Estimation of cluster functionals for regularly varying time series

Youssouph Cissokho

Thesis submitted as part of the requirements of the PhD program
In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematics and Statistics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa.

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

The classical Extreme Value Theory deals with independent random variables. If random variables are dependent, large values tend to cluster (that is, one large value is followed by a series of large values). It is of interest to describe probabilistically the clustering and estimate the relevant cluster functionals. We consider disjoint blocks, sliding blocks and runs estimators of cluster indices. Using a modern theory of multivariate, regularly varying time series, we obtain consistency results and central limit theorems under conditions that can be easily verified for a large class of short-range dependent models. In particular, we show that in the Peak-over-Threshold framework, all the estimators have the same limiting variances. This solves a longstanding open problem and is in contrast to the Block Maxima method. Our findings are illustrated by simulation experiments.

Acknowledgments

First of all, I would like to thank the Almighty God for his grace and mercy he has given me to be able to finish my project.

I would like to express my deep gratitude to Professor **Rafal Kulik**, my research supervisor, for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been really appreciated. He is not only a great and brilliant professional academics but also an amazing person willing to see his students to succeed.

I am thankful to the members of my thesis committee :

1. **Dr. Stanislav Volgushev** (University of Toronto)

2. **Dr. Mohamedou Ould-Haye** (Carleton University)

3. **Dr. Raluca Balan** (University of Ottawa)

4. **Dr. Mahmoud Zarepour** (University of Ottawa)

for your work as thesis examiners.

Finally, my grateful thanks are also extended to all the members of Ottawa-Carleton community.

Dedications

I cannot finish thanking without saying something to my own parent. I am very grateful to my mother **Adama Sabaly**, my father **Seni Cissokho**, my sisters and bothers: **Fatou Bintou, Mansata, Seydou, Momodou, Mouskouta**, for their support and for taking care of me since I was young, and to all the **Cissokho family**. May God bless them.

I also own my deepest gratitude to my lovely Princess **Isatou Choi** and affectionately dedicated this thesis to her. You are a gift from heaven and one in a million.

Contents

Chapter 1

Introduction

1.1. Motivation and Goals

Stationary, multivariate time series often exhibits dependence between its coordinates and over time. As such, the analysis of dependence between large observations is a suitable approach to understand extremal behaviour of stationary time series. So-called cluster indices (to be defined formally in the subsequent chapters) are a crucial tool for this purpose. Of interest, is the estimation of cluster indices thanks to a modern theory of multivariate, regularly varying time series.

The classical extreme value theory deals with independent random variables. If $\{X_j, j \in \mathbb{Z}\}$ are i.i.d. nonnegative random variables with the marginal distribution function F and $\mathbb{P}(X_0 > x) = x^{-\alpha}L(x)$, $\alpha > 0$, where L is a slowly varying function, then there exists a sequence $a_n = F^{\leftarrow}(1 - 1/n)$, where $F^{\leftarrow}$ is the quantile function, such that

$$\lim_{n\to\infty} \mathbb{P}\left(a_n^{-1} \max_{j=1,\dots,n} X_j \le x\right) = \exp(-x^{-\alpha}) \ , \quad x > 0 \ .$$

If random variables are dependent (but have the same marginal distribution), large values tend to cluster. In particular, under the appropriate weak dependence conditions

$$\lim_{n\to\infty} \mathbb{P}\left(a_n^{-1} \max_{j=1,\dots,n} X_j \le x\right) = \exp(-\theta x^{-\alpha}) \ , \quad x > 0 \ ,$$

where $\theta \in (0, 1]$ is called the *extremal index* (whenever exists). See e.g. [KS20, Chapter 7].

The extremal index describes the amount of clustering and is one of the examples of so-called cluster indices. Also, it plays an important role in statistical inference (for example, the extremal index is an important tool to estimate *return levels* (in hydrology) and *risk measures* (in finance)). As such, it is of interest to describe probabilistically clustering and estimate the relevant cluster indices. Let $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a multivariate, $\mathbb{R}^d$-real valued time series. Informally speaking, a cluster is a triangular array $(\boldsymbol{X}_1/u_n, \dots, \boldsymbol{X}_{r_n}/u_n)$ with $r_n, u_n \to \infty$ that converges in distribution. Cluster indices are obtained by applying the appropriate functional H to the cluster. The cluster functionals are defined on $(\mathbb{R}^d)^{\mathbb{Z}}$, the space of $\mathbb{R}^d$-valued sequences, and are such that their values do not depend on coordinates that are equal to zero. More precisely, for $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z}$, we denote

$$\boldsymbol{X}_{i,j} = (\boldsymbol{X}_i, \dots, \boldsymbol{X}_j).$$

Then, we consider the cluster functional applied to the block $H(\boldsymbol{X}_{i,j})$, by identifying $H(\boldsymbol{X}_{i,j})$ with $H(\boldsymbol{0}, \boldsymbol{X}_{i,j}, \boldsymbol{0})$. Here, $(\boldsymbol{0}, \boldsymbol{X}_{i,j}, \boldsymbol{0})$ denotes a random element with values in $(\mathbb{R}^d)^{\mathbb{Z}}$ with zeros at all the coordinates smaller than i and larger than j. For example, the functional $H(\boldsymbol{x}) =$

