January 5, 2024

ON THE ASYMPTOTICS OF EXTREMAL ℓ^p-BLOCKS CLUSTER INFERENCE

GLORIA BURITICÁ AND OLIVIER WINTENBERGER

Abstract. Extremes occur in stationary regularly varying time series as short periods with several large observations, known as extremal blocks. We study cluster statistics summarizing the behavior of functions acting on these extremal blocks. Examples of cluster statistics are the extremal index, cluster size probabilities, and other cluster indices. The purpose of our work is twofold. First, we state the asymptotic normality of block estimators for cluster inference based on consecutive observations with large ℓ^p-norms, for $p > 0$. Second, we verify the conditions we require on classic models such as linear models and solutions of stochastic recurrence equations. Regarding linear models, we prove that the asymptotic variance of classical index cluster-based estimators is null as first conjectured in [26]. We illustrate our findings on simulations.

1. Introduction

We study stationary heavy-tailed time series with regularly varying distributions; cf. [5]. In this framework, extremal observations cluster: an extreme value triggers a short period with numerous large observations. This behavior is known to perturb classical inference procedures tailored for independent observations like high quantile inference; see [20]. This clustering effect can be summarized with the *extremal index*, initially introduced in [34] and [35]. We can interpret it as the inverse of the mean number of consecutive exceedances above a high threshold in a short period of time. In this article, we aim to infer such properties of the clustering effect by letting functionals act on consecutive observations with extremal behavior. For example, we can recover the extremal index from this setting and other important indices of the extremes of the series.

We consider cluster statistics of regularly varying time series $(\mathbf{X}_t)$ with values in $(\mathbb{R}^d, |\cdot|)$, and tail index $\alpha > 0$; a formal definition is conferred to Section 2.2. For cluster inference, we consider a sample $\mathbf{X}_{[1,n]}$ together with the sequence (b_n), and we define the sample of disjoint blocks $(\mathcal{B}_j)_{j=1,\dots,m_n}$ as blocks of consecutive observations with

$$\mathcal{B}_j \;:=\; (\mathbf{X}_{(j-1)b_n+1}, \dots, \mathbf{X}_{jb_n}) \;=\; \mathbf{X}_{(j-1)b_n+[1\,:\,b_n]}, \tag{1.1}$$

such that $b_n \to \infty$, $m_n = n/b_n \to \infty$, as $n \to \infty$. We follow the p-clusters theory developed in [12] for a fixed $p \in (0, \infty]$. The extremal behavior of

2020 *Mathematics Subject Classification.* Primary 60G70 ; Secondary 60F10 62G32 60F05 60G57.

regularly varying time series is modeled by the conditional behavior of a block $\mathcal{B}_j$ given that its ℓ^p-norm is large:

$$\mathbb{P}(\mathcal{B}_1/x_{b_n} \in A \,|\, \|\mathcal{B}_1\|_p > x_{b_n}) \xrightarrow{w} \mathbb{P}(Y\mathbf{Q}^{(p)} \in A), \qquad n \to \infty, \tag{1.2}$$

such that Y is independent of $\mathbf{Q}^{(p)} \in \ell^p$, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. The weak convergence holds for a family of shift-invariant continuity sets $A \subset \ell^p$, and (x_n) is a suitable sequence satisfying $\mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n}) \to 0$, as $n \to \infty$. The spectral $p-$cluster process $\mathbf{Q}^{(p)}$ models the short period's behavior under the rare event that its ℓ^p-norm is large. We study p-cluster statistics of the form

$$f_p^{\mathbf{Q}} = \mathbb{E}[f_p(Y\mathbf{Q}^{(p)})], \tag{1.3}$$

for suitable ℓ^p-continuity functions $f_p : \ell^p \to \mathbb{R}$, invariant to the shift operator. To infer the cluster statistic (1.3), we use the disjoint blocks estimators proposed in [12], and defined as

$$\widehat{f_p^{\mathbf{Q}}} := \frac{1}{k}\sum_{t=1}^{m_n} f_p(\mathcal{B}_t/\|\mathcal{B}\|_{p,(k+1)})\mathbb{1}(\|\mathcal{B}_t\|_p > \|\mathcal{B}\|_{p,(k+1)}), \tag{1.4}$$

where $\|\mathcal{B}\|_{p,(1)} \geqslant \|\mathcal{B}\|_{p,(2)} \geqslant \ldots \geqslant \|\mathcal{B}\|_{p,(m_n)}$, denotes the sequence of order statistics of the ℓ^p-norms of blocks defined in (1.1).

The main goal of this article is to establish the asymptotic normality of the block estimators in Equation (1.4), tailored for cluster inference. We state moment, mixing and bias assumptions yielding the existence of a sequence (k_n), satisfying $k = k_n \to \infty$, $m_n/k_n \to \infty$ such that

$$\sqrt{k}\big(\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}\big) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(f_p(Y\mathbf{Q}^{(p)}))), \quad n \to \infty, \tag{1.5}$$

and the limit is a centered Gaussian distribution. As a result, we see that the asymptotic variance of the blocks estimator can be computed in terms of the $p-$cluster $\mathbf{Q}^{(p)}$. In general, it is possible to obtain numerous representations of one cluster statistics in (1.3) combining the choice of p with f_p. This follows by the change-of-norms equation given in [12] that we recall in (2.12). For example, the extremal index admits distinct $p-$clusters representations in the form of (1.3) if we use $p = \infty$ or $p = \alpha$, the tail-index of the series; cf. [13]. This strategy points to new inference methodologies to estimate the same statistic. The asymptotic result in (1.5) allows us to compare the variances of these inference procedures tuned with different p.

We show that introducing ℓ^p-norm block order statistics in (1.4), instead of order statistics of the sample $(|\mathbf{X}_t|)$ as in [18, 14], can lead to a better asymptotic variance for cluster inference. We give examples of variance reduction in the case of linear models with short-range dependence, for inference of classical indices. In our examples, the asymptotic variance $\mathrm{Var}(f_p(Y\mathbf{Q}^{(p)}))$ is null because of the deterministic properties of the spectral p-cluster process of linear models. For linear models, the advantage of replacing thresholds with block maxima records was previously investigated

in [26]. Existing works [17, 18, 14, 33] following [26] focus on cluster of exceedances inference such that $p = \infty$. Our asymptotic result comforts and extends the heuristics presented in [26] for $p = \infty$ and linear models to the case $p < \infty$ and general models. To prove the asymptotic normality of block estimators, we rely on the asymptotics of disjoint block estimators studied in Theorem 2.10. in [17], and the theory of empirical processes therein. We also follow the modern overview in [33]. To handle the asymptotics of extremal ℓ^p-blocks, we build on the large deviation principles studied in [12], and appeal to the p-cluster processes theory therein.

The blocks estimator in Equation (1.4) compares favourably to state-of-the-art methodologies for cluster inference in terms of asymptotic variance. Considering inference through extremal ℓ^p-blocks with $p < \infty$ has already proven in [12] to be advantageous in terms of bias. Choosing $p = \alpha$ can be useful in practice to make tuning the parameters in the blocks estimator less susceptible to time dependencies; see [12]. We illustrate on simulations that the ℓ^α–block estimator is competitive both in terms of bias and variance for finite sample sizes. This approach performs well even when the tail index $\alpha > 0$ has to be estimated with a Hill-type procedure.

The article is organized as follows. Preliminaries on mixing coefficients, regular variation, and the p-clusters theory of stationary time series are compiled in Section 2. In Section 3 we present our main result in Theorem 3.2, stating the asymptotic normality of the block estimators introduced in Equation (1.4). We work under mixing, moment, and bias conditions on the series that we also present in Section 3. Section 4 studies examples of extremal cluster inference such as estimation of the extremal index, the cluster size probabilities, and the cluster index for sums. We conclude by verifying our conditions on classical models such as linear processes and stochastic recurrence equations in Section 5. In the case of linear models with short-range dependence, Theorem 5.6 states that the ℓ^p-block estimators of all the aforementioned quantities have null-asymptotic variance. Thereby, they are super-efficient for cluster inference of important indices as conjectured by [26] for $p = \infty$. We illustrate the finite-sample performances of our estimators in Section 6. All proofs are deferred to Section A.

