figure 2, but the RMSE of the projection based estimator is much more strongly reduced than that of the forward estimators for x not too close to 0. So, in this setting, $\hat{p}_{n,A}$ overall performs best as an estimator of the pre-asymptotic cdf.

In addition to the GARCHt and the SRE model, here we examine a model with trivial tail process:

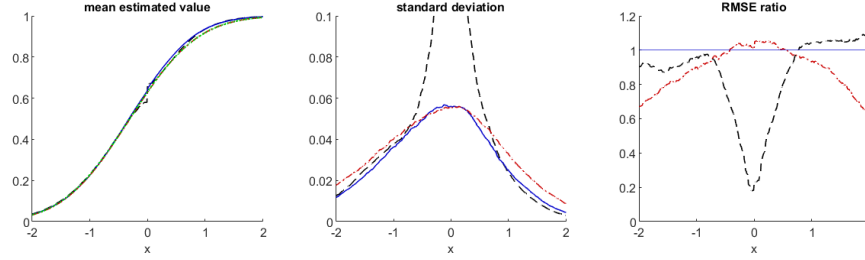


FIGURE 5. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right), of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) as estimators of the pre-asymptotic cdf of $X_i/|X_0|$ given $|X_0| > F_{|X|}^{\leftarrow}(0.95)$ in the SRE model for lag $i = 1$; the true cdf is indicated by the green dotted line.

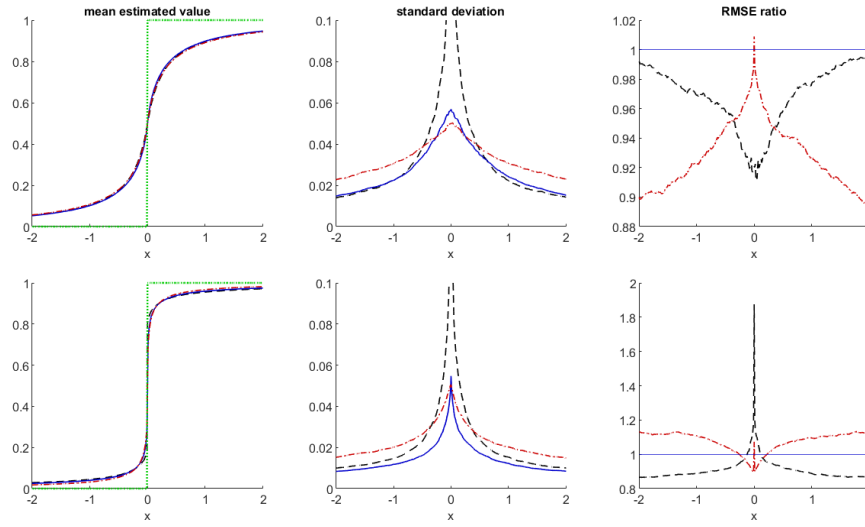


FIGURE 6. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right), of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) in the SV model for lags $i = i$ (top) and $i = 10$ (bottom); the true cdf is indicated by the green dotted line.

- SV: Consider the stationary stochastic volatility model $X_t = \sigma_t \epsilon_t$, $t \in \mathbb{Z}$, with $\log(\sigma_t) = 0.9 \log(\sigma_{t-1}) + Z_t$ where Z_t are i.i.d. standard normal random variables and ϵ_t are i.i.d. with Student's $t_{2.6}$ -distribution, independent of $(Z_t)_{t \in \mathbb{Z}}$. Then $(X_t)_{t \in \mathbb{Z}}$ is a stationary regularly varying time series with $\alpha = 2.6$ (cf. Davis et al. (2018), Section 4). Since the extremal behavior of $(X_t)_{t \in \mathbb{Z}}$ is dominated by the i.i.d. heavy-tailed innovations ϵ_t , we have $P\{\Theta_i = 0\} = 1$ for all $i \neq 0$, see Davis & Mikosch (2009).

Note that the conditions of Section 2 can only be fulfilled for the family of sets $(-\infty, x]$ if a neighborhood of 0 is omitted from the range of x -values. Nevertheless, here we present the results for the full range $[-2, 2]$ (and lags $i \in \{1, 10\}$) in Figure 6. For lag $i = 1$, the projection based estimator performs best, but the loss of efficiency of the other estimators is less than 10%. In contrast, for $i = 10$, the $\hat{p}_{n,A}^f$ has usually a moderately smaller RMSE than $\hat{p}_{n,A}$, which in turn is preferable to $\hat{p}_{n,A}^b$ outside a tiny neighborhood of the origin.

To sum up, usually the projection based estimator has the smallest variance, especially for larger lags. If pre-asymptotic probabilities significantly differ from the corresponding

limit probabilities, $\hat{p}_{n,A}$ may have a larger bias and RMSE than the forward estimator and sometimes also than the backward estimator. Overall, though, $\hat{p}_{n,A}$ shows a more robust performance than both these estimators.

Recall that in the definition of the projection based estimator one has to choose $s_n > i$ which determines the length of the blocks. Next, we examine the sensitivity of the estimators to changes of the block length.

If blocks are chosen too small than they do not capture the dependence structure of a typical cluster of extremes. Moreover, a point mass at 0 is introduced as an artefact of the artificial clipping of the cluster. On the other hand, if the block length is chosen too

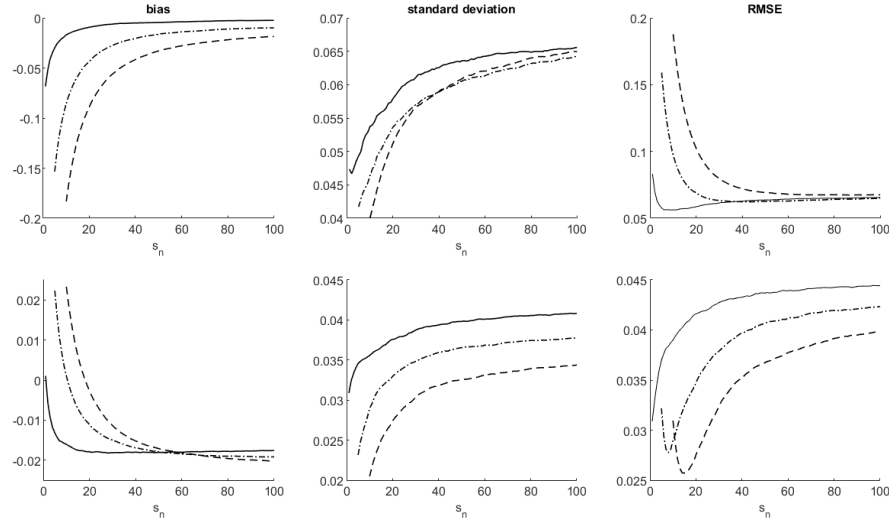


FIGURE 7. Bias (left), standard deviation (middle) and RMSE (right) of $\hat{p}_{n,A}$ versus s_n , for lag 1 (solid line), lag 5 (dashed-dotted line) and lag 10 (dashed line) with $x=0$ (top) and $x=1/2$ (bottom) in the GARCHt model.

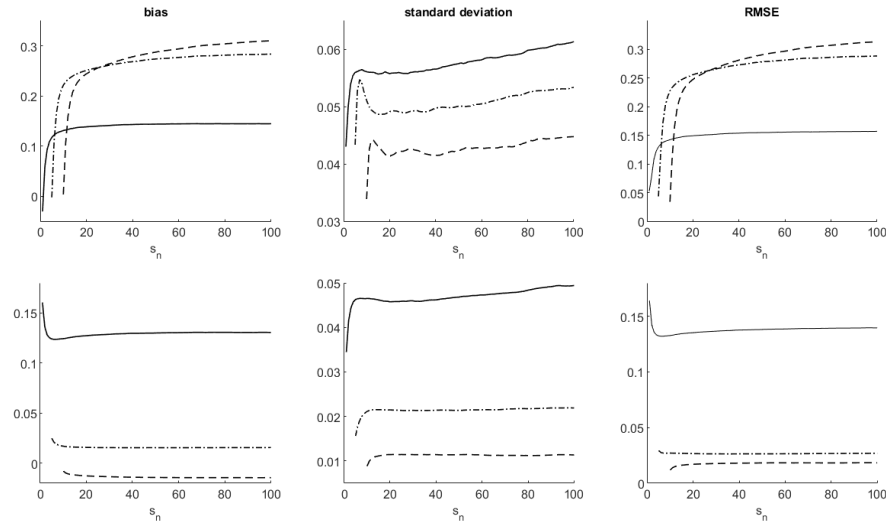


FIGURE 8. Bias (left), standard deviation (middle) and RMSE (right) of $\hat{p}_{n,A}$ versus s_n , for lag 1 (solid line), lag 5 (dashed-dotted line) and lag 10 (dashed line) with $x=0$ (top) and $x=1/2$ (bottom) in the SRE model.

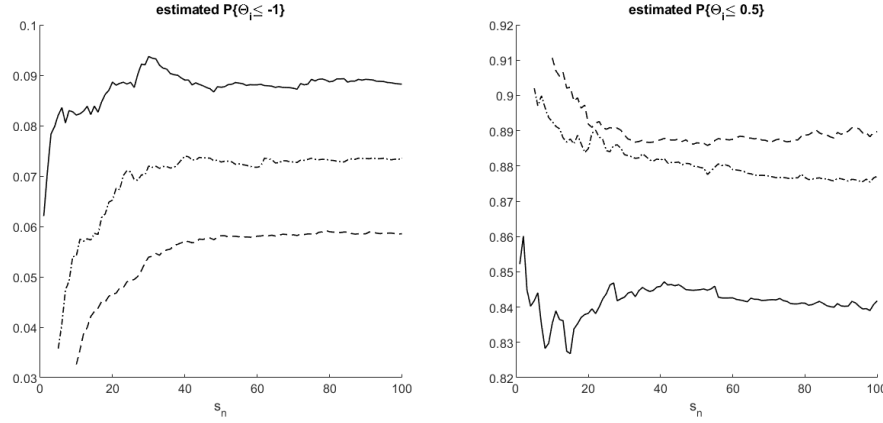


FIGURE 9. $\hat{p}_{n,A}$ versus s_n for a single realization of the GARCHt model for lag 1 (solid line), lag 5 (dashed-dotted line) and lag 10 (dashed line) with $x = -1$ (left) and $x = 0.5$ (right).

large, then almost independent clusters of large observations are compounded, which may lead to a bias and increase the variance, too.