$\mathbb{1}\{\max_{j\in\mathbb{Z}} |\boldsymbol{x}_j| > 1\}$, $\boldsymbol{x} \in (\mathbb{R}^d)^{\mathbb{Z}}$, depends on the coordinates large than 1 only, and leads to the aforementioned extremal index.
When dealing with extremes, there are two main estimation methods: Peaks-over-Threshold (PoT) and Block Maxima (BM). For the former, we look only at the data points which exceed a high threshold. The latter, consists of dividing the data into (non)-overlapping intervals and considering the largest observation in each interval.

In the PoT framework, there are two basic estimation methods: disjoint blocks and sliding blocks. As the names suggest, in the first case we divide data into non-overlapping blocks, apply the appropriate function to each block and then average over all blocks. The second case consists of sliding a window of a given size through the data, resulting in overlapping blocks. As such we have more blocks, but we introduce an additional dependence. Mathematically speaking, for the disjoint blocks estimators we consider $H(\boldsymbol{X}_{(i-1)r_n+1,ir_n})$, $1 \le i \le [n/r_n]$, while for the sliding blocks estimators we deal with $H(\boldsymbol{X}_{i+1,i+r_n})$, $1 \le i \le n-r_n+1$. In principle, disjoint and sliding blocks estimators may have a different asymptotic behaviour.

Several studies have been done in both cases. In the Block Maxima (BM) framework, we would like to point out [BS18b, BS18a] and the recent paper by [ZVB20]. Since this thesis deals with the PoT framework in what follows we focus on the recent literature that uses this methodology. Although some special cases were considered (estimation of the extremal index in [Hsi91] and [SW94]; tail array sums in [RLdH98]), the general theory was developed in [DR10]. There the authors provided a general methodology to deal with statistics for disjoint blocks of cluster functionals. Using the same methodology, in [DN20], the authors studied the sliding blocks and showed that (under some conditions) the limiting variance of the sliding blocks estimator never exceeds that of the disjoint blocks estimator. In case of the extremal index, both variances were proven to be equal. The latter result is in contrast to the BM method, where the variances of the sliding blocks estimators are typically smaller than those of the disjoint ones. The comprehensive discussion and up-to-date results can be found in the recent monograph [KS20, Chapters 9 and 10]. The other relevant references are included in each chapter.

Up until recently, there is no thorough explanation of these phenomena and no formal comparison between PoT and BM framework. See [FdH15] for some partial results and [BZ18] for a recent review.

In this thesis, we deal with multivariate, regularly varying time series and consider the disjoint blocks *statistics*

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{m_n} H\left(\boldsymbol{X}_{(j-1)r_n+1,jr_n}/u_n\right) \,,$$

where $m_n = [n/r_n]$ is the number of disjoint blocks; the sliding blocks statistics

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) := \frac{1}{q_n r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \,, \tag{1.1.1}$$

where $q_n = n - r_n$ is the number of sliding blocks; and the runs statistics

$$\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=r_n+1}^{n-r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \,, \tag{1.1.2}$$

with an arbitrary choice of the so-called anchoring map $\mathcal{C}$ (see Section (2.4)), where

$$H^{\mathcal{C}}(\boldsymbol{x}) = H(\boldsymbol{x})\mathbb{1}\{\mathcal{C}(\boldsymbol{x}) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > 1\} \,.$$

Also, we consider the corresponding *estimators*, where the unknown sequence u_n is replaced with intermediate order statistics.

The goal is to obtain the consistency and the asymptotic normality of both statistics and the corresponding estimators. Our focus is on providing the conditions that can be easily verified for a variety of time series models. At the same time, we will show that the limiting variance of both sliding blocks and runs estimators are the same as the disjoint blocks estimators. In the general framework of multivariate, regularly varying time series, this is done for the first time in the literature.

1.2. Methodology

The starting point for our approach is the seminal paper by [DR10], where the authors consider disjoint blocks statistics in a very general, abstract framework. This theory was further extended, summarized, and developed under very mild conditions in [KS20]. From there, a natural question arises: what is the asymptotic behavior of the sliding blocks statistics? This line of work is also extended to the so-called runs estimators.