1.1. **Notation.** We consider stationary time series $(\mathbf{X}_t)$ taking values in $\mathbb{R}^d$, that we endow with a norm $|\cdot|$. Let $p > 0$, and $(\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$. Define the p-modulus function $\|\cdot\|_p : (\mathbb{R}^d)^{\mathbb{Z}} \to [0, +\infty]$ as

$$\|(\mathbf{x}_t)\|_p^p \;:=\; \textstyle\sum_{t\in\mathbb{Z}} |\mathbf{x}_t|^p \,,$$

and define the sequential space ℓ^p as

$$\ell^p \;:=\; \{(\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}} : \|(\mathbf{x}_t)\|_p^p < +\infty\} \,,$$

with the convention that, for $p = \infty$, the space ℓ^∞ refers to sequences with finite supremum norm. For any $p \in (0, +\infty]$, the p-modulus functions induce a distance d_p in ℓ^p, and for $p \in [1, +\infty)$, it defines a norm. Abusing notation, we call them ℓ^p-norms for $p \in (0, +\infty]$. Let $\tilde{\ell}^p = \ell^p/\sim$ be the shift-invariant

quotient space where: $(\mathbf{x}_t) \sim (\mathbf{y}_t)$ if and only if there exists $k \in \mathbb{Z}$ such that $\mathbf{x}_{t-k} = \mathbf{y}_t$, $t \in \mathbb{Z}$. We also consider the metric space $(\tilde{\ell}^p, \tilde{d}_p)$ such that for $[\mathbf{x}], [\mathbf{y}] \in \tilde{\ell}^p$,

$$\tilde{d}_p([\mathbf{x}], [\mathbf{y}]) \quad = \quad \inf_{k\in\mathbb{Z}} \{d_p(\mathbf{x}_{t-k}, \mathbf{y}_t), (\mathbf{x}_t) \in [\mathbf{x}], (\mathbf{y}_t) \in [\mathbf{y}]\},$$

and without loss of generality, we write an element $[\mathbf{x}]$ in $\tilde{\ell}^p$ also as $(\mathbf{x}_t)$. Further details on the shift-invariant spaces are deferred to [12, 4].

The operator norm for $d \times d$ matrices, $\mathbf{A} \in \mathbb{R}^{d\times d}$, is defined as $|\mathbf{A}|_{op} := \sup_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|$. The truncation operations of $(\mathbf{x}_t)$ at the level ϵ, for $\epsilon > 0$, are defined by

$$(\overline{\mathbf{x}}^{\varepsilon}_t) \quad = \quad (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|\leqslant\varepsilon}), \qquad\qquad (\underline{\mathbf{x}}_{t_\epsilon}) \quad = \quad (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|>\varepsilon}).$$

The notation $a \wedge b$ denotes the minimum between two constants $a, b \in \mathbb{Z}$, and $a \vee b$ denotes its maximum. We write $\log^+(x) = \log(x) \vee 0$, for $x \in (0, \infty)$. We sometimes write $\mathbf{x}$ for the sequence $\mathbf{x} := (\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$. Furthermore, for $a, b, \in \mathbb{R}$, and $a \leqslant b$, we write as $\mathbf{x}_{[a,b]}$ the vector $(\mathbf{x}_t)_{t=a,\cdots,b}$ taking values in $(\mathbb{R}^d)^{b-a+1}$. We sometimes write $\mathbf{x}_{[a,b]} \in \tilde{\ell}^p$, which means we take the natural embedding of $\mathbf{x}_{[a,b]}$ in $\tilde{\ell}^p$ defined by assigning zeros to undefined coefficients. It will be convenient to write $\mathcal{G}_+(\tilde{\ell}^p)$ for the continuous non-negative functions on $(\tilde{\ell}^p, \tilde{d}_p)$ which vanish in a neighborhood of the origin.

2. Preliminaries

2.1. **Mixing coefficients.** Let $(\mathbf{X}_t)$ be an $\mathbb{R}^d$-valued strictly stationary time series defined over a probability space $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P})$. The properties of stationary sequences are usually studied through mixing coefficients. Denote the past and future $\sigma-$algebras by

$$\mathcal{F}_{t\leqslant 0} \quad := \quad \sigma((\mathbf{X}_t)_{t\leqslant 0}), \qquad \mathcal{F}_{t\geqslant h} \quad := \quad \sigma((\mathbf{X}_t)_{t\geqslant h}), \qquad h \geqslant 1,$$

respectively. We recall the definition of mixing coefficients $(\rho_h), (\beta_h)$, below

$$\rho_h \quad = \quad \sup_{f\in L^2(\mathcal{F}_{t\leqslant 0}), g\in L^2(\mathcal{F}_{t\geqslant h})} |\mathrm{Corr}(f, g)|,$$

$$\beta_h \quad = \quad d_{TV}\big(\mathbb{P}_{\mathcal{F}_{t\leqslant 0}\otimes\mathcal{F}_{t\geqslant h}}, \mathbb{P}_{\mathcal{F}_{t\leqslant 0}} \otimes \mathbb{P}_{\mathcal{F}_{t\geqslant h}}\big),$$

where $d_{TV}(\cdot, \cdot)$ is the total variation distance between two probability measures: $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P}_1)$, $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P}_2)$, and $\mathbb{P}_1 \otimes \mathbb{P}_2(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$, for $A, B \in \mathcal{A}$. For a summary on mixing conditions see [8, 16, 42].

Remark 2.1. The $\rho-$mixing coefficients $(\rho_t)_{t\geqslant 0}$ were introduced in [32], and popularized due to the Ibragimov central limit theorem for dependent stationary sequences in [28]. This theorem states that sufficient conditions for the central limit of stationary sequences to hold are the mixing condition: $\sum_{t=1}^{\infty} \rho_{2^t} < \infty$, together with a moment assumption of order $\kappa > 2$. The aforementioned mixing condition was studied in detail in [7, 40, 46, 48]; see [8] for a review. However, aside from the m_0-dependent case and the

Gaussian case (where $\rho_t = 0$ for $t > m_0$ and $\rho_t \leqslant \pi\beta_t$, respectively) there is no general recipe for computing ρ–mixing rates.

Remark 2.2. A detailed interpretation of the $\beta-$mixing coefficients (β_h) in terms of the total variation distance can be found in Chapter 1.2 in [16]. These mixing coefficients are well adapted while working with Markov processes. Indeed, a strictly stationary Harris recurrent Markov chain $(\mathbf{X}_t)$, satisfies $\beta_t \to 0$ as $t \to \infty$; see Theorem 3.5 in [8].

2.2. **Regular variation.** We consider stationary time series $(\mathbf{X}_t)$ taking values in $(\mathbb{R}^d, |\cdot|)$ and that it is regularly varying with tail index $\alpha > 0$: all its finite-dimensional vectors are multivariate regularly varying of the same index. In this case we write $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$. Borrowing the ideas in [5], $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$ if and only if, for all $h \geqslant 0$, there exists a vector $(\boldsymbol{\Theta}_t)_{|t|\leqslant h}$, taking values in $(\mathbb{R}^d)^{2h+1}$ such that

$$\mathbb{P}(x^{-1}(\mathbf{X}_t)_{|t|\leqslant h} \in \cdot \,|\, |\mathbf{X}_0| > x) \quad \xrightarrow{d} \quad \mathbb{P}(Y(\boldsymbol{\Theta}_t)_{|t|\leqslant h} \in \cdot), \quad x \to +\infty, \tag{2.6}$$

where Y is independent of $(\boldsymbol{\Theta}_t)_{|t|\leqslant h}$ and $\mathbb{P}(Y > y) = y^{-\alpha}, y > 1$. We call the sequence $(\boldsymbol{\Theta}_t)$, taking values in $(\mathbb{R}^d)^{\mathbb{Z}}$, the spectral tail process.

The time series $(\boldsymbol{\Theta}_t)$ does not inherit the stationarity property of the series. Instead, the time-change formula of [5] holds: for any $s, t \in \mathbb{Z}, s \leqslant 0 \leqslant t$ and for any measurable bounded function $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$,

$$\mathbb{E}[f(\boldsymbol{\Theta}_{s-i}, \ldots, \boldsymbol{\Theta}_{t-i})\mathbb{1}(|\boldsymbol{\Theta}_{-i}| \neq 0)] = \mathbb{E}[|\boldsymbol{\Theta}_i|^\alpha f(\boldsymbol{\Theta}_s/|\boldsymbol{\Theta}_i|, \ldots, \boldsymbol{\Theta}_t/|\boldsymbol{\Theta}_i|)]. \tag{2.7}$$

2.3. **p-cluster processes.** Let $(\mathbf{X}_t)$ be a stationary time series satisfying $\mathbf{RV}_\alpha$. For $p > 0$, we say the series admits a p-cluster process $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$ if there exists a well-chosen sequence (x_n), satisfying

$$\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_p > x_n) \quad \sim \quad n\, c(p) \mathbb{P}(|\mathbf{X}_1| > x_n), \quad n \to \infty \tag{2.8}$$

with $c(p) \in (0, \infty)$, $n\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$, and

$$\mathbb{P}(\mathbf{X}_{[1,n]}/x_n \in \cdot \,|\, \|\mathbf{X}_{[1,n]}\|_p > x_n) \quad \xrightarrow{w} \quad \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot), \quad n \to \infty, \tag{2.9}$$

where Y is independent of $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., and the limit in (2.9) holds in $(\tilde{\ell}^p, \tilde{d}_p)$. We study below the anti-clustering and vanishing-small values conditions noted $\mathbf{AC}$, $\mathbf{CS}_p$, respectively, which guarantee the existence of $p-$clusters. We rephrase next the Theorem 2.1. of [12].