Figure 7 shows the bias, the standard deviation and the RMSE of $\hat{p}_{n,A}$ as function of s_n for the lags $i \in \{1, 5, 10\}$ and $x \in \{0, 0.5\}$ for the GARCHt model and Figure 8 does the same for the SRE model. For the GARCHt model, especially for larger lags, the estimator is clearly biased if s_n is chosen too small (see top left plot in Figure 7), but for $s_n \geq 30$ the bias is small and the RMSE is quite stable. In the SRE model, in most cases the bias is caused by the deviation of the pre-asymptotic cdf from the limit cdf, discussed above. This negative bias happens to (partly) cancel a positive bias caused by too short a block out if s_n is chosen small, leading to a very small RMSE. Apart from this very specific effect, the RMSE is again rather stable if s_n is not too small. We thus recommend to choose the block length not too small (in particular for larger lags i), but unless an excessively large value is used, the performance of the estimator is not very sensitive to this tuning parameter.

To guide the selection of a reasonable value, a graph analogous to a Hill plot may be useful. In Figure 9, for a single time series according to the GARCHt model, the value of $\hat{p}_{n,A}$ is plotted versus s_n , for $A = (-\infty, -1]$ (left) and $A = (-\infty, 1/2]$ (right). We suggest to choose s_n in a range where the plots starts to stabilize, which in this case again leads to a value of about 30.

As always when the peaks-over-threshold approach is employed, all three estimators under consideration depend on the selected threshold u_n . So far, we have fixed this tuning parameter to the empirical 95%-quantile of $|X_0|$. Figure 10 displays the performance of all estimators for the cdf at $x = 0.1$ in the GARCHt and the SRE model for lag 1 for different quantile levels. Unless the threshold is chosen very high, the performance of the projection based estimator and forward estimator seems quite stable. To a lesser extent, this is also true for the backward estimator, but here the performance deteriorates a bit faster. (Note that here we consider a set A whose boundary x is close to the origin, while the backward estimator is known to perform better for sets clearly bounded away from 0.)

To conclude, apparently the newly proposed estimator is quite insensitive to the choice of the block length, and it is not more sensitive to threshold selection than the direct forward estimator.

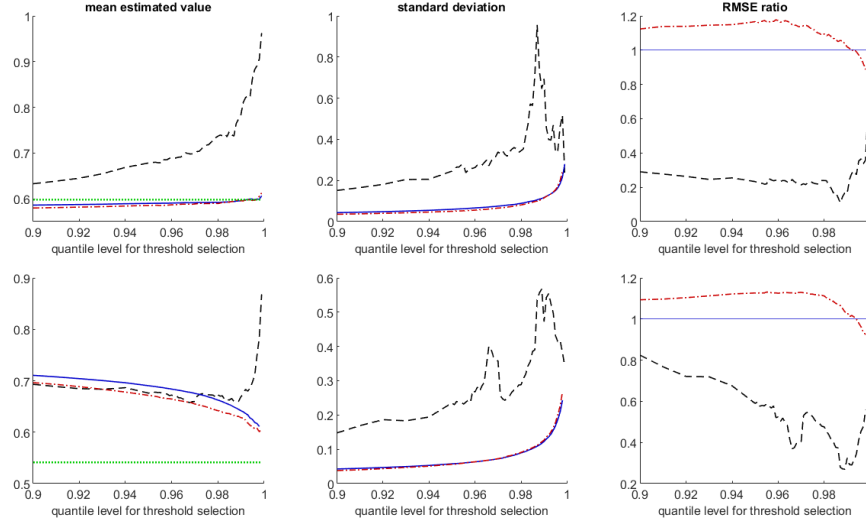


FIGURE 10. Mean (left), standard deviation (middle) and relative efficiency w.r.t. $\hat{p}_{n,A}$ (right) versus $F_{|X|}(u_n)$, of $\hat{p}_{n,A}$ (blue solid line), $\hat{p}_{n,A}^b$ (black dashed line) and $\hat{p}_{n,A}^f$ (red dashed-dotted line) in the GARCHt model (top) and in the SRE model (bottom) for lag $i = 1$ and $x = 0.1$; the true cdf is indicated by the green dotted line.

5. Example: comparison of asymptotic variances. In this section we give further details for the calculation of the covariances in Example 2.4 of the main paper. Recall that we defined the process $W = (W_t)_{t \in \mathbb{Z}}$ by

$$P\{W_0 = a^{-1}, W_1 = -1, W_t = 0, \forall t \notin \{0, 1\}\} = p,$$

$$P\{W_0 = 1, W_1 = b^{-1}, W_t = 0, \forall t \notin \{0, 1\}\} = 1 - p$$

for some $a > 1$, $b > 1$ and $p = [0, 1]$ and we set $\alpha = 1$ and define $\Theta = (\Theta_t)_{t \in \mathbb{Z}}$ such that $P^\Theta = (P^W)^{RS}$, which results in

$$P\{\Theta \in D\} = E \left[\frac{\|W_0\|}{\|W_0\| + \|W_1\|} \mathbf{1}_D \left(\frac{W}{\|W_0\|} \right) + \frac{\|W_1\|}{\|W_0\| + \|W_1\|} \mathbf{1}_D \left(\frac{(W_{t+1})_{t \in \mathbb{Z}}}{\|W_1\|} \right) \right].$$

The specific choice $D = \{(y_t)_{t \in \mathbb{Z}} | y_{-1} = a^{-1}, y_0 = -1, y_t = 0 \forall t \notin \{-1, 0\}\}$ yields

$$P\{\Theta_{-1} = a^{-1}, \Theta_0 = -1, \Theta_t = 0, \forall t \notin \{-1, 0\}\} = p \frac{a}{a+1}.$$

Likewise, we obtain

$$P\{\Theta_0 = 1, \Theta_1 = -a, \Theta_t = 0, \forall t \notin \{0, 1\}\} = p \frac{1}{a+1},$$

$$P\{\Theta_{-1} = b, \Theta_0 = 1, \Theta_t = 0, \forall t \notin \{-1, 0\}\} = (1-p) \frac{1}{1+b},$$

$$P\{\Theta_0 = 1, \Theta_1 = b^{-1}, \Theta_t = 0, \forall t \notin \{0, 1\}\} = (1-p) \frac{b}{1+b}.$$

According to Theorem 2.3, for sets A such that its boundary is disjoint to $B := \{-a, 0, a^{-1}, b^{-1}, 1, b\}$, the asymptotic variance of $\hat{p}_{n,A}$ equals

$$\text{Var}(Z(A) - (p_A - \alpha d_A)Z(\mathbb{R}) - \alpha^2 d_A Z_\phi),$$

with $p_A = P\{\Theta_i \in A\}$ and d_A, Z and Z_ϕ given in Theorem 2.3 and (2.2).

Likewise, according to [Davis et al. \(2018\)](#), Theorem 3.1, for $A = (x, \infty)$ with $x \notin B$, the asymptotic variance of $\hat{p}_{n,A}^b$ is given by

$$\text{Var}(\tilde{Z}(A) - p_A Z(\mathbb{R}) + (\alpha^2 Z_\phi - \alpha Z(\mathbb{R})) E[\log(\|\Theta_i\|) \mathbf{1}_A(\Theta_i)])$$

where $(\tilde{Z}, Z_\phi, Z(\mathbb{R}))$ is a centered Gaussian process with covariance specified by the formulas in Theorem 2.3 and

$$\text{Var}(\tilde{Z}(A)) = E \left[\sum_{j \in \mathbb{Z}} (\|\Theta_j\|^\alpha \wedge 1) \left(\frac{\|\Theta_{j-i}\|^\alpha}{\|\Theta_j\|^\alpha} \mathbf{1}_A \left(\frac{\Theta_j}{\|\Theta_{j-i}\|} \right) \right) \left(\|\Theta_{-i}\|^\alpha \mathbf{1}_A \left(\frac{\Theta_0}{\|\Theta_{-i}\|} \right) \right) \right],$$

$$\text{Cov}(\tilde{Z}(A), Z(\mathbb{R})) = E \left[\sum_{j \in \mathbb{Z}} (\|\Theta_j\|^\alpha \wedge 1) \left(\|\Theta_{-i}\|^\alpha \mathbf{1}_A \left(\frac{\Theta_0}{\|\Theta_{-i}\|} \right) \right) \right]$$

$$\text{Cov}(\tilde{Z}(A), Z_\phi) = E \left[\sum_{j \in \mathbb{Z}} (\|\Theta_j\|^\alpha \wedge 1) (\log(\|\Theta_j\| \vee 1) + \alpha^{-1}) \|\Theta_{-i}\|^\alpha \mathbf{1}_A \left(\frac{\Theta_0}{\|\Theta_{-i}\|} \right) \right].$$

Finally, from [Davis et al. \(2018\)](#), Theorem 3.1, it is known that in this setting the variance of the forward estimator $\hat{p}_{n,A}^f$ equals

$$\sum_{j \in \mathbb{Z}} E \left[(\|\Theta_j\|^\alpha \wedge 1) \left(p_A - \mathbf{1}_A \left(\frac{\Theta_{j+i}}{\|\Theta_j\|} \right) \right) (p_A - \mathbf{1}_A(\Theta_i)) \right].$$

All these expectations can be easily calculated numerically, since the spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$ vanishes for $|t| > 1$ and its distribution is discrete with just 4 points of mass.