Proposition 2.3. *Let $(\mathbf{X}_t)$ be a stationary time series satisfying $\mathbf{RV}_\alpha$. Let (x_n) be a sequence such that $n\,\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$, as $n \to \infty$ and $p > 0$. For all $\delta > 0$, assume*

$$\mathbf{AC}: \ \lim_{s\to\infty} \limsup_{n\to\infty} \mathbb{P}(\|\mathbf{X}_{[s,n]}\|_\infty > \epsilon\, x_n \,|\, |\mathbf{X}_1| > \epsilon\, x_n) = 0, \qquad \epsilon > 0,$$

$$\mathbf{CS}_p: \ \lim_{\epsilon\downarrow 0} \limsup_{n\to\infty} \frac{\mathbb{P}\left(\|\overline{\mathbf{X}_{[1,n]}/x_n}^{\epsilon}\|_p^p > \delta\right)}{n\mathbb{P}(|\mathbf{X}_1| > x_n)} = 0.$$

Then, if $p \geqslant \alpha$, Equation (2.8) holds with $c(\infty) \leqslant c(p) \leqslant c(\alpha) = 1$, and $(\mathbf{X}_t)$ admits a $p-$cluster process $\mathbf{Q}^{(p)}$ in the sense of (2.9). If $p < \alpha$, existence of the $p-$cluster process holds if $\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] < \infty$. In this case, Equation (2.8) holds with $c(p) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha]$.

From Proposition 2.3 we see that assuming **AC**, and $\mathbf{CS}_\alpha$, the time series $(\mathbf{X}_t)$ admits an α-cluster $\mathbf{Q}^{(\alpha)}$, where $\alpha > 0$, denotes the tail index. In this case, appealing to Proposition 3.1. in [12], we have

$$\mathbf{Q}^{(\alpha)} \stackrel{d}{=} \boldsymbol{\Theta}/\|\boldsymbol{\Theta}\|_\alpha, \quad \in \tilde{\ell}^\alpha\,, \tag{2.10}$$

where $(\boldsymbol{\Theta}_t)$ is the spectral tail process from Equation (2.6). Moreover, if $\mathbf{CS}_p$, $\mathbf{CS}_{p'}$, and $\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] + \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_{p'}^\alpha] < \infty$ also hold, then the p, p'-clusters exist and are related by the change-of-norms formula below

$$\begin{aligned} &\mathbb{P}(\mathbf{Q}^{(p)} \in \cdot) \\ &= c(p)^{-1}\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha \mathbb{1}(\mathbf{Q}^{(\alpha)}/\|\mathbf{Q}^{(\alpha)}\|_p \in \cdot)] \qquad (2.11)\\ &= \frac{c(p')}{c(p)}\mathbb{E}[\|\mathbf{Q}^{(p')}\|_p^\alpha \mathbb{1}(\mathbf{Q}^{(p')}/\|\mathbf{Q}^{(p')}\|_p \in \cdot)]. \qquad (2.12)\end{aligned}$$

Since $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. for any $p > 0$, then $c(\alpha) = 1$, and $\mathbb{E}[\|\mathbf{Q}^{(p')}\|_p^\alpha] = c(p)/c(p')$, where $c(p), c(p')$, are as in Equation (2.8).

Remark 2.4. *We can check readily that if $\sum_{t=1}^\infty \rho_t < \infty$, then* **AC** *holds for all sequences of levels (x_n) satisfying $n\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$, as $n \to \infty$.*

Remark 2.5. *Using the monotonicty of norms, we see straightforwardly that $\mathbf{CS}_p$ implies the condition $\mathbf{CS}_{p'}$, for $p' > p > 0$. If $p > \alpha$, where α is the tail index, the condition $\mathbf{CS}_p$ is always satisfied for sequences (x_n) such that $n\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$, as $n \to \infty$. In the case $\alpha/2 < p \leqslant \alpha$, if $\sum_{t=0}^\infty \rho_t < \infty$, then condition $\mathbf{CS}_p$ holds for sequences (x_n) such that there exists $\kappa > 0$, satisfying $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, as $n \to \infty$. This follows by Remarks 5.1. and 5.2. in* [12].

3. Asymptotic normality

3.1. **Main result.** Let $(\mathbf{X}_t)$ be an $\mathbb{R}^d$-valued stationary time series satisfying $\mathbf{RV}_\alpha$. Assume the conditions of Proposition 2.3 hold for $p > 0$, the series admits a p-cluster process $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$, and (2.9) holds for a sequence of high levels (x_n) satisfying $\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_p > x_n) \to 0$. Recall the block estimator in (1.4) is tuned with the block lengths (b_n), and the number (k_n) of extremal blocks. The total number of disjoint blocks in a sample is denoted (m_n) with $m_n = \lfloor n/b_n \rfloor$. We assume the relation between (k_n) and (b_n)

$$\begin{aligned} k \quad := \quad k_n \quad &= \quad \left\lfloor m_n \mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n}) \right\rfloor \\ &\sim \quad n\,c(p)\mathbb{P}(|\mathbf{X}_1| > x_{b_n}), \quad n \to \infty, \end{aligned} \tag{3.13}$$

holds, where $c(p) \in (0,\infty)$ are as in (2.8). In what follows, if the sequences (x_n), (b_n), (k_n), (m_n), appear then they coincide with the ones mentioned here.

Our main result is presented in this section in Theorem 3.2. It establishes the asymptotic normality of the block estimator in (1.4) under the moment, bias, and mixing assumptions that we introduce below.

L: Let $f_p : \tilde{\ell}^p \to \mathbb{R}$, be such that $f_p \in \mathcal{G}_+(\tilde{\ell}^p)$. Moreover assume $u \mapsto f_p((\mathbf{x}_t)/u)$ is a non-increasing function, and assume there exists $\delta > 0$ such that for all $u > 0$,

$$\frac{m_n}{k_n}\mathbb{E}[f_p(\mathcal{B}_1/u\, x_{b_n})^{2+\delta}] \;=\; O(1), \tag{3.14}$$

as $n \to \infty$, and $\mathbb{E}[f_p(Y\mathbf{Q}^{(p)})^2] < \infty$.

B: Let $f_p : \tilde{\ell}^p \to \mathbb{R}$ satisfy **L**. Fix $\epsilon > 0$, and assume the bias conditions

$$0 \;=\; \lim_{n\to\infty}\sqrt{k}\sup_{u\in[1-\varepsilon,1+\varepsilon]}\Big|\frac{\mathbb{E}[f_p(\mathcal{B}_1/u\, x_{b_n})\mathbb{1}(\|\mathcal{B}_1/x_{b_n}\|_p > u)]}{\mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n})} - u^{-\alpha}\, f_p^{\mathbf{Q}}\Big|, \tag{3.15}$$

$$0 \;=\; \lim_{n\to\infty}\sqrt{k}\sup_{u\in[1-\varepsilon,1+\varepsilon]}\Big|\frac{\mathbb{P}(\|\mathcal{B}_1/x_{b_n}\|_p > u)}{\mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n})} - u^{-\alpha}\Big|, \tag{3.16}$$

where $f_p^{\mathbf{Q}}$ is as in (1.3).

MX: Assume the mixing coefficients (β_t) satisfy the condition

$$\lim_{n\to\infty} m_n\beta_{b_n} \;=\; 0. \tag{3.17}$$

MX$_\beta$: Let $f_p : \tilde{\ell}^p \to \mathbb{R}$ satisfy **L**, and let $\delta > 0$ be such that (3.14) hold. Assume there exists a sequences (ℓ_n), satisfying $\ell_n \to \infty$, and the mixing coefficients (β_t) satisfy $m_n\beta_{\ell_n}/k_n \to 0$, $\ell_n/b_n \to 0$, as $n \to \infty$, and

$$\lim_{n\to\infty}\textstyle\sum_{t=1}^{m_n}(m_n\beta_{tb_n}/k_n)^{\frac{\delta}{2+\delta}} \;=\; 0. \tag{3.18}$$

If f_p is bounded, we assume $\sum_{t=1}^{m_n} m_n\beta_{tb_n}/k_n \to 0$ instead of (3.18).

MX$_\rho$: Assume the correlation coefficients (ρ_t) satisfy $\sum_{t=1}^{m_n}\rho_{tb_n} \to 0$ and

$$\lim_{s\to\infty}\lim_{b_n\to\infty}\textstyle\sum_{t=s}^{b_n}\mathbb{P}(|\mathbf{X}_t| > \epsilon\, x_{b_n}\,|\,|\mathbf{X}_0| > \epsilon\, x_{b_n}) \;=\; 0. \tag{3.19}$$

We state in Theorem 3.2 below our main result on the asymptotic normality of the blocks estimator. We defer its proof to Section A.

Theorem 3.1. *Let $(\mathbf{X}_t)$ be a stationary time series satisfying $\mathbf{RV}_\alpha$. Assume the conditions of Proposition 2.3 hold, such that the series admits a p-cluster*

process $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$. *Consider* $f_p : \tilde{\ell}^p \to \mathbb{R}$ *such that* **L** *and* **B** *hold. Assume also* **MX** *holds, and assume either* $\mathbf{MX}_\beta$ *or* $\mathbf{MX}_\rho$ *hold for the same sequence* (k_n). *Then we have*

$$\sqrt{k}\big(\,\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}\,\big) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mathrm{Var}(f_p(Y\mathbf{Q}^{(p)}))\,), \quad n \to \infty. \tag{3.20}$$

Here we follow the notation from equations (1.3) *and* (1.4).

Theorem 3.2. *Assume in addition that for* $q \in [\alpha - \epsilon, \alpha + \epsilon]$,

$$\mathbb{E}\Big[\|\mathbf{Q}^{(\alpha)}\|_q^{2\alpha} \sum_{t\in\mathbb{Z}} \frac{\|\mathbf{Q}_t^{(\alpha)}|^{q\wedge\alpha}}{\|\mathbf{Q}^{(\alpha)}\|_q^{q\wedge\alpha}} \log\Big(\frac{\|\mathbf{Q}^{(\alpha)}\|_q}{|\mathbf{Q}_t^{(\alpha)}|}\Big)\Big] < \infty.$$

Then, if $\widehat{\alpha}$ *is the Hill estimator of the series tuned with* k_h *higher order statistics of the sample* $(|\mathbf{X}_t|)$, *then*

$$\sqrt{k}\big(\,\widehat{f_{\widehat{\alpha}}^{\mathbf{Q}}} - f_\alpha^{\mathbf{Q}}\,\big) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mathrm{Var}(f_\alpha(Y\mathbf{Q}^{(\alpha)})) + c_h\sigma_\alpha^2\,), \quad n \to \infty, \tag{3.21}$$

where $\sqrt{k/k_h} \to c_h$, *and* $\sigma_\alpha^2 > 0$. *Here we follow the notation from equations* (1.3) *and* (1.4).

Remark 3.3. *The proof of Theorem 3.2 follows the functional central limit theorem stated in Theorem C.4.5 in* [33]. *Condition* **L** *restricts the family of functions we can consider for inference. Equation* (3.14) *entails a Lindeberg-type condition holds.*

4. Cluster statistics

In view of Theorem 3.2, we derive asymptotic normality of classical cluster index estimators in extreme value theory.

4.1. The extremal index.

Let $(\mathbf{X}_t)$ be a stationary time series in $(\mathbb{R}^d, |\cdot|)$ satisfying $\mathbf{RV}_\alpha$. The extremal index $\theta_{|\mathbf{X}|}$ of the series $(|\mathbf{X}_t|)$ is a measure of serial clustering introduced in [34] and [35]. We recall the extremal index estimator proposed in [13], based on extremal $\ell^\alpha-$blocks.

Corollary 4.1. *Consider* $f_p : \tilde{\ell}^p \to \mathbb{R}$ *to be the function* $\mathbf{x} \mapsto \|\mathbf{x}\|_\infty^\alpha / \|\mathbf{x}\|_\alpha^\alpha$. *Assume the conditions of Theorem 3.2 hold for* $p = \alpha$. *Let* $\theta_{|\mathbf{X}|} = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha]$, *hence we deduce an estimator*

$$\widehat{\theta}_{|\mathbf{X}|} \quad = \quad \frac{1}{k}\sum_{t=1}^{m} \frac{\|\mathcal{B}_t\|_\infty^\alpha}{\|\mathcal{B}_t\|_\alpha^\alpha} \mathbb{1}(\|\mathcal{B}_t\|_\alpha > \|\mathcal{B}\|_{\alpha,(k+1)}), \tag{4.22}$$

such that

$$\sqrt{k}\big(\,\widehat{\theta}_{|\mathbf{X}|} - \theta_{|\mathbf{X}|}\,\big) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mathrm{Var}(\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha)), \quad n \to \infty\,.$$

Proof. The proof of Corollary 4.1 follows directly as $f_p \in \mathcal{G}_+(\tilde{\ell}^p)$ is a bounded continuous function satisfying **L**. □

For comparison we review the blocks estimator based on extremal $\ell^\infty-$blocks proposed in [25]:

$$\widehat{\theta}^B_{|\mathbf{X}|} \;=\; \frac{1}{k}\sum_{t=1}^{m_n} \mathbb{1}(\|\mathcal{B}_t\|_\infty > |\mathbf{X}|_{(k+1)})\,. \tag{4.23}$$

Direct computations from Example 10.4.2 in [33] yield

$$\sqrt{k}(\widehat{\theta}^B_{|\mathbf{X}|} - \theta_{|\mathbf{X}|}) \;\xrightarrow{d}\; \mathcal{N}(0,\sigma^2_\theta), \quad n\to\infty,$$

where $\sigma^2_\theta \in [0,+\infty)$, and

$$\begin{aligned}\sigma^2_\theta &= \theta^2_{|\mathbf{X}|}\sum_{j\in\mathbb{Z}}\mathbb{E}[|\mathbf{\Theta}_j|^\alpha\wedge 1] - \theta_{|\mathbf{X}|}\\ &= \theta^2_{|\mathbf{X}|}\sum_{j\in\mathbb{Z}}\sum_{t\in\mathbb{Z}}\mathbb{E}[|\mathbf{Q}^{(\alpha)}_{j+t}|^\alpha\wedge|\mathbf{Q}^{(\alpha)}_t|^\alpha] - \theta_{|\mathbf{X}|}. \end{aligned}\tag{4.24}$$

The last equality follows appealing to the time-change formula in (2.7) and Equation (2.10). As a result, we can compare the asymptotic variances of $\widehat{\theta}_{|\mathbf{X}|}$ and $\widehat{\theta}^B_{|\mathbf{X}|}$ in the cases where $\mathbf{Q}^{(\alpha)}$ is known. This is the topic of Section 5.

4.2. **The cluster index for sums.** Let $(\mathbf{X}_t)$ be a stationary time series with values in $(\mathbb{R}^d,|\cdot|)$ satisfying $\mathbf{RV}_\alpha$. We recall [39] coined the constant $c(1)$ in (2.8) as the cluster index for sums. We review a cluster-based estimator of it, introduced in [12], based on extremal ℓ^α-blocks.

Corollary 4.2. *Consider $f_p:\tilde{\ell}^p\to\mathbb{R}$ to be the function $\mathbf{x}\mapsto\|\mathbf{x}\|_1^\alpha/\|\mathbf{x}\|_\alpha^\alpha$. Assume the conditions of Theorem 3.2 hold for $p=\alpha\wedge 1$, and $\alpha<2$. Let $c(1)=\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_1^\alpha]<\infty$, hence one deduces an estimator*

$$\widehat{c}(1) \;=\; \frac{1}{k}\sum_{t=1}^{m}\frac{\|\mathcal{B}_t\|_1^\alpha}{\|\mathcal{B}_t\|_\alpha^\alpha}\mathbb{1}(\|\mathcal{B}_t\|_\alpha > \|\mathcal{B}\|_{\alpha,(k+1)}), \tag{4.25}$$

such that

$$\sqrt{k}\big(\,\widehat{c}(1)-c(1)\,\big) \;\xrightarrow{d}\; \mathcal{N}(0,\mathrm{Var}(\|\mathbf{Q}^{(\alpha)}\|_1^\alpha))\,, \quad n\to\infty,$$

and $c(1)$ is as in (2.8) with $p=1$.

Proof. The proof of Corollary 4.2 follows directly from Theorem 3.2 as $f_p\in\mathcal{G}_+(\tilde{\ell}^p)$ is a bounded continuous function satisfying $\mathbf{L}$. □

Another sums index cluster-based estimator we can consider is the one proposed in [33] based on extremal ℓ^∞-blocks:

$$\widehat{c}^B(1) \;=\; \frac{1}{kb_n}\sum_{t=1}^{m_n}\mathbb{1}(\|\mathcal{B}_t\|_1 > |\mathbf{X}|_{(k+1)}). \tag{4.26}$$

Then, relying on Example 10.4.2 in [33],

$$\sqrt{k}(\widehat{c}^B - c(1)) \;\xrightarrow{d}\; \mathcal{N}(0,\sigma^2_{c(1)}), \quad n\to\infty.$$

for a constant $\sigma^2_{c(1)} \in [0, +\infty)$ defined by

$$\sigma^2_{c(1)} \;=\; c(1)^2 \sum_{j\in\mathbb{Z}} \sum_{t\in\mathbb{Z}} \mathbb{E}[|\mathbf{Q}^{(\alpha)}_{j+t}|^\alpha \wedge |\mathbf{Q}^{(\alpha)}_t|^\alpha] - c(1). \tag{4.27}$$

Similarly as in Example 4.1, whenever $\mathbf{Q}^{(\alpha)}$ is known, we can compare directly the asymptotic variances relative to the estimators $\widehat{c}(1)$ and $\widehat{c}^{\,B}(1)$. Section 5 covers this topic for classical models where the cluster process is known.