Note that the asymptotic variances of the backward estimator and the projection based estimator with known α can be calculated likewise, by omitting all terms involving d_A or $E[\log(\|\Theta_i\|) \mathbf{1}_{\{\Theta_i \in A\}}]$.

6. Verification of (BC), (BC') and (M) (i) for solutions to stochastic recurrence equations. In Subsection 2.1 of the main paper we discussed stationary solutions to stochastic recurrence equations

$$X_t = C_t X_{t-1} + D_t, \quad t \in \mathbb{Z},$$

where C_t are random $d \times d$ -matrices with non-negative entries and D_t are $[0, \infty)^d$ -valued random vectors so that (C_t, D_t) , $t \in \mathbb{Z}$ are i.i.d. The Conditions (SRE) and (SRE') have been used to conclude the existence of a stationary regular varying solution and that $(X_t)_{t \in \mathbb{Z}}$ is an aperiodic, positive Harris recurrent, P^{X_0} -irreducible Feller process. Recall

$$(6.1) \quad X_t = \Pi_{1,t} X_0 + R_t$$

where $\Pi_{j,k} := C_k \cdots C_j$ and $R_t := \sum_{j=1}^t \Pi_{j+1,t} D_j$ are independent of X_0 .

Here we give more details why Conditions (BC), (BC') and (M) (i) are satisfied in this setting. As in the paper, $\|\cdot\|$ denotes both the Euclidean norm of a vector and the corresponding operator norm of a matrix. Recall that for all $q \in (0, \alpha)$ one has for sufficient large m

$$\kappa := E(\|\Pi_{1,m}\|^q) < 1$$

and

$$(6.2) \quad E[\|\Pi_{1,k}\|^q] \leq c_m \tilde{\kappa}^k$$

for all $k \in \mathbb{N}$, some $c_m > 0$ and $\tilde{\kappa} := \kappa^{1/m} < 1$.

Fix some $c \in (0, 1]$ and let $v_{n,c} := P\{\|X_0\| > cu_n\}$. In view of (6.1)

$$\begin{aligned} & P(\|X_k\| > cu_n \mid \|X_0\| > cu_n) \\ & \leq \frac{1}{v_{n,c}} P\{\|X_0\| > cu_n, \|R_k\| + \|\Pi_{1,k}X_0\| > cu_n\} \\ & \leq \frac{1}{v_{n,c}} \left(P\left\{\|X_0\| > cu_n, \|R_k\| > \frac{cu_n}{2}\right\} + P\left\{\|X_0\| > cu_n, \|\Pi_{1,k}\| \|X_0\| > \frac{cu_n}{2}\right\} \right) \\ & = P\left\{\|R_k\| > \frac{cu_n}{2}\right\} + \frac{1}{v_{n,c}} \int_{cu_n}^{\infty} P\left\{\|\Pi_{1,k}\| > \frac{cu_n}{2t}\right\} P^{\|X_0\|}(dt). \end{aligned}$$

Since we assume that C and D only have non-negative entries, the first summand can be bounded by $P\{\|X_k\| > cu_n/2\} = v_{n,c/2}$. For the second term, the generalized Markov inequality and (6.2) yield

$$P\left\{\|\Pi_{1,k}\| > \frac{cu_n}{2t}\right\} \leq E[\|\Pi_{1,k}\|^q] \left(\frac{2t}{cu_n}\right)^q \leq c_m \tilde{\kappa}^k \left(\frac{2t}{cu_n}\right)^q.$$

Hence,

$$\frac{1}{v_{n,c}} \int_{cu_n}^{\infty} P\left\{\|\Pi_{1,k}\| > \frac{cu_n}{2t}\right\} P^{\|X_0\|}(dt) \leq 2^q c_m \tilde{\kappa}^k E\left(\left(\frac{\|X_0\|}{cu_n}\right)^q \mid \|X_0\| > cu_n\right)$$

for all $k \in \mathbb{N}$. To sum up, we have shown

$$P\{\|X_k\| > cu_n \mid \|X_0\| > cu_n\} \leq v_{n,c/2} + 2^q c_m \tilde{\kappa}^k E\left(\left(\frac{\|X_0\|}{cu_n}\right)^q \mid \|X_0\| > cu_n\right) =: e_{n,c}(k)$$

which is (2.11) in the main paper.

Because of regular variation of $\|X_0\|$, $q < \alpha$ and $r_n v_n \rightarrow 0$, we have $E((\|X_0\|/cu_n)^q \mid \|X_0\| > cu_n) \rightarrow E(\|Y_0\|^q)$ and $r_n v_{n,c/2} = r_n v_n (v_{n,c/2}/v_n) \rightarrow 0$. Therefore

$$e_{\infty,c}(k) = \lim_{n \rightarrow \infty} e_{n,c}(k) = 2^q c_m \tilde{\kappa}^k E(\|Y_0\|^q) < \infty$$

and

$$\begin{aligned} \sum_{k=1}^{r_n} e_{n,c}(k) &= r_n v_{n,c/2} + 2^q c_m E\left(\left(\frac{\|X_0\|}{cu_n}\right)^q \mid \|X_0\| > cu_n\right) \sum_{k=1}^{r_n} \tilde{\kappa}^k \\ &\rightarrow 2^q c_m E(\|Y_0\|^q) \left(\frac{1}{1-\tilde{\kappa}} - 1\right) = \sum_{k=1}^{\infty} e_{\infty,c}(k) < \infty, \end{aligned}$$

i.e. condition (BC).

Next observe that, for $p \in (0, \alpha)$, from the drift condition $E(\|X_1\|^p \mid X_0 = y) \leq \beta \|y\|^p + b$, i.e. Assumption 2.1 (iii) of Kulik et al. (2019) which has been established in Subsection 2.1 with some $\beta \in (0, 1)$ and some $b > 0$, one may conclude by induction that $E(\|X_k\|^p \mid X_0 = y) \leq \beta^k \|y\|^p + b/(1-\beta)$ (Douc et al. (2018), Proposition 14.1.8). For all $p, \tilde{p} > 0$ such that $p + \tilde{p} < \alpha$, it follows

$$\begin{aligned} & E\left(\left(\frac{\|X_k\|}{u_n}\right)^p \left(\frac{\|X_0\|}{u_n}\right)^{\tilde{p}} \mathbf{1}_{\{\|X_k\| > u_n\}} \mid \|X_0\| > u_n\right) \\ & \leq v_n^{-1} \int_{u_n}^{\infty} u_n^{-(p+\tilde{p})} E(\|X_k\|^p \mid \|X_0\| = y) \|y\|^{\tilde{p}} P^{\|X_0\|}(dy) \\ & \leq \beta^k E\left(\left(\frac{\|X_0\|}{u_n}\right)^{p+\tilde{p}} \mid \|X_0\| > u_n\right) + \frac{b}{1-\beta} u_n^{-p} E\left(\left(\frac{\|X_0\|}{u_n}\right)^{\tilde{p}} \mid \|X_0\| > u_n\right) \end{aligned}$$

$$(6.3) \quad \leq 2\beta^k E(\|Y_0\|^{p+\tilde{p}}) + \frac{2b}{1-\beta} u_n^{-p} E(\|Y_0\|^{\tilde{p}})$$

for sufficiently large n .

Define $\psi(x) := \max(\log \|x\|, \mathbf{1}_{[1,\infty)}(\|x\|))$, which for all $p > 0$ can be bounded in absolute value by $\tilde{c}_p \|x\|^p \mathbf{1}_{\{\|x\| > 1\}}$ for some constant $\tilde{c}_p$. The uniform moment bound (6.3) readily shows that the random variables $\psi(X_0/u_n)\psi(X_k/u_n)/v_n$ are uniformly integrable, so that the definition of the tail process yields

$$e'_n(k) := E(\psi(X_0/u_n)\psi(X_k/u_n) \mid \|X_0\| > u_n) \rightarrow e'_\infty(k) := E[\psi(Y_0)\psi(Y_k)] < \infty$$

for all $k \in \mathbb{N}$. Moreover, the representation of the forward tail process $Y_k = \|Y_0\| \Theta_0 \Pi_{1,k}$ and (6.2) show that the $e'_\infty(k)$ are summable:

$$\sum_{k=1}^{\infty} e'_\infty(k) \leq \tilde{c}_{q/2}^2 E[\|Y_0\|^q] \sum_{k=1}^{\infty} E[\|\Pi_{1,k}\|^q] \leq \tilde{c}_{q/2}^2 c_m E[\|Y_0\|^q] \sum_{k=1}^{\infty} \tilde{\kappa}^k < \infty.$$

Hence, for all $\varepsilon > 0$, one can find some $L \in \mathbb{N}$ such that $\sum_{k=L+1}^{\infty} e'_\infty(k) < \varepsilon$. Using (6.3) and convergence (2.10), one may conclude for sufficiently large $L' \geq L$

$$\limsup_{n \rightarrow \infty} \left| \sum_{k=1}^{r_n} e'_n(k) - \sum_{k=1}^{\infty} e'_\infty(k) \right| \leq \limsup_{n \rightarrow \infty} \left| \sum_{k=1}^{L'} e'_n(k) - \sum_{k=1}^{L'} e'_\infty(k) \right| + 2\varepsilon = 2\varepsilon.$$

Let ε tend to 0 to obtain (BC').