4.3. **The cluster sizes.** In general, a classical approach to model serial exceedances is using point processes as in [35] and [26]. For the levels (a_n), satisfying $n\mathbb{P}(|X_1| > a_n) \to 1$, as $n \to \infty$, and for every fixed $x > 0$ consider the point process of exceedances with state space $(0, 1]$:

$$\eta_{n,x}(\cdot) \;:=\; \overline{N}_n\big(\{\mathbf{y} : |\mathbf{y}| > x\} \times \cdot\big) \;=\; \sum_{i=1}^{n} \varepsilon_{i/n}(\cdot)\, \mathbb{1}(|\mathbf{X}_i| > x\, a_n)\,.$$

Under mixing and anti-clustering conditions, for fixed $x > 0$, we can express the limiting point process in [26] such as

$$\begin{aligned} \eta_{n,x}(\cdot) \xrightarrow{d} \eta_x(\cdot) \;&:=\; \overline{N}\big(\{\mathbf{y} : |\mathbf{y}| > x\} \times \cdot\big) \\ &=\; \sum_{i=1}^{\infty} \sum_{j\in\mathbb{Z}} \mathbb{1}\big(\Gamma_i^{-1/\alpha} |\mathbf{Q}^{(\alpha)}_{ji}| > x\big)\, \varepsilon_{U_i}(\cdot)\,, \end{aligned}$$

where the points (U_i) are iid uniformly distributed on $(0, 1)$, (Γ_i) are the points of a standard homogeneous Poisson process, and $(\mathbf{Q}^{(\alpha)}_{\cdot i})$ are iid copies of the cluster process $\mathbf{Q}^{(\alpha)}$. Using the independence among these three processes, one can easily rewrite the limit as

$$\eta_x((0, t]) \;=\; \sum_{i=1}^{N_x(t)} \xi_i\,, \qquad 0 < t \leqslant 1\,, \tag{4.28}$$

where

- N_x is a homogeneous Poisson process on $(0, 1]$ with intensity $x^{-\alpha}$,
- for an iid sequence (Y_i) of Pareto(α)-distributed random variables which is also independent of $(\mathbf{Q}^{(\alpha)}_i)$,
$$\xi_i \;=\; \sum_{j\in\mathbb{Z}} \mathbb{1}(Y_i\, |\mathbf{Q}^{(\alpha)}_{ij}| > 1)\,,$$
- N_x, (ξ_i) are independent.

Relying on the point process of exceedances representation in (4.28), the random variables (ξ_i) can be interpreted as counts of serial exceedances from one cluster. Furthermore, we deduce the relation $\mathbb{P}(\xi_1 > 0) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|^\alpha_\infty] = \theta_{|\mathbf{X}|}$, and also get an expression for the cluster size probabilities

$$\mathbb{P}(\xi_1 = j) \;=\; \mathbb{E}[|\mathbf{Q}^{(\alpha)}|^\alpha_{(j)} - |\mathbf{Q}^{(\alpha)}|^\alpha_{(j+1)}] = \;\; \pi_j\,, \qquad j \geqslant 1\,. \tag{4.29}$$

The statistic π_j can be understood as the probability of recording a cluster of length j. The blocks estimator provide natural estimators

$$\widehat{\pi}_j \;=\; \frac{1}{k}\sum_{t=1}^{m}\frac{|\mathcal{B}_t|_{(j)}^{\alpha}-|\mathcal{B}_t|_{(j+1)}^{\alpha}}{\|\mathcal{B}_t\|_{\alpha}^{\alpha}}\mathbb{1}(\|\mathcal{B}_t\|_{\alpha}>\|\mathcal{B}\|_{\alpha,(k+1)}), \tag{4.30}$$

$|\mathcal{B}_t|_{(1)}\geqslant|\mathcal{B}_t|_{(2)}\geqslant\ldots\geqslant|\mathcal{B}_t|_{(b)}$ are the order statistics of $\mathcal{B}_t$, the t-th block.

Corollary 4.3. *Consider the function $\pi_j^{\mathbf{Q}}:\tilde{\ell}^p\to\mathbb{R}$ defined by $\pi_j^{\mathbf{Q}}(\mathbf{x}):=(|\mathbf{x}|_{(j)}^{\alpha}-|\mathbf{x}|_{(j+1)}^{\alpha})/\|\mathbf{x}\|_{\alpha}^{\alpha}$, where $|\mathbf{x}|_{(1)}\geqslant|\mathbf{x}|_{(2)}\geqslant\ldots$. Assume the conditions of Theorem 3.2 hold for $p=\alpha$. Then, for all $j\geqslant 1$ we have*

$$\sqrt{k}\big(\widehat{\pi}_j-\pi_j\big)\;\xrightarrow{d}\;\mathcal{N}(0,\mathrm{Var}(\pi_j^{\mathbf{Q}}(\mathbf{Q}^{(\alpha)})))\,,\quad n\to\infty\,. \tag{4.31}$$

Corollary 4.3 provides a novel procedure for estimating cluster size probabilities based on extremal $\ell^{\alpha}-$blocks. As in the previous examples, the asymptotic variance can be computed as long as $\mathbf{Q}^{(\alpha)}$ is known. This allows for comparison with the other cluster-based inference procedures provided in [25, 21, 45]. One advantage of our methodology is that we can straightforwardly infer the asymptotic variances of cluster sizes since we express them as cluster statistics in (4.31). Moreover, inference through extremal ℓ^{α}-blocks has already proven to be useful in [12] for fine-tuning the hyperparameters of the estimators, see also the discussion in Section 6.

5. Models

5.1. **Linear m_0–dependent sequences.** We consider $(\mathbf{X}_t)$ to be a m_0–dependent time series with values in $(\mathbb{R}^d,|\cdot|)$ satisfying $\mathbf{RV}_{\alpha}$.

Example 5.1. *The time series $(\mathbf{X}_t)$ is a linear moving average of order $m_0\geqslant 1$ if it satisfies*

$$\mathbf{X}_t \;:=\; \mathbf{Z}_t+\varphi_1\mathbf{Z}_{t-1}+\cdots+\varphi_{m_0}\mathbf{Z}_{t-m_0},\quad t\in\mathbb{Z}, \tag{5.32}$$

with $\mathbb{R}^d$-variate iid innovations $(\mathbf{Z}_t)$ satisfying $\mathbf{RV}_{\alpha}$, and $(\varphi_j)\in\mathbb{R}^{m_0}$.

Alternatively, the max-moving average of order $m_0\geqslant 1$ satisfies

$$X_t \;:=\; \max\{Z_t,\varphi_1 Z_{t-1},\ldots,\varphi_{m_0}Z_{t-m_0}\},\quad t\in\mathbb{Z}, \tag{5.33}$$

with $\mathbb{R}_+$-variate iid innovations (Z_t) satisfying $\mathbf{RV}_{\alpha}$, and $(\varphi_j)\in\mathbb{R}_+^{m_0}$.

Then both moving averages satisfy $\mathbf{RV}_{\alpha}$ with $|\mathbf{Q}^{(\alpha)}|$ admitting the same deterministic expression $(|\varphi_t|/\|(\varphi_j)\|_{\alpha})$ in $\tilde{\ell}^{\alpha}$, see for instance Proposition 3.1. in [12] and Chapter 5 of [33].

Let $p>\alpha/2$. For all $\kappa>0$, a sequence satisfying $x_n=O(b_n^{\kappa+1/(p\wedge\alpha)})$ verifies **AC** and $\mathbf{CS}_p$. This is a consequence of Remark 2.5. Choosing (x_n) in this way implies there exist $\kappa'>0$, and (k_n) satisfying

$$k_n \;=\; O(n\,b_n^{-\kappa'-\alpha/(p\wedge\alpha)}), \tag{5.34}$$

such that Equation (3.13) holds from an application of Potter's bound. Since κ can be chosen arbitrarily small, κ' can also be chosen arbitrarily close to

zero. Keeping this in mind, we can state the Proposition below. The proof is postponed to Section B.

Proposition 5.2. *Consider $(\mathbf{X}_t)$ to be a m_0–dependent time series with values in $(\mathbb{R}^d, |\cdot|)$. Consider $p > \alpha/2$, and a sequence (k_n) satisfying (5.34), such that $k = k_n \to \infty$, $m_n/k_n \to \infty$. Consider $f_p : \tilde{\ell}^p \to \mathbb{R}$, and assume* **L** *and* **B** *hold. Then,*

$$\sqrt{k}(\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mathrm{Var}(\, f_p(Y\mathbf{Q}^{(p)})\,)), \quad n \to \infty.$$

In particular, the α-cluster based estimators from Section 4 in (4.22) (4.25), and (4.30), are asymptotically normally distributed. Their asymptotic variances are null in the case of the moving averages of Example 5.1.