Finally, (6.3) with $p \in (\alpha(1+\delta)/(1+\zeta), \alpha)$, $\delta < \zeta$, and $\tilde{p} \in (0, \alpha - p)$ shows that the sum on the left-hand side of Condition (M) (i) can be bounded by a multiple of

$$\begin{aligned} & \sum_{k=1}^{r_n} \left(\beta^k E[\|Y_0\|^{p+\tilde{p}}] + u_n^{-p} E[\|Y_0\|^{\tilde{p}}] \right)^{1/(1+\delta)} \\ & \leq \sum_{k=1}^{r_n} (\beta^{1/(1+\delta)})^k (E[\|Y_0\|^{p+\tilde{p}}])^{1/(1+\delta)} + r_n u_n^{-p/(1+\delta)} (E[\|Y_0\|^{\tilde{p}}])^{1/(1+\delta)} \\ & = O(1), \end{aligned}$$

provided $r_n^{1+\zeta} v_n$ is bounded, because u_n is of larger order than $v_n^{\eta-1/\alpha}$ for all $\eta > 0$.

7. Modified sliding blocks limit theorem. In this section we present a slightly modified version of a limit theorem for unbounded functions of sliding blocks introduced by Drees & Neblung (2021) (D&N), Theorem 2.4, which is applied in the proof of Theorem 2.3 of the main paper. To this end, assume that $(X_{n,i})_{1-s_n \leq i \leq n+s_n, n \in \mathbb{N}}$ is a triangular array of row-wise stationary random variables, consider a set $\mathcal{G}$ of functionals defined on vectors of arbitrary length and define $W_{n,j} := (X_{n,j-s_n}, \dots, X_{n,j+s_n})$ for some sequence s_n ,

$$\begin{aligned} V_{n,i}(g) &:= \frac{1}{b_n(g)} \sum_{j=1}^{r_n} g(W_{n,(i-1)r_n+j}), \\ \tilde{V}_{n,i}(g) &:= \frac{1}{b_n(g)} \sum_{j=1}^{r_n-l_n} g(W_{n,(i-1)r_n+j}), \\ \bar{Z}_n(g) &:= \frac{1}{\sqrt{p_n} b_n(g)} \sum_{j=1}^n (g(W_{n,j}) - E g(W_{n,j})), \\ Z_n(g) &:= \frac{1}{\sqrt{p_n}} \sum_{i=1}^{\lfloor n/r_n \rfloor} (V_{n,i}(g) - E V_{n,i}(g)), \quad g \in \mathcal{G}, \end{aligned}$$

and $p_n := P\{\exists g \in \mathcal{G} : V_n(g) \neq 0\}$ for some suitable sequence $r_n = o(n)$; here V_n denotes a random variable with the same distribution as $V_{n,1}$. Moreover, we use the following conditions taken from Section 2 and Appendix A of D&N:

- (A1) $(X_{n,i})_{1 \leq i \leq n}$ is stationary for all $n \in \mathbb{N}$.
- (A2) The sequences $l_n, r_n, s_n \in \mathbb{N}$, p_n , and $b_n(g) > 0$, $g \in \mathcal{G}$, satisfy $s_n \leq l_n = o(r_n)$, $r_n = o(n)$, $p_n \rightarrow 0$ and $r_n = o(\sqrt{p_n} \inf_{g \in \mathcal{G}} b_n(g))$.
- (D0) The processes V_n , $n \in \mathbb{N}$, are separable.
- (MX) $m_n \beta_{n, l_n - (2s_n + 1)}^X \rightarrow 0$ for $m_n := \lfloor n/r_n \rfloor$.
- (C) There exists a function $c : \mathcal{G}^2 \rightarrow \mathbb{R}$ such that

$$\frac{m_n}{p_n} \text{Cov}(V_n(g), V_n(h)) \rightarrow c(g, h), \quad \forall g, h \in \mathcal{G}.$$

- (Δ) $\Delta_n := V_n - \tilde{V}_n$ satisfies
 - (i) $E[(\Delta_n(g) - E[\Delta_n(g)])^2 \mathbf{1}_{\{|\Delta_n(g) - E[\Delta_n(g)]| \leq \sqrt{p_n}\}}] = o(p_n/m_n)$, $\forall g \in \mathcal{G}$,
 - (ii) $P\{|\Delta_n(g) - E[\Delta_n(g)]| > \sqrt{p_n}\} = o(1/m_n)$, $\forall g \in \mathcal{G}$.
 - (iii) $E[(\Delta_n(g) - E[\Delta_n(g)]) \mathbf{1}_{\{|\Delta_n(g) - E[\Delta_n(g)]| \leq \tau \sqrt{p_n}\}}] = o(\sqrt{p_n}/m_n)$, $\forall g \in \mathcal{G}$
 for some $\tau > 0$.
- (L) $E[(V_n(g) - E[V_n(g)])^2 \mathbf{1}_{\{|V_n(g) - E[V_n(g)]| > \epsilon \sqrt{p_n}\}}] = o(p_n/m_n)$, $\forall g \in \mathcal{G}, \epsilon > 0$.
- (D1) There exists a semi-metric ρ on $\mathcal{G}$ such that $\mathcal{G}$ is totally bounded and

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{g, h \in \mathcal{G}, \rho(g, h) < \delta} \frac{m_n}{p_n} E[(V_n(g) - V_n(h))^2] = 0.$$

(D2)

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \int_0^\delta \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{G}, L_2^n)} d\epsilon = 0,$$

where $N_{[\cdot]}(\epsilon, \mathcal{G}, L_2^n)$ denotes the ϵ -bracketing number of $\mathcal{G}$ w.r.t. L_2^n , i.e. the smallest number N_ϵ such that for each $n \in \mathbb{N}$ there exists a partition $(\mathcal{G}_{n,k}^\epsilon)_{1 \leq k \leq N_\epsilon}$ of $\mathcal{G}$ satisfying

$$\frac{m_n}{p_n} E^* \left[\sup_{g, h \in \mathcal{G}_{n,k}^\epsilon} (V_n(g) - V_n(h))^2 \right] \leq \epsilon^2, \quad \forall 1 \leq k \leq N_\epsilon.$$

(Here E^* denotes the outer expectation.)

(D3) Denote by $N(\epsilon, \mathcal{G}, d_n)$ the ϵ -covering number of $\mathcal{G}$ w.r.t. the random semi-metric

$$d_n(g, h) = \left(\frac{1}{p_n} \sum_{i=1}^{m_n} (V_{n,i}^*(g) - V_{n,i}^*(h))^2 \right)^{1/2}$$

with $V_{n,i}^*$, $1 \leq i \leq m_n$, independent copies of $V_{n,1}$, i.e. $N(\epsilon, \mathcal{G}, d_n)$ is the smallest number of balls with respect to d_n with radius ϵ that is needed to cover $\mathcal{G}$. We assume

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} P^* \left\{ \int_0^\delta \sqrt{\log(N(\epsilon, \mathcal{G}, d_n))} d\epsilon > \tau \right\} = 0, \quad \forall \tau > 0.$$

Condition (Δ) is always fulfilled if

$$(7.1) \quad E[(\Delta_n(g))^2] = o(p_n/m_n), \quad \forall g \in \mathcal{G}.$$

Then we can modify Theorem 2.4 of D&N as follows.

THEOREM 7.1. (i) Suppose the conditions (A1), (A2), (D0), (MX) and (C) are met. Moreover, assume condition (L) is satisfied and

$$(7.2) \quad E \left[\left(\sum_{i=1}^{r_n} |g(W_{n,i})| \right)^2 \right] = O \left(\frac{p_n b_n^2(g)}{m_n} \right), \quad \forall g \in \mathcal{G}.$$

Then the fidis of $(Z_n(g))_{g \in \mathcal{G}}$ and of $(\bar{Z}_n(g))_{g \in \mathcal{G}}$ converge to the fidis of the Gaussian process $(Z(g))_{g \in \mathcal{G}}$ defined in Theorem 2.1 of D&N.

(ii) If, in addition, $g_{\max} = \sup_{g \in \mathcal{G}} |g|$ is measurable, $b_n(g) = b_n$ is the same for all $g \in \mathcal{G}$, (7.2) holds for $g = g_{\max}$ and the conditions (D1) and (D2) or the conditions (D1) and (D3) are fulfilled, then the processes $(Z_n(g))_{g \in \mathcal{G}}$ and $(\bar{Z}_n(g))_{g \in \mathcal{G}}$ converge weakly to $(Z(g))_{g \in \mathcal{G}}$ uniformly.

So, in comparison with Theorem 2.4 of D&N, basically condition (2.8) in that paper, which is the analog to (7.2) with some exponent strictly greater than 2, is replaced with the weaker condition (7.2), a weak technical condition is omitted, and as a compensation condition (L) is added.

PROOF OF THEOREM 7.1. The proof is basically the same as for Theorem 2.4 in D&N. We apply Theorem A.1 of D&N to the observations $X'_{n,t} := X_{n,t-s_n}$, $1 \leq t \leq n' := n + 2s_n$, and block lengths $s'_n := 2s_n + 1$ and $l'_n := l_n + s_n + 1$ to establish fidi-convergence of $(Z_n(g))_{g \in \mathcal{G}}$. Only Condition (Δ) must be verified, because (L) is assumed and the remaining conditions follow as in the proof of Theorem 2.1 of D&N. Since

$$\begin{aligned} E \left[\left(\sum_{i=1}^{r_n} |g(W_{n,i})| \right)^2 \right] &\geq \sum_{j=1}^{\lfloor r_n/l_n \rfloor} E \left[\left(\sum_{i=1}^{l_n} |g(W_{n,(j-1)l_n+i})| \right)^2 \right] \\ &= \lfloor r_n/l_n \rfloor E \left[\left(\sum_{i=1}^{l_n} |g(W_{n,i})| \right)^2 \right], \end{aligned}$$

(7.2) and $l_n = o(r_n)$ imply

$$\begin{aligned} E(\Delta_n(g)^2) &\leq \frac{1}{b_n^2(g)} E \left[\left(\sum_{i=1}^{l_n} |g(W_{n,i})| \right)^2 \right] \leq \frac{1}{b_n^2(g) \lfloor r_n/l_n \rfloor} E \left[\left(\sum_{i=1}^{r_n} |g(W_{n,i})| \right)^2 \right] \\ &= O \left(\frac{l_n}{r_n b_n^2(g)} \frac{p_n b_n^2(g)}{m_n} \right) = o \left(\frac{p_n}{m_n} \right). \end{aligned}$$

Hence, equation (7.1) holds, which in turn yields (Δ). Now, the convergence of the fidis of $(Z_n(g))_{g \in \mathcal{G}}$ follows from Theorem A.1 of D&N. Similarly,

$$E \left((\bar{Z}_n(g) - Z_n(g))^2 \right) \leq \frac{1}{p_n b_n^2(g)} E \left[\left(\sum_{j=r_n m_n + 1}^{n-s_n} |g(W_{n,j})| \right)^2 \right] = O \left(\frac{1}{m_n} \right) \rightarrow 0,$$

so that the fidi-convergence of $(\bar{Z}_n(g))_{g \in \mathcal{G}}$ follows, too.