5.2. **Linear processes.** In this section we consider stationary linear processes $(\mathbf{X}_t)$ with values in $(\mathbb{R}^d, |\cdot|)$ satisfying $\mathbf{RV}_\alpha$.

Example 5.3. *Consider $(\mathbf{X}_t)$ to be an $\mathbb{R}^d-$variate sequence satisfying*

$$\mathbf{X}_t \quad = \quad \sum_{t\in\mathbb{Z}} \varphi_j \mathbf{Z}_{t-j}, \quad t \in \mathbb{Z}, \tag{5.35}$$

for a sequence of iid innovations $(\mathbf{Z}_t)$ satisfying $\mathbf{RV}_\alpha$, and a sequence (φ_j) in $\mathbb{R}^{\mathbb{Z}}$. Moreover, assume there exists $\kappa > 0$ such that $\|(\varphi_j)\|_{(\alpha-\kappa)\wedge 2} < \infty$.

In the setting of Example 5.3, a stationary solution $(\mathbf{X}_t)$ exists and satisfies $\mathbf{RV}_\alpha$ (c.f. [15, 37]). Proposition 5.4 below demonstrates conditions **AC, CS**$_p$ hold for $p > \alpha/2$, and a suitable sequence (x_n) such that $n\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$ as $n \to \infty$. Therefore, the time series $(\mathbf{X}_t)$ admits an $\alpha-$cluster process $\mathbf{Q}^{(\alpha)}$, which we can compute in terms of the filter (φ_j), and the spectral measure of the random variable $\mathbf{Z}_0$, denoted by $\boldsymbol{\Theta}_0^{\mathbf{Z}}$, with $|\boldsymbol{\Theta}_0^{\mathbf{Z}}| = 1$ a.s. We obtain the expression, cf. Chapter 5 of [33],

$$\mathbf{Q}^{(\alpha)} \quad \stackrel{d}{=} \quad (\varphi_t/\|(\varphi_j)\|_\alpha)\, \boldsymbol{\Theta}_0^{\mathbf{Z}}, \quad \in \tilde{\ell}^\alpha. \tag{5.36}$$

Note again that the norm of the $\alpha-$cluster process, i.e., $|\mathbf{Q}^{(\alpha)}|$, is deterministic in $\tilde{\ell}^\alpha$. Assuming $\|(\varphi_j)\|_p < \infty$, we can compute the indices $c(p)$ in (2.8) by

$$c(p) \quad = \quad \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] \; = \; \|(\varphi_j)\|_p^\alpha / \|(\varphi_j)\|_\alpha^\alpha \; < \; \infty. \tag{5.37}$$

Classic examples of these heavy-tailed linear models are auto-regressive moving averages, i.e., ARMA processes, with iid regularly varying noise; cf. [10].

The proposition below guarantees that the assumptions of Proposition 2.3 hold. We defer its proof to Section B.1.

Proposition 5.4. *Consider $(\mathbf{X}_t)$ to be a linear process with values in $(\mathbb{R}^d, |\cdot|)$, as in Example 5.3. Consider $p > \alpha/2$, and a sequence (x_n) such that $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, $n \to \infty$, for some $\kappa > 0$. Then it holds for all $\delta > 0$*

$$\lim_{s\to\infty} \limsup_{n\to\infty} \frac{\mathbb{P}(\|\mathbf{X}_{[1,n]}/x_n - \mathbf{X}^{(s)}_{[1,n]}/x_n\|_p^p > \delta)}{n\mathbb{P}(|\mathbf{X}_1| > x_n)} \quad = \quad 0, \tag{5.38}$$

where $\mathbf{X}_t^{(s)} := \sum_{|j|\leqslant s} \varphi_j \mathbf{Z}_{t-j}$. *Thus* **AC** *and* $\mathbf{CS}_p$ *are satisfied.*

We now review the mixing properties of a linear process. We recall below the statement in Theorem 2.1. in [41] (see Lemma 15.3.1. in [33]).

Proposition 5.5. *Consider* $(\mathbf{X}_t)$ *to be a causal linear process with values in* $(\mathbb{R}^d, |\cdot|)$, *as in Example 5.3 with* $\varphi_j = 0$, *for* $j < 0$. *Assume the distribution of* $\mathbf{Z}_0$ *is absolutely continuous with respect to the Lebesgue measure in* $\mathbb{R}^d$, *and has a density* g_Z *satisfying*

i) $\int |g(\mathbf{x}-\mathbf{y}) - g(\mathbf{x})| d\mathbf{x} = O(|\mathbf{y}|)$, *for all* $\mathbf{y} \in \mathbb{R}^d$,
ii) $\varphi_t = O(t^{-\rho})$, *for* $t \geqslant 0$, *and* $\rho > 2 + 1/\alpha$,
iii) $\sum_{j=0}^{\infty} \varphi_j \mathbf{x}^j \neq 0$, *for all* $\mathbf{x} \in \mathbb{R}^d$ *with* $|\mathbf{x}| < 1$,

Then, for all $0 < \kappa < \alpha$, *the mixing coefficients* (β_t) *satisfy*

$$\beta_t \;=\; O\Big(t^{1-\frac{(\rho-1)(\alpha-\kappa)}{1+\alpha-\kappa}}\Big). \tag{5.39}$$

Combining Propositions 5.4 and 5.5, we state below the asymptotic normality of the p-cluster based estimators for linear processes in Theorem 5.6. We defer its proof to Section B.2.

Theorem 5.6. *Consider* $(\mathbf{X}_t)$ *to be a causal linear process with values in* $(\mathbb{R}^d, |\cdot|)$, *as in Example 5.3. Let* $\rho > 0$, *and assume the conditions of Proposition 5.5 hold with* $\varphi_t = O(t^{-\rho})$, *for* $t > 0$. *Consider* $p > \alpha/2$, *and a sequence* (k_n) *satisfying* (5.34), *such that* $k = k_n \to \infty$, $m_n/k_n \to \infty$. *Consider* $f_p : \tilde{\ell}^p \to \mathbb{R}$, *and assume* **L** *and* **B** *hold. Furthermore, assume*

i) for $\delta > 0$ *as in* (3.14),

$$\rho \;>\; 3 + \tfrac{2}{\alpha} + \tfrac{2}{\delta}(1 + \tfrac{1}{\alpha}),$$

ii) for all $\kappa > 0$, $n\, b_n^{-\frac{(\rho-1)(\alpha-\kappa)}{1+\alpha-\kappa}} \to 0$, *as* $n \to \infty$.

If f_p *is bounded, condition i) can be replaced by* $\rho > 3 + 2/\alpha$. *Then,*

$$\sqrt{k}(\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}) \;\xrightarrow{d}\; \mathcal{N}\big(0, \mathrm{Var}(\, f_p(Y\mathbf{Q}^{(p)})\,)\big), \quad n \to \infty.$$

In particular, the α*-cluster based estimators from Section 4 in* (4.22) (4.25), *and* (4.30), *are asymptotically normally distributed and their asymptotic variances are null.*

Regarding cluster inference in the case of linear models, the α-cluster approach has an optimal asymptotic variance for shift-invariant functionals since we use the $\ell^\alpha-$norm order statistics. For this reason, it compares favourably with state-of-the-art blocks estimator. For example, for the extremal index, the super-efficient estimator in (4.22) has a lower asymptotic variance than the blocks estimator in (4.23). Indeed the asymptotic variance σ_θ^2 of the latter, computed in (4.24), is not necessarily null. For example, for the autoregressive process of order one AR(1) one has $\sigma_\theta^2 = 1 - \theta_{|\mathbf{X}|} > 0$.

The main drawback of the α-cluster estimator is that we must infer α. We propose to use a consistent and unbiased Hill estimator of $1/\alpha$; see [25, 44]. Numerical experiments in Section 6 show this works fine in practice.

5.3. **Affine stochastic recurrence equation solution under Kesten's conditions.** In this section we focus on the causal solution to the affine stochastic recurrence equation SRE under Kesten's conditions. To guarantee the existence of a solution $(\mathbf{X}_t)$, with values in $(\mathbb{R}^d, |\cdot|)$ as in (5.40) satisfying $\mathbf{RV}_\alpha$, we rely on Theorem 2.1. and Theorem 2.4 in [2]. For an overview, we refer to [11]. In what follows, we study time series $(\mathbf{X}_t)$ as in the Example 5.7 below.

Example 5.7. *Consider $(\mathbf{X}_t)$ to be a sequence with values in $\mathbb{R}^d$ satisfying*

$$\mathbf{X}_t \;=\; \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{B}_t, \qquad t \in \mathbb{Z}, \tag{5.40}$$

where $((\mathbf{A}_t, \mathbf{B}_t))$ is an iid sequence of non-negative random $d \times d$ matrices with generic element $\mathbf{A}$, and non-negative random vectors with generic element $\mathbf{B}$ taking values in $\mathbb{R}^d$. For the existence of a causal stationary solution, we assume

i) $\mathbb{E}[\log^+ |\mathbf{A}|_{op}] + \mathbb{E}[\log^+ |\mathbf{B}|] \;<\; \infty$,

ii) under i), assume the Lyapunov exponent of $(\mathbf{A}_t)$, denoted γ, satisfies

$$\gamma \;:=\; \lim_{n\to\infty} n^{-1} \log |\mathbf{A}_t \cdots \mathbf{A}_1|_{op} \;<\; 0, \quad a.s.$$

To guarantee the heavy-tailedness condition $\mathbf{RV}_\alpha$, we also assume

iii) $\mathbf{B} \neq \mathbf{0}$ *a.s., and* $\mathbf{A}$ *has no zero rows a.s.*

iv) there exists $\kappa > 0$ such that $\mathbb{E}[|\mathbf{A}|_{op}^{\kappa}] < 1$,

v) the set Γ from Equation (5.41) generates a dense group on $\mathbb{R}$,

$$\begin{aligned}\Gamma \;=\; \{&\log |\mathbf{a}_n \cdots \mathbf{a}_1|_{op} : n \geqslant 1,\, \mathbf{a}_n \cdots \mathbf{a}_1 > 0,\\ &\mathbf{a}_n, \ldots, \mathbf{a}_1 \text{ are in the support of } \mathbf{A}\text{'s distribution } \},\end{aligned} \tag{5.41}$$

vi) there exists $\kappa_1 > 0$ such that $\mathbb{E}[(\min_{i=1,\ldots,d} \sum_{t=1}^d A_{ij})^{\kappa_1}] \geqslant d^{\kappa_1/2}$, and $\mathbb{E}[|\mathbf{A}|_{op}^{k_1} \log^+ |\mathbf{A}|_{op}] < \infty$.

vii) under i) – vi), there exists a unique $\alpha > 0$ such that

$$\lim_{n\to\infty} n^{-1} \log \mathbb{E}\big[|\mathbf{A}_n \cdots \mathbf{A}_1|_{op}^{\alpha}\big] \;=\; 0, \tag{5.42}$$

and $\mathbb{E}[|\mathbf{B}|^\alpha] < \infty$. If $d > 1$ assume α is not an even integer.

The $\mathbb{R}^d$-variate series $(\mathbf{X}_t)$, satisfying (5.40) and i) – vii), admits a causal stationary solution and satisfies $\mathbf{RV}_\alpha$, with $\alpha > 0$ as in Equation (5.42).

The previous example is motivated by the seminal Kesten's paper [31]. We follow Theorem 2.1. in [2] to state conditions $i) - ii)$ of Example 5.7. Under the conditions $i) - ii)$, the unique solution $(\mathbf{X}_t)$ of (5.40) has the a.s. causal representation

$$\mathbf{X}_t \;=\; \sum_{i\geqslant 0} \mathbf{A}_{t-i+1} \ldots \mathbf{A}_t\, \mathbf{B}_{t-i}, \qquad t \in \mathbb{Z}, \tag{5.43}$$

where the first summand is $\mathbf{B}_t$ for $i = 0$ by convention; for an overview see [11].

One of the main reasons why the solutions to SRE as in Example 5.7 have received strong interest, is because $(\mathbf{X}_t)$ satisfies $\mathbf{RV}_\alpha$ even when the innovations $((\mathbf{A}_t, \mathbf{B}_t))$ are light-tailed. This feature was first noticed in [31] where the original Kesten's assumptions were introduced. In Kesten's framework, a causal stationary solution to the SRE exists as in (5.43), and the extremes of the series occur due to the sums of infinitely many terms of growing length products appearing in (5.43); see [6] for a review. Further, the community adopted the simplified Kesten's conditions stated by Goldie in [22] for univariate SRE. These conditions also aim to capture the heavy-tailed feature under lighter-tailed innovations. In Example 5.7, we borrow the conditions *iii*) – *vii*) established for the multivariate setting from Theorem 2.4 and Corollary 2.7. in [2]. Then, a solution $(\mathbf{X}_t)$ as in Example 5.7 satisfies $\mathbf{RV}_\alpha$, for $\alpha > 0$, and the index of regular variation α is the unique solution to the Equation (5.42). We are also interested in Example 5.7 because it models classic econometric time series such as the squared ARCH(p), and the volatility of GARCH(p, q) processes; see [11].

Concerning the extremes of $(\mathbf{X}_t)$ in Example 5.7, the forward spectral tail process satisfies the relation

$$\boldsymbol{\Theta}_t \;=\; \mathbf{A}_t \cdots \mathbf{A}_1 \boldsymbol{\Theta}_0, \quad t \geqslant 0,$$

where $(\mathbf{A}_t)$ is an iid sequence distributed as $\mathbf{A}$; see [30]. The backward spectral tail process has a cumbersome representation that we omit here; c.f. [30]. We state in Proposition 5.8 sufficient conditions on $(\mathbf{A}, \mathbf{B})$ yielding assumptions **AC**, $\mathbf{CS}_p$ hold for $p > \alpha/2$, and a suitable sequence (x_n) such that $n\mathbb{P}(|\mathbf{X}_1| > x_n) \to 0$ as $n \to \infty$. In this case the time series $(\mathbf{X}_t)$ admits an $\alpha-$cluster process $\mathbf{Q}^{(\alpha)}$. We recall the identity from Equation (8.6) of [12]: $c(p) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] = \mathbb{E}[\|(\boldsymbol{\Theta}_t)_{t\geqslant 0}\|_p^\alpha - \|(\boldsymbol{\Theta}_t)_{t\geqslant 1}\|_p^\alpha]$, for $c(p)$ as in (2.8). Then, letting $p = \alpha/2$, a straightforward computation yields

$$\begin{aligned} c(p) \;&\leqslant\; 2\,\mathbb{E}[\|(\boldsymbol{\Theta}_t)_{t\geqslant 0}\|_p^{\alpha-p}] \;=\; 2\,\mathbb{E}[\textstyle\sum_{t\geqslant 0}|\mathbf{A}_t \cdots \mathbf{A}_1 \boldsymbol{\Theta}_0|^p] \\ &\leqslant\; 2\,s\textstyle\sum_{t\geqslant 0}(\mathbb{E}[|\mathbf{A}_s \cdots \mathbf{A}_1|_{op}^p])^t, \end{aligned}$$

and $\mathbb{E}[|\mathbf{A}_s \cdots \mathbf{A}_1|_{op}^p] < 1$, for $p < \alpha$ and $s \geqslant 1$ fixed sufficiently large in the setting of Example 5.7. Hence, for $p \in (\alpha/2, \alpha)$, $c(p) < \infty$ in (2.8), and then the series admits a $p-$cluster process $\mathbf{Q}^{(p)}$.

We state now Proposition 5.8 which verifies conditions **AC**, $\mathbf{CS}_p$ for the SRE equation. The proof is postponed to Section B.3.

Proposition 5.8. *Let $(\mathbf{X}_t)$ be a stationary time series with values in $(\mathbb{R}^d, |\cdot|)$, as in Example 5.7. Let $p > \alpha/2$, and consider (x_n) such that there exists $\kappa > 0$ satisfying $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, as $n \to \infty$. Then, (x_n) satisfies conditions* **AC** *and* $\mathbf{CS}_p$.

In the setting of SRE equations, condition **AC** has been shown in Theorem 4.17 in [38]. In [38], the authors already considered a condition similar to

$\mathbf{CS}_p$. Parallel to their setting, we propose a proof of Proposition 5.8 which shows $\mathbf{CS}_p$ holds over uniform regions $\Lambda_n = (x_n, \infty)$ such that $n/x_n^p \to 0$, as $n \to \infty$, in the sense of (2.77). Thereby, our proof extends Theorem 4.17 in [38] to uniform regions Λ_n not having an upper bound.

Concerning the mixing properties of $(\mathbf{X}_t)_{t\geqslant 0}$ as in Example 5.7, we use that it is a Markov chain and that $\mathbf{X}_0$ has the stationary distribution. As mentioned in Remark 2.2, we can then use Markov chain's theory to compute its mixing coefficients; cf. [36]. We review Theorem 2.8. in [2], yielding an exponential decay of the mixing-coefficients (β_t) of the series. For a general treatment see Chapter 4.2 in [11].