Part (ii) of the assertion can be concluded by the same arguments as in the proof of Theorem 2.4 of D&N. $\square$

8. Proof of (2.1) and (A.4). First we want to establish (2.1) of the paper, i.e. the convergence

$$(8.1) \quad E(g_A(W_{n,0}) \mid \|X_0\| > u_n) \rightarrow E(g_A(Y)) = E \left(\|\Theta\|_\alpha^{-\alpha} \sum_{h \in \mathbb{Z}} \|\Theta_h\|^\alpha \mathbf{1}_A(\Theta_{h+i}/\|\Theta_h\|) \right)$$

for $g_A : l_\alpha \rightarrow [0, 1]$,

$$g_A((w_h)_{h \in \mathbb{Z}}) := \frac{\mathbf{1}_{\{\|w_0\| > 1\}}}{\sum_{h \in \mathbb{Z}} \|w_h\|^\alpha} \sum_{h \in \mathbb{Z}} \|w_h\|^\alpha \mathbf{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right)$$

and $W_{n,t} := (X_{n,t+h})_{|h| \leq s_n}$, $X_{n,t} := X_t/u_n$. To this end, we define the approximating functions $g_A^{(m)} : l_\alpha \rightarrow \mathbb{R}$,

$$g_A^{(m)}((w_h)_{h \in \mathbb{Z}}) := \frac{\mathbf{1}_{\{\|w_0\| > 1\}}}{\sum_{|h| \leq m} \|w_h\|^\alpha} \sum_{|h| \leq m} \|w_h\|^\alpha \mathbf{1}_A\left(\frac{w_{h+i}}{\|w_h\|}\right)$$

for all $m \in \mathbb{N}$. As a finite sum of continuous functions, $g_A^{(m)}$ is $P^{(Y_{t+j})_{t \in \mathbb{Z}}}$ -a.s. continuous for $j \in \mathbb{Z}$ if $P\{\exists h \in \mathbb{Z} : Y_{h+i}/\|Y_h\| \in \partial A, \|Y_h\| > 0\} = 0$ and $P\{\|Y_j\| = 1\} = 0$. While the former equality is ensured by (C Θ) and Lemma A.3, the latter follows from $Y_j = \Theta_j\|Y_0\|$, where $\|\Theta_j\|$ and $\|Y_0\|$ are independent and $\|Y_0\|$ has a Pareto(α)-distribution. Here we only need the continuity for $j = 0$, but below the general version is used in the proof of (A.4) for the continuity of f .

Since $g_A^{(m)}$ is bounded by 1, the weak convergence $\mathcal{L}((X_{n,h})_{|h| \leq m}) \mid \|X_0\| > u_n \rightarrow \mathcal{L}((Y_h)_{|h| \leq m})$ defining the tail process implies

$$(8.2) \quad E(g_A^{(m)}(W_{n,0}) \mid \|X_0\| > u_n) \rightarrow E(g_A^{(m)}(Y)).$$

One has $g_A^{(m)}(w) \rightarrow g_A(w)$ as $m \rightarrow \infty$ for all $w = (w_h)_{h \in \mathbb{Z}} \in l_\alpha$. Since $|g_A^{(m)}| \leq 1$ for all $m \in \mathbb{N}$, dominated convergence implies $E(g_A^{(m)}(Y)) \rightarrow E(g_A(Y))$ as $m \rightarrow \infty$.

Next, we prove that the difference between $g_A^{(m)}(W_{n,j})$ and $g_A(W_{n,j})$ is asymptotically negligible for all $j \in \mathbb{Z}$ as n and m tend to ∞ . While here this result is only needed for $j = 0$, the general version is used in the proof of (A.4) below. Using (A.18) of the paper, one has for sufficiently large n (such that $s_n \geq m + |i| + |j|$)

$$\begin{aligned} & |g_A^{(m)}(W_{n,j}) - g_A(W_{n,j})| \\ &= \left| \frac{\sum_{|h| \leq m} \|X_{n,h+j}\|^\alpha \mathbf{1}_A\left(\frac{X_{h+j+i}}{\|X_{h+j}\|}\right) \sum_{m < |k| \leq s_n} \|X_{n,k+j}\|^\alpha}{\sum_{|h| \leq m} \|X_{n,h+j}\|^\alpha \sum_{|k| \leq s_n} \|X_{n,k+j}\|^\alpha} \right. \\ & \quad \left. - \frac{\sum_{m < |h| \leq s_n} \|X_{n,h+j}\|^\alpha \left(\mathbf{1}_{\{h \in H_n\}} \mathbf{1}_A\left(\frac{X_{h+j+i}}{\|X_{h+j}\|}\right) + \mathbf{1}_{\{h \in H_n^c\}} \mathbf{1}_A(0) \right) \sum_{|k| \leq m} \|X_{n,k+j}\|^\alpha}{\sum_{|h| \leq m} \|X_{n,h+j}\|^\alpha \sum_{|k| \leq s_n} \|X_{n,k+j}\|^\alpha} \right| \\ &\leq 2 \frac{\sum_{m < |h| \leq s_n} \|X_{h+j}\|^\alpha}{\sum_{|h| \leq s_n} \|X_{h+j}\|^\alpha}. \end{aligned}$$

For $c \in (0, 1)$ from (TC), Condition (BC) implies

$$\begin{aligned} & E\left(\frac{\sum_{m < |h| \leq s_n} \|X_{h+j}\|^\alpha \mathbf{1}_{\{\|X_{h+j}\| > cu_n\}}}{\sum_{|h| \leq s_n} \|X_{h+j}\|^\alpha} \mid \|X_0\| > u_n\right) \\ &\leq \sum_{m < |h| \leq s_n} P(\|X_{h+j}\| > cu_n \mid \|X_0\| > cu_n) \frac{P(\|X_0\| > cu_n)}{P(\|X_0\| > u_n)} \\ &\leq 4c^{-\alpha} \sum_{m < |h| \leq s_n} e_{n,c}(h+j) \end{aligned}$$

for sufficiently large n , due to regular variation of $\|X_0\|$. Therefore,

$$\begin{aligned} & E(|g_A^{(m)}(W_{n,j}) - g_A(W_{n,j})| \mid \|X_0\| > u_n) \\ & \leq 4c^{-\alpha} \sum_{m < |h| \leq s_n} e_{n,c}(h+j) + E\left(\frac{\sum_{m < |h| \leq s_n} \|X_{h+j}\|^\alpha \mathbf{1}_{\{\|X_{h+j}\| \leq cu_n\}}}{\sum_{|h| \leq s_n} \|X_{h+j}\|^\alpha} \mid \|X_0\| > u_n\right) \end{aligned}$$

for sufficiently large $n \in \mathbb{N}$ and all $j \in \mathbb{Z}$. Thus, condition (BC) and (TC) yield

$$(8.3) \quad \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} E(|g_A^{(m)}(W_{n,j}) - g_A(W_{n,j})| \mid \|X_0\| > u_n) = 0.$$

Combine this with (8.2) and $E(g_A^{(m)}(Y)) \rightarrow E(g_A(Y))$ to conclude (8.1), i.e. (2.1) of the paper.

Next we verify (A.4) of the paper, i.e.

$$(8.4) \quad E(f(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n) \rightarrow E(f((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}}))$$

for all $j \in \mathbb{Z}$, with f defined as $f = f_{A,B} : l_\alpha \times l_\alpha \rightarrow [0, 1]$ with

$$\begin{aligned} f((y_t)_{t \in \mathbb{Z}}, (z_t)_{t \in \mathbb{Z}}) &:= \mathbf{1}_{\{\|y_0\| > 1\}} \mathbf{1}_{\{\|z_0\| > 1\}} \\ &\quad \times \left(\sum_{h \in \mathbb{Z}} \frac{\|z_h\|^\alpha}{\sum_{k \in \mathbb{Z}} \|z_k\|^\alpha} \mathbf{1}_A\left(\frac{z_{h+i}}{\|z_h\|}\right) \right) \left(\sum_{l \in \mathbb{Z}} \frac{\|y_l\|^\alpha}{\sum_{k \in \mathbb{Z}} \|y_k\|^\alpha} \mathbf{1}_B\left(\frac{y_{l+i}}{\|y_l\|}\right) \right). \end{aligned}$$

Again we define approximating functions $f^{(m)} : l_\alpha \times l_\alpha \rightarrow \mathbb{R}$ by

$$\begin{aligned} f^{(m)}((y_t)_{t \in \mathbb{Z}}, (z_t)_{t \in \mathbb{Z}}) &:= \mathbf{1}_{\{\|y_0\| > 1\}} \mathbf{1}_{\{\|z_0\| > 1\}} \\ &\quad \times \left(\sum_{|h| \leq m} \frac{\|z_h\|^\alpha}{\sum_{|k| \leq m} \|z_k\|^\alpha} \mathbf{1}_A\left(\frac{z_{h+i}}{\|z_h\|}\right) \right) \left(\sum_{|l| \leq m} \frac{\|y_l\|^\alpha}{\sum_{|k| \leq m} \|y_k\|^\alpha} \mathbf{1}_B\left(\frac{y_{l+i}}{\|y_l\|}\right) \right), \end{aligned}$$

for all $m \in \mathbb{N}$. Observe that $f^{(m)}(y, z) = g_B^{(m)}(y)g_A^{(m)}(z)$ and $f(y, z) = g_B(y)g_A(z)$, so that $f^{(m)}$ is $P^{((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}})}$ -a.s. continuous and bounded by 1. Thus, the definition of the tail process implies

$$(8.5) \quad E(f^{(m)}(W_{n,0}, W_{n,j}) \mid \|X_0\| > u_n) \rightarrow E(f^{(m)}((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}}))$$

for all $m \in \mathbb{N}$. Moreover,

$$\begin{aligned} & |f^{(m)}(W_{n,0}, W_{n,j}) - f(W_{n,0}, W_{n,j})| \\ & \leq g_A^{(m)}(W_{n,0})|g_B^{(m)}(W_{n,j}) - g_B(W_{n,j})| + |g_A^{(m)}(W_{n,0}) - g_A(W_{n,0})|g_B(W_{n,j}) \\ & \leq |g_B^{(m)}(W_{n,j}) - g_B(W_{n,j})| + |g_A^{(m)}(W_{n,0}) - g_A(W_{n,0})|. \end{aligned}$$

Thus, in view of (8.3), for all $j \in \mathbb{Z}$,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} E(|f^{(m)}(W_{n,0}, W_{n,j}) - f(W_{n,0}, W_{n,j})| \mid \|X_0\| > u_n) = 0.$$

Combining this with (8.5) and $E(f^{(m)}((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}})) \rightarrow E(f((Y_t)_{t \in \mathbb{Z}}, (Y_{t+j})_{t \in \mathbb{Z}}))$ as $m \rightarrow \infty$, which holds due to dominated convergence and $|f^{(m)}| \leq 1$ for all $m \in \mathbb{N}$, yields (8.4), i.e. (A.4) of the paper.

9. Proof of Remark 2.2. We start with part (a). Since $Y_h = \Theta_h \|Y_0\|$ and $|a + b|^{1+\delta} \leq 2^{1+\delta}(|a|^{1+\delta} + |b|^{1+\delta})$ for $a, b \in \mathbb{R}$ and $\delta \geq 0$, we conclude

$$\begin{aligned} & E\left(\frac{\sum_{|h| \leq m} |\log \|Y_h\||^{1+\delta} \|Y_h\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha}\right) \\ &= E\left(\frac{\sum_{|h| \leq m} |\log \|\Theta_h\| + \log \|Y_0\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha}\right) \\ &\leq 2^{1+\delta} \left(E\left(\frac{\sum_{|h| \leq m} |\log \|\Theta_h\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha}\right) + E\left(\frac{\sum_{|h| \leq m} |\log \|Y_0\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha}\right) \right) \\ &= 2^{1+\delta} \left(E\left(\frac{\sum_{|h| \leq m} |\log \|\Theta_h\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha}\right) + E(|\log \|Y_0\||^{1+\delta}) \right) < \infty, \end{aligned}$$

where in the last expression the first expectation is finite due to (M) (ii) and the second expectation, because $\log \|Y_0\|$ is exponentially distributed. Thus, Condition (M) (ii) implies (2.5) of the paper.

Conversely, (2.5) of the paper implies

$$\begin{aligned} & E\left(\frac{\sum_{|h| \leq m} |\log \|\Theta_h\||^{1+\delta} \|\Theta_h\|^\alpha}{\sum_{|k| \leq m} \|\Theta_k\|^\alpha}\right) \\ &\leq 2^{1+\delta} \left(E\left(\frac{\sum_{|h| \leq m} |\log \|Y_h\||^{1+\delta} \|Y_h\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha}\right) + E(|\log \|Y_0\||^{1+\delta}) \right) < \infty. \end{aligned}$$

Thus, Condition (M) (ii) is equivalent to (2.5) of the main paper.

Next we turn to part (b) of Remark 2.2, which states that (M') implies (M) (ii) and (TC). According to (a), it suffices to establish (2.5) of the paper, which follows from the definition of the tail process if we prove uniform integrability of

$$\frac{\mathbb{1}_{\{\|X_0\| > u_n\}}}{v_n} \cdot \frac{\sum_{|h| \leq (s_n \wedge m)} |\log \|X_{n,h}\||^{1+\delta} \|X_{n,h}\|^\alpha}{\sum_{|k| \leq (s_n \wedge m)} \|X_{n,k}\|^\alpha}, \quad n, m \in \mathbb{N}.$$

This, in turn, is implied by the uniform moment bound

$$(9.1) \quad \sup_{m \in \mathbb{N}} \sup_{n \in \mathbb{N}} E\left(\left(\frac{\sum_{|h| \leq (s_n \wedge m)} |\log \|X_{n,h}\||^{1+\delta} \|X_{n,h}\|^\alpha}{\sum_{|k| \leq (s_n \wedge m)} \|X_{n,k}\|^\alpha}\right)^{1+\eta} \middle| \|X_0\| > u_n\right) < \infty$$

for $\eta = (\delta' - \delta)/(1 + \delta) > 0$. By Jensen's inequality, the expectation can be bounded by

$$E\left(\frac{\sum_{|h| \leq (s_n \wedge m)} |\log \|X_{n,h}\||^{(1+\delta)(1+\eta)} \|X_{n,h}\|^\alpha}{\sum_{|k| \leq (s_n \wedge m)} \|X_{n,k}\|^\alpha} \middle| \|X_0\| > u_n\right).$$

On the set $\{\|X_0\| > u_n\} = \{\|X_{n,0}\| > 1\}$, the term in the expectation can be bounded by

$$\begin{aligned} & \sup_{|h| \leq (s_n \wedge m)} (\log \|X_{n,h}\|)^{1+\delta'} \mathbb{1}_{\{\|X_{n,h}\| > 1\}} + \sum_{|h| \leq (s_n \wedge m)} |\log \|X_{n,h}\||^{1+\delta'} \|X_{n,h}\|^\alpha \mathbb{1}_{\{\|X_{n,h}\| \leq 1\}} \\ &\leq \sup_{|h| \leq s_n} (\log^+ \|X_{n,h}\|)^{1+\delta'} + \sum_{h=-s_n}^{s_n} (\log^- \|X_{n,h}\|)^{1+\delta'} \|X_{n,h}\|^\alpha. \end{aligned}$$

Thus, Conditions (M') (i) and (ii) imply (9.1), and hence in turn condition (M) (ii).

To verify (TC) under (M') (ii), choose $c < e^{-1}$ so that $\mathbf{1}_{\{\|X_h\| \leq cu_n\}} \leq (\log^{-}(\|X_h\|/u_n))^{1+\delta'}$. Since, for $s_n \geq |j|$, the denominator is at least u_n^α , it directly follows

$$\begin{aligned} & E\left(\frac{\sum_{m < |h| \leq s_n} \|X_{h+j}\|^\alpha \mathbf{1}_{\{\|X_{h+j}\| \leq cu_n\}}}{\sum_{|h| \leq s_n} \|X_{h+j}\|^\alpha} \mid \|X_0\| > u_n\right) \\ & \leq \sum_{|k| \leq s_n+j} E\left(\left(\log^{-} \frac{\|X_k\|}{u_n}\right)^{1+\delta'} \left(\frac{\|X_k\|}{u_n}\right)^\alpha \mid \|X_0\| > u_n\right). \end{aligned}$$

Because $x \mapsto (\log^{-} x)^{1+\delta'} x^\alpha$ is a bounded function on $(0, \infty)$ and thus the sum over k with $s_n < |k| \leq s_n + j$ is bounded, Condition (M') (ii) ensures that (TC) holds.

10. Proof of Lemma A.2. Next, we give the proof for Lemma A.2 of the main paper. This lemma shows that the well-known anticlustering condition is implied by our condition (BC) which also restricts the size of clusters of extremes.

LEMMA 10.1. *If (RV) and (BC) holds, then the so-called anticlustering condition*

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\max_{m \leq |t| \leq r_n} \|X_t\| > cu_n \mid \|X_0\| > cu_n\right) = 0$$

is satisfied for all $c \in (0, \infty)$.

PROOF. For $c \leq 1$, the assertion is an immediate consequence of (BC):

$$\begin{aligned} & \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\max_{m \leq |t| \leq r_n} \|X_t\| > cu_n \mid \|X_0\| > cu_n\right) \\ & \leq \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{m \leq |t| \leq r_n} P(\|X_t\| > cu_n \mid \|X_0\| > cu_n) \leq \lim_{m \rightarrow \infty} 2 \sum_{t=m}^{\infty} e_{\infty, c}(t) = 0. \end{aligned}$$

For $c > 1$ and sufficiently large n , stationarity, regular variation and (BC) yield

$$\begin{aligned} P(\|X_k\| > cu_n \mid \|X_0\| > cu_n) & \leq P(\|X_{|k|}\| > u_n \mid \|X_0\| > u_n) \frac{P\{\|X_0\| > u_n\}}{P\{\|X_0\| > cu_n\}} \\ & \leq 2c^\alpha e_{n,1}(|k|) \end{aligned}$$

for all $k \in \mathbb{Z}$. Thus, (BC) implies

$$\begin{aligned} & \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\max_{m \leq |t| \leq r_n} \|X_t\| > cu_n \mid \|X_0\| > cu_n\right) \\ & \leq \lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{m \leq |t| \leq r_n} P(\|X_t\| > cu_n \mid \|X_0\| > cu_n) \leq \lim_{m \rightarrow \infty} 4c^\alpha \sum_{t=m}^{\infty} e_{\infty,1}(t) = 0. \quad \square \end{aligned}$$

11. Proof of Lemma A.9, part (ii). Part (i) of Lemma A.9 was already established in the main paper. Here we prove the second assertion by similar arguments. First, recall the Lemma A.9 (ii):

LEMMA 11.1. *Under Conditions (RV), (S) and (M) (ii)*

$$\lim_{m \rightarrow \infty} \sup_{A \in \mathcal{A}} |d_A^{(m)} - d_A| = 0$$

with $d_A^{(m)} = E(f_A^{(m)}(Y))$ and d_A defined in Theorem 2.3.