Proposition 5.9. *Consider a time series $(\mathbf{X}_t)$ with values in $(\mathbb{R}^d, |\cdot|)$, as in Example 5.7. Assume there exists a Borel measure μ on $(\mathbb{R}^d, |\cdot|)$, such that the Markov chain $(\mathbf{X}_t)_{t\geqslant 0}$ is μ-irreducible, i.e., for all $C \subset \mathbb{R}^d$ with $\mu(C) > 0$,*

$$\sum_{t=0}^{\infty} \mathbb{P}(\mathbf{X}_t \in C \,|\, \mathbf{X}_0 = \mathbf{x}) > 0, \quad \mathbf{x} \in \mathbb{R}. \tag{5.44}$$

Then $(\mathbf{X}_t)$ has mixing coefficients (β_t) satisfying $\beta_t = O(\rho^t)$ for some $\rho \in (0,1)$, and we say it is strongly mixing with geometric rate. Moreover, $(\mathbf{X}_t)_{t\geqslant 0}$ is irreducible with respect to the Lebesgue measure if $(\mathbf{A}, \mathbf{B})$ admits a density.

We can now state the asymptotic normality of cluster-based estimator for SRE solutions in Theorem 5.10 below. The proof is postponed to Section 2.4.

Theorem 5.10. *Consider $(\mathbf{X}_t)$ to be the causal solution to the SRE in (5.40) with values in $(\mathbb{R}^d, |\cdot|)$, as in Example 5.7. Assume the conditions of Proposition 5.9 hold. Consider $p > \alpha/2$, and a sequence (k_n) satisfying (5.34), such that $k = k_n \to \infty$, $m_n/k_n \to \infty$. Consider $f_p : \tilde{\ell}^p \to \mathbb{R}$, and assume $\mathbf{L}$ and $\mathbf{B}$ hold. Assume $\log(n)/b_n \to 0$, as $n \to \infty$, Then,*

$$\sqrt{k}(\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(f(Y\mathbf{Q}^{(p)}))), \quad n \to \infty.$$

In particular, the α-cluster based estimators from Section 4 in (4.22) (4.25), and (4.30), are asymptotically normally distributed.

Remark 5.11. In this example, the asymptotic variances of the α-cluster based estimators from Section 4 in (4.22) (4.25), and (4.30) are non-null. The limiting variances in Theorem 5.10 are difficult to compare with the existing ones in the literature because of the complexity of the distribution of $\mathbf{Q}^{(p)}$. However, we provide simple ℓ^α-block estimators of the asymptotic variances in Section 6.

6. Numerical experiments

This section aims to illustrate the finite-sample performance of the α-cluster estimators on time series $(\mathbf{X}_t)$ with tail-index $\alpha > 0$. In all the models we consider in Section 5, we work under the assumption that the

tuning parameters of the α-cluster satisfy (5.34). We take $\kappa' = 1$ in (5.34) which yields $b = \sqrt{n/k}$. In this case, the implementation of our estimators can be written solely as a function of k. Recall $k = k_n$ must satisfy $k \to \infty$ and $m/k \to \infty$ with $m = [n/b]$. Numerical comparisons of our α-cluster based approach with other existing estimators for the extremal index and the cluster index are at the advantage of our approach; see [13] and [12]. The code of all numerical experiments is available at: https://github.com/GBuritica/cluster_functionals.git.

6.1. **Cluster size probabilities.** We reviewed in Section 4.3 how cluster sizes play a key role to model the serial behavior of exceedances. In this section, we implement the cluster size probabilites estimation procedure from Equation (4.30) in an example of a solution to the SRE under Kesten's conditions.

Example 6.1. *Consider the non-negative univariate random variables A, B, defined by $\log A = N - 0.5$, where N denotes a standard Gaussian random variable, and B is uniformly distributed in $[0, 1]$. Let (X_t) be the solution to the SRE in (5.40). Then, (X_t) satisfies $\mathbf{RV}_\alpha$ with $\alpha = 1$. If (A_j) is a sequence of iid random variables with generic element A, then*

$$Q_t^{(\alpha)} \stackrel{d}{=} \Pi_t / \|(\Pi_j)\|_\alpha\,, \qquad t \in \mathbb{Z}\,,$$

with

$$\Pi_t \stackrel{d}{=} \begin{cases} A_t \cdots A_1 & \text{if} \quad t \geqslant 1, \\ A_t \cdots A_{-1} & \text{if} \quad t \leqslant -1, \\ 1 & \text{if} \quad t = 0\,. \end{cases}$$

This follows by Example 6.1 in [30], and Proposition 3.1 in [12]. Then, for $p > \alpha/2$, the p-cluster based estimators (4.22) (4.25), and (4.30) are asymptotically normally distributed.

Recall the cluster sizes $\pi_1, \pi_2, \ldots$, defined in (4.29). We infer the cluster sizes of Example 6.1 using α-cluster estimates. To illustrate Theorem 5.10, we run a Monte–Carlo simulation experiment based on 1 000 samples $(X_t)_{t=1,\ldots,n}$ of length $n = 12\,000$ from Example 6.1. For each sampled trajectory, we obtain estimates $\widehat{\pi}_1, \widehat{\pi}_2, \ldots$, letting $k = 8$ and $b = 38$ in (4.30). We use the known value of the tail-index $\alpha = 1$. We also estimate the extremal index θ_X of the series from Equation (4.22). Theorem 5.10 yields, for $j \geqslant 0$,

$$\mathrm{Var}(\sqrt{k}\,(\widehat{\pi}_j - \pi_j^{\mathbf{Q}})) \quad \to \quad \mathrm{Var}(\pi_j^{\mathbf{Q}}(\mathbf{Q}^{(\alpha)})), \quad n \to \infty, \tag{6.45}$$

where $\pi_j^{\mathbf{Q}}$ are the cluster functional yielding the cluster sizes π_j with the notation in (4.31). Notice that the asymptotic variance of our cluster sizes estimate is again a cluster statistic that we can infer. We compute an estimate of the asymptotic variance in (6.45) using cluster-based estimates, and compare this estimate with the empirical variance obtained from the

Monte-Carlo simulation study. Figure 1 plots the profile of the limit Gaussian distribution where the asymptotic variance is computed in these two ways. As expected from Equation (6.45), the curves overlap, even if k is small.

In the case of SRE equations, the cluster sizes were studied in detail in [24]. The authors proposed a method to approximate the true values when the tail-index α, and the random variable A are known. We approximate true values using Equation 3.5 in [24], and a Monte-Carlo study with 10 000 samples of length 500 000. The obtained values are pointed out in red in Figure 1. We see that this choice of k yields estimates centered around the true value.

6.2. **Replacing α by $\widehat{\alpha}$.** So far we have used known values of α. In this section, we illustrate that the $\alpha-$cluster-based estimators perform well in simulation when we replace α with a Hill-type estimator $\widehat{\alpha}$. We use the bias-correction procedure in [23] and write an estimator of the tail-index as $k' \mapsto \widehat{\alpha}(k')$ where the tuning parameter k' is the number of higher order statistics of $(|\mathbf{X}_t|)$ that we use for inference. For consistency of the tail-index estimator, we must take $k' = k'_n$ satisfying $k' \to \infty$ and $n/k' \to \infty$. Therefore, we recommend choosing the tuning sequence of the tail-index and of the cluster estimators as (k'_n), (k_n), respectively, such that $k'/k \to \infty$. Roughly speaking, the cluster statistics capture the block extremal behavior whereas the tail-index points to an extremal property of margins. As a consequence, the variance of the Hill procedure should not affect the variance of the cluster estimates asymptotically. To illustrate this point, we simulate 500 samples $(X_t)_{t=1,\dots,n}$ of an AR(φ) model with absolute value student(α) noise for $n = 12\,000$, $\alpha = 1$ and $\varphi \in \{0.5, 0.7\}$, and for samples of Example 5.7. We estimate the extremal index $\widehat{\theta}_X(k)$ as in (4.22) where we replace α by $\widehat{\alpha}(k')$. Recall that for an AR(φ) model the asymptotic variances of the extremal index estimator are null. We see in Figures 2, 3 and 4 that in practice we have to choose k small to reduce the bias of the estimator. Moreover, the estimation procedure is robust with respect to k' therefore we recommend taking k' large to reduce variance. Similar results were found for $n = 3\,000$, $n = 5\,000$, and $n = 8\,000$ and these are available upon request. To conclude, we see in Figures 2, 3 and 4 that standard deviations are small, and thus the error of cluster inference is mainly due to bias. We recommend choosing k small and k' larger in all settings.

6.3. **Conclusion.** Our main theoretical result in Theorem 3.2 states asymptotic normality of α-cluster-based disjoint blocks estimators $\widehat{f}_\alpha^{\mathbf{Q}}(k)$, based on k extremal ℓ^α-blocks, where α is the tail index of the series. The advantage of α-cluster-based methods is that the choice of k is robust to time dependencies; see [12] who have shown already consistency of $\widehat{f}_\alpha^{\mathbf{Q}}$ estimators. Equation (3.21) characterizes their asymptotic variance in terms of a cluster statistic that we can infer. We further show in Section 4 that many