PROOF. In a first step we prove

$$E[f_A(Y)] = d_A.$$

Direct calculations using $Y_h = \Theta_h \|Y_0\|$ with Θ_h and $\|Y_0\|$ independent, $P\{\|Y_0\| > y\} = y^{-\alpha} \wedge 1$ yield

$$\begin{aligned}
& E[f_A(Y)] \\
&= E \left[\sum_{h \in \mathbb{Z}} \left(\frac{\log(\|Y_h\|) \|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} - \frac{\|Y_h\|^\alpha \sum_{k \in \mathbb{Z}} \log(\|Y_k\|) \|Y_k\|^\alpha}{\|Y\|_\alpha^{2\alpha}} \right) \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \left(\frac{\log(\|\Theta_h\|) \|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} + \log(\|Y_0\|) \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\|\Theta_h\|^\alpha \sum_{k \in \mathbb{Z}} \log(\|\Theta_k\|) \|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}} \right. \right. \\
&\quad \left. \left. - \log(\|Y_0\|) \frac{\|\Theta_h\|^\alpha \sum_{k \in \mathbb{Z}} \|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^{2\alpha}} \right) \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \left(\log(\|\Theta_h\|) \frac{\sum_{k \in \mathbb{Z}} \|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} - \frac{\sum_{k \in \mathbb{Z}} \log(\|\Theta_k\|) \|\Theta_k\|^\alpha}{\|\Theta\|_\alpha^\alpha} \right) \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \left(- \frac{\sum_{k \in \mathbb{Z}} \log(\|\Theta_k\| / \|\Theta_h\|) (\|\Theta_k\| / \|\Theta_h\|)^\alpha}{\|\Theta\|_\alpha^\alpha} \right) \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \right] \\
&\quad (11.1) \\
&= E \left[- \sum_{k \in \mathbb{Z}} \log(\|\Theta_k\|) \|\Theta_k\|^\alpha \|\Theta\|_\alpha^{-\alpha} \mathbf{1}_A(\Theta_i) \right] = d_A,
\end{aligned}$$

because the last two expectations can be interpreted as expectations of the same function on l_α w.r.t. $(P^\Theta)^{RS}$ and P^Θ , respectively, and these two distributions coincide. The remaining part of the proof is similar to the proof of part (i) of Lemma A.9. Direct calculations show

$$\begin{aligned}
& |d^{(m)}(A) - d(A)| \leq E(|f_A^{(m)}(Y) - f_A(Y)|) \\
&\leq E \left| \frac{\sum_{|h| \leq m} \log(\|Y_h\|) \|Y_h\|^\alpha \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right)}{\sum_{|k| \leq m} \|Y_k\|^\alpha} - \frac{\sum_{h \in \mathbb{Z}} \log(\|Y_h\|) \|Y_h\|^\alpha \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right)}{\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right| \\
&\quad + E \left| \frac{\sum_{|h| \leq m} \|Y_h\|^\alpha \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \sum_{|k| \leq m} \log(\|Y_k\|) \|Y_k\|^\alpha}{(\sum_{|k| \leq m} \|Y_k\|^\alpha)^2} \right. \\
&\quad \left. - \frac{\sum_{h \in \mathbb{Z}} \|Y_h\|^\alpha \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \sum_{k \in \mathbb{Z}} \log(\|Y_k\|) \|Y_k\|^\alpha}{(\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha)^2} \right| \\
&=: T_1^Y + T_2^Y.
\end{aligned}$$

Similar as for the bound of T_1 in the proof of part (i), we obtain

$$\begin{aligned}
T_1^Y &\leq E \left(\frac{\sum_{|h| \leq m} |\log \|Y_h\|| \|Y_h\|^\alpha \sum_{|k| > m} \|Y_k\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha \sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right) + E \left(\frac{\sum_{|h| > m} |\log \|Y_h\|| \|Y_h\|^\alpha}{\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right) \\
&=: T_{1,1}^Y + T_{1,2}^Y.
\end{aligned}$$

Applying the Hölder inequality for expectations and for sums, we can bound $T_{1,1}^Y$ by

$$\left(E \left(\frac{\sum_{|h| \leq m} |\log \|Y_h\||^{1+\delta} \|Y_h\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha} \right) \right)^{1/(1+\delta)} \left(E \left(\left(\frac{\sum_{|k| > m} \|Y_k\|^\alpha}{\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right)^{(1+\delta)/\delta} \right) \right)^{\delta/(1+\delta)}.$$

According to Remark 2.2 (a), the first expectation is bounded if Condition (M) (ii) is met. The second expectation converges to 0 as $m \rightarrow \infty$ by monotone convergence. Hence $\lim_{m \rightarrow \infty} T_{1,1}^Y = 0$. Since (2.5) in combination with $|\log \|Y_h\|| \leq |\log \|Y_h\||^{1+\delta} + 1$ also implies that $T_{1,2}^Y < \infty$ for all $m \in \mathbb{N}$, monotone convergence yields $\lim_{m \rightarrow \infty} T_{1,2}^Y = 0$, too.

By analogous arguments as used for T_2 in the proof of part (i), one can show that

$$\begin{aligned} T_2^Y &\leq 2E \left(\frac{\sum_{|k| \leq m} |\log \|Y_k\|| \|Y_k\|^\alpha \sum_{|h| > m} \|Y_h\|^\alpha}{\sum_{|k| \leq m} \|Y_k\|^\alpha \sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right) \\ &\quad + E \left(\frac{\sum_{|k| > m} |\log \|Y_k\|| \|Y_k\|^\alpha}{\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha} \right) + E \left(\frac{\sum_{|h| > m} \|Y_h\|^\alpha \sum_{|k| \leq m} |\log \|Y_k\|| \|Y_k\|^\alpha}{(\sum_{k \in \mathbb{Z}} \|Y_k\|^\alpha)^2} \right) \\ &\leq 3T_{1,1}^Y + T_{1,2}^Y, \end{aligned}$$

so that $\lim_{m \rightarrow \infty} T_2^Y = 0$. Since the bounds on T_1^Y and T_2^Y do not depend on A , all convergences hold uniformly in A , which proves assertion (ii). $\square$

12. Proof of Lemma A.10. Lemma A.10 of the main paper establishes bounds on the solutions to two optimization problems. For convenience, the lemma is restated here.

LEMMA 12.1. *For $m \in \mathbb{N}$ and $a = (a_1, \dots, a_m) \in [0, 1]^m$ let*

$$M(a) := \frac{\sum_{k=1}^m a_k \log^2 a_k}{1 + \sum_{k=1}^m a_k}, \quad \tilde{M}(a) := \frac{\sum_{k=1}^m a_k |\log a_k|}{1 + \sum_{k=1}^m a_k}$$

with $0 \log^i(0) := 0$ for $i \in \{1, 2\}$. Then $\sup_{a \in [0, 1]^m} M(a) = O(\log^2 m)$ and $\sup_{a \in [0, 1]^m} \tilde{M}(a) = O(\log m)$ as $m \rightarrow \infty$.

PROOF. We only establish the bound on M , as the second assertion follows by similar arguments.

First note that $x \mapsto x \log^2 x$ is decreasing on the interval $[e^{-2}, 1]$. Thus, if $a_h > e^{-2}$ for some $h \in \{1, \dots, m\}$, then replacing a_h by e^{-2} increases $M(a)$, i.e. any point of maximum must belong to $[0, e^{-2}]^m$.

Check that the partial derivative

$$\frac{\partial}{\partial a_h} M(a) = \frac{\log^2 a_h + 2 \log a_h}{1 + \sum_{k=1}^m a_k} - \frac{\sum_{k=1}^m a_k \log^2 a_k}{(1 + \sum_{k=1}^m a_k)^2}$$

tends to ∞ as $a_h \downarrow 0$. Hence all points of maximum a must be in $(0, e^{-2}]^m$ and all partial derivatives must vanish at such points. In particular, $\log^2 a_h + 2 \log a_h = \log^2 a_1 + 2 \log a_1$ for all $h \in \{2, \dots, m\}$, which is equivalent to $(\log(a_h) + 1)^2 = (\log(a_1) + 1)^2$ and thus to $a_h = a_1$ or $a_h = 1/(e^2 a_1)$. Because the second case contradicts $a_1, a_h \leq e^{-2}$, we have shown so far that all coordinates of a point of maximum must be equal.

The condition on the partial derivatives then reads as

$$\frac{\log^2 a_1 + 2 \log a_1}{1 + m a_1} - \frac{m a_1 \log^2 a_1}{(1 + m a_1)^2} = 0 \quad \Leftrightarrow \quad |\log a_1| = 2(1 + m a_1).$$

This equation has a unique solution a_1^* . It is easily seen that a_1^* is of smaller order than $m^{\delta-1}$ and of larger order than $m^{-\delta-1}$ as $m \rightarrow \infty$ for all $\delta > 0$. In particular, $|\log a_1^*| \sim \log m$ and thus

$$a_1^* \sim \frac{\log m}{2m}$$

as $m \rightarrow \infty$. The maximum of M is given by

$$M(a_1^*, \dots, a_1^*) = \frac{ma_1^* \log^2 a_1^*}{1 + ma_1^*} = 4ma_1^*(1 + ma_1^*) \sim \log^2 m$$

as $m \rightarrow \infty$. □

13. Proof of Lemma A.12. Lemma A.12 of the paper contains the limit behavior of the covariances between $(Z(A))_{A \in \mathcal{A}}$ and Z_ϕ and the variance of Z_ϕ . The proof uses similar ideas as the proof of Lemma A.4.

LEMMA 13.1. *Suppose the conditions (RV), (S), (BC), (TC), (CΘ), and (BC') are satisfied. Then*

(i)

$$\frac{1}{r_n v_n} \text{Var} \left(\sum_{t=1}^{r_n} \phi(W_{n,t}) \right) = \alpha^{-1} \sum_{k \in \mathbb{Z}} E \left[(1 \wedge \|\Theta_k\|^\alpha) (|\log(\|\Theta_k\|)| + 2\alpha^{-1}) \right],$$

(ii)

$$\begin{aligned} & \frac{1}{r_n v_n} \text{Cov} \left(\sum_{j=1}^{r_n} \phi(W_{n,j}), \sum_{t=1}^{r_n} g_A(W_{n,t}) \right) \\ &= \sum_{k \in \mathbb{Z}} E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) (1 \wedge \|\Theta_k\|^\alpha) (\log^+ \|\Theta_k\| + \alpha^{-1}) \right] \end{aligned}$$

for all $A \in \mathcal{A}$.

PROOF. Regular variation implies

$$E(\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n) \rightarrow E[\log^+(Y_0) \log^+(Y_k)]$$

$$E(\log^+ \|X_{n,0}\| \mid \|X_0\| > u_n) \rightarrow E[\log^+ Y_0]$$

for all $k \in \mathbb{Z}$ (cf. Kulik & Soulier (2020), Section 2.3.3). Therefore, by stationarity we obtain

$$\begin{aligned} & \frac{1}{r_n v_n} \text{Var} \left(\sum_{t=1}^{r_n} \phi(W_{n,t}) \right) \\ &= \frac{1}{v_n} \sum_{k=-r_n}^{r_n} \left(1 - \frac{|k|}{r_n} \right) E[\phi(W_{n,0}) \phi(W_{n,k})] - \frac{(r_n v_n)^2}{r_n v_n} (E(\log^+ \|X_{n,0}\| \mid \|X_0\| > u_n))^2 \\ &= \sum_{k=-r_n}^{r_n} \left(1 - \frac{|k|}{r_n} \right) E(\log^+(\|X_{n,0}\|) \log^+(\|X_{n,k}\|) \mid \|X_0\| > u_n) + o(1) \\ &\rightarrow \sum_{k \in \mathbb{Z}} E[\log^+(\|Y_0\|) \log^+(\|Y_k\|)], \end{aligned}$$

where the last step follows from Pratt's Lemma, which may be applied because of (BC'). Check that

$$\begin{aligned} \int_u^\infty \log(y) \alpha y^{-(\alpha+1)} dy &= u^{-\alpha} (\log u + \alpha^{-1}) \\ \int_u^\infty \log^2(y) \alpha y^{-(\alpha+1)} dy &= u^{-\alpha} (\log^2 u + 2\alpha^{-1} \log u + 2\alpha^{-2}). \end{aligned}$$

Direct calculations similar to (11.1), using $Y_k = \Theta_k \|Y_0\|$ with Θ_k and $\|Y_0\|$ independent and $P\{\|Y_0\| > y\} = y^{-\alpha} \wedge 1$ yield

$$\begin{aligned} &E[\log^+(\|Y_0\|) \log^+(\|Y_k\|)] \\ &= E[\log^+(\|Y_0\|) \log^+(\|\Theta_k\| \|Y_0\|)] \\ &= E\left[\int_{1 \vee \|\Theta_k\|^{-1}}^\infty \log(y) \log(\|\Theta_k\| y) \alpha y^{-\alpha-1} dy\right] \\ &= E\left[(1 \vee \|\Theta_k\|^{-1})^{-\alpha} (\log \|\Theta_k\| (\log(1 \vee \|\Theta_k\|^{-1}) + \alpha^{-1}) \right. \\ &\quad \left. + \log^2(1 \vee \|\Theta_k\|^{-1}) + 2\alpha^{-1} \log(1 \vee \|\Theta_k\|^{-1}) + 2\alpha^{-2})\right] \\ &= E\left[(1 \wedge \|\Theta_k\|^\alpha) (\log \|\Theta_k\| (-\log(1 \wedge \|\Theta_k\|) + \alpha^{-1}) \right. \\ &\quad \left. + \log^2(1 \wedge \|\Theta_k\|) - 2\alpha^{-1} \log(1 \wedge \|\Theta_k\|) + 2\alpha^{-2})\right] \\ &= \alpha^{-1} E\left[(1 \wedge \|\Theta_k\|^\alpha) (\log \|\Theta_k\| + 2\alpha^{-1})\right] \end{aligned}$$

for all $k \in \mathbb{Z}$, which proves assertion (i).

Next we prove

$$\begin{aligned} (13.1) \quad &E(g_A(W_{n,0}) \log^+ \|X_{n,k}\| \mid \|X_0\| > u_n) \rightarrow E[g_A(Y) \log^+ \|Y_k\|] \\ &= E\left[\sum_{h \in \mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|^\alpha} \mathbf{1}_A\left(\frac{Y_{h+i}}{\|Y_h\|}\right) \log^+ \|Y_k\|\right] \end{aligned}$$

for all $k \in \mathbb{Z}$, using similar techniques as in the proof of (8.1). In particular, by the same arguments and uniform integrability one obtains, for all $m \in \mathbb{N}, k \in \mathbb{Z}$,

$$(13.2) \quad \lim_{n \rightarrow \infty} E(g_A^{(m)}(W_{n,0}) \log^+ \|X_{n,k}\| \mid \|X_0\| > u_n) = E[g_A^{(m)}(Y) \log^+ \|Y_k\|],$$

$$(13.3) \quad \lim_{m \rightarrow \infty} E[g_A^{(m)}(Y) \log^+ \|Y_k\|] = E[g_A(Y) \log^+ \|Y_k\|].$$

Furthermore, because $g_A \in [0, 1]$, the Cauchy-Schwarz inequality yields

$$\begin{aligned} &E(|g_A^{(m)}(W_{n,0}) - g_A(W_{n,0})| \log^+ \|X_{n,k}\| \mid \|X_0\| > u_n) \\ &\leq \left[E(|g_A^{(m)}(W_{n,0}) - g_A(W_{n,0})| \mid \|X_0\| > u_n) E((\log^+ \|X_{n,k}\|)^2 \mid \|X_0\| > u_n)\right]^{1/2}. \end{aligned}$$

Because the second expectation converges to $E[(\log^+ \|Y_k\|)^2]$, (8.3) shows that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} E(|g_A^{(m)}(W_{n,0}) - g_A(W_{n,0})| \log^+ \|X_{n,k}\| \mid \|X_0\| > u_n) = 0,$$

which, in combination with (13.2) and (13.3), yields (13.1) by standard arguments.

Now, by stationarity and Pratt's lemma,

$$\begin{aligned}
& \frac{1}{r_n v_n} \text{Cov} \left(\sum_{j=1}^{r_n} \phi(W_{n,j}), \sum_{t=1}^{r_n} g_A(W_{n,t}) \right) \\
&= \sum_{k=-r_n}^{r_n} \left(1 - \frac{|k|}{r_n} \right) E \left(\log^+(\|X_{n,k}\|) g_A(W_{n,0}) \mid \|X_0\| > u_n \right) + o(1) \\
&\rightarrow \sum_{k \in \mathbb{Z}} E \left[\log^+(\|Y_k\|) \sum_{h \in \mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \right].
\end{aligned}$$

Similarly as above, we obtain

$$\begin{aligned}
& E \left[\sum_{h \in \mathbb{Z}} \frac{\|Y_h\|^\alpha}{\|Y\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{Y_{h+i}}{\|Y_h\|} \right) \log^+(\|Y_k\|) \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \log^+(\|\Theta_k\| \|Y_0\|) \right] \\
&= E \left[\int_{1 \vee \|\Theta_k\|^{-1}}^\infty \sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) \log(\|\Theta_k\| y) \alpha y^{-\alpha-1} dy \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) (1 \vee \|\Theta_k\|^{-1})^{-\alpha} (\log \|\Theta_k\| + \log(1 \vee \|\Theta_k\|^{-1}) + \alpha^{-1}) \right] \\
&= E \left[\sum_{h \in \mathbb{Z}} \frac{\|\Theta_h\|^\alpha}{\|\Theta\|_\alpha^\alpha} \mathbf{1}_A \left(\frac{\Theta_{h+i}}{\|\Theta_h\|} \right) (1 \wedge \|\Theta_k\|^\alpha) (\log^+ \|\Theta_k\| + \alpha^{-1}) \right],
\end{aligned}$$

which concludes the proof of assertion (ii). □

REFERENCES

- Davis, R. A., Drees, H., Segers, J. & Warchoř, M. (2018), ‘Inference on the tail process with application to financial time series modeling’, *Journal of Econometrics* **205**(2), 508–525.
- Davis, R. A. & Mikosch, T. (2009), Extremes of stochastic volatility models, in ‘Handbook of Financial Time Series’, Springer, pp. 355–364.
- Douc, R., Moulines, E., Priouret, P. & Soulier, P. (2018), *Markov Chains*, Springer.
- Drees, H. & Neblung, S. (2021), ‘Asymptotics for sliding blocks estimators of rare events’, *Bernoulli* . to appear.
- Kulik, R. & Soulier, P. (2020), *Heavy-Tailed Time Series*, Springer.
- Kulik, R., Soulier, P. & Wintenberger, O. (2019), ‘The tail empirical process of regularly varying functions of geometrically ergodic Markov chains’, *Stochastic Processes and their Applications* **129**(11), 4209–4238.