In order to answer the research questions, we use the modern theory of multivariate, regularly varying time series along with empirical process techniques.

1.3. Contribution and Structure

This thesis is based on three papers that are published either as the journal articles (Electronic Journal of Statistics) or preprints. In all the papers, some overlap and redundancy may be found in the introductory and preliminary sections of each chapter. Indeed, they share generally the same theoretical part, have common terminology and notation.

The thesis is organized as follows:

1.3.1. Chapter 2: Preliminaries

Chapter 2 includes all the tools needed to establish our asymptotic results. In particular, we fix the notation, we introduce the relevant classes of functions, cluster functionals and anchoring maps. We introduce the important object, called the tail process. It is a distributional limit of the stationary time series, conditionally on large value at time 0. With its help we define the cluster measure and the cluster indices.

Convergence of cluster measure requires so-called anticlustering condition. The condition itself guarantees vague convergence for functionals that vanish around zero. If this is not the case, small jumps have to be controlled.

Finally, we provide several examples of time series for which our theorems hold.

1.3.2. Chapter 3: Consistency of disjoint blocks estimators

This chapter discusses the conditions for consistency of the disjoint blocks statistics $\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H)$. It is entirely based on [KS20, Chapter 10].

1.3.3. Chapter 4: Consistency of sliding blocks estimators

In Chapter 4, we establish consistency of the sliding blocks statistics under very mild assumptions on the temporal dependence structure. In particular, consistency of sliding blocks estimators holds for long memory moving averages. This has never been done in the literature before. The chapter is based on a preprint that we intend to submit soon.

1.3.4. Chapter 5: Central limit theorem for disjoint blocks estimators

This chapter deals with the central limit theorem for disjoint blocks statistics and estimators. Again, it is based on [KS20, Chapter 10]. The limiting results are valid under anticlustering and mixing conditions that can be relatively easily verified for many time series models. In the subsequent chapters we use the methodology that stems from this one.

1.3.5. Chapter 6: Central limit theorem for sliding blocks estimators

In Chapter 6 we consider sliding blocks statistics and estimators. They have been studied for some specific functionals H (leading primarily to the extremal index), however there has been no unified theory available. Recently, [DN20] used the framework of [DR10] and showed that the limiting variance of the sliding blocks estimator never exceeds that of the disjoint blocks estimator. In case of the extremal index, both variances were proven to be equal.
The main result of this chapter is the asymptotic normality of the appropriately normalized statistics in 1.1.1 and the corresponding estimator based on the intermediate order statistics. The most important (and somehow surprising) conclusion is that both sliding and disjoint blocks estimators yield the same the limiting variance.

The corresponding paper is published in the Electronic Journal of Statistics as [CK21].

1.3.6. Chapter 7: Central limit theorem for runs estimators

In the spirit of Chapter 6, Chapter 7 tackles the runs statistics and estimators. We prove the asymptotic normality of the appropriately normalized statistics and the corresponding data-based estimator. We show, in particular, that the limiting variance agrees with the one for the disjoint blocks and sliding blocks estimators. Furthermore, we proved that variance reduction cannot be achieved by considering a linear combination of runs estimators with a different choice of anchoring maps.

The corresponding paper is published in the Electronic Journal of Statistics as [CK22]

1.3.7. Chapter 8: Conclusions and future

In this chapter we summarize our findings (this time including some technical terms). We also discuss potential directions of future research, including asymptotic expansions for the disjoint and sliding blocks estimators, limit theorems for long memory sequences, piecewise stationary time series among others.

1.3.8. Chapter 9: Appendix

Appendix includes the foundational material on VC-classes. It is needed to establish functional convergence.

1.4. Summary of the Contribution

To the best of our knowledge, the following results are new:

- Consistency of sliding blocks statistics, including long memory moving averages. This has never been done in the literature before. In fact, there are very few results on statistical inference for extremes in case of long memory sequences.
- The results for disjoint blocks estimators in Chapter 6 are new and are published in the Electronic Journal of Statistics in [CK21].
- The results for runs estimators in Chapter 7 are new and is published in Electronic Journal of Statistics as well in [CK22]

Chapter 2

Preliminaries

In this section, we fix the notation and introduce the relevant classes of functions (Sections 2.1 and 2.3). In Section 2.4 we define anchoring maps while in Section 2.2 we recall the notion of the tail and the spectral tail process (cf. [BS09]) along with the definition of cluster indices; see [KS20, Chapter 5] for a detailed introduction. Section 2.6 contain results about the convergence of cluster measures.

For this chapter, the appropriate references can be found at the end of this thesis, on page 131.

2.1. Notation

Let $|\cdot|$ be a norm on $\mathbb{R}^d$. For a sequence $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z} \cup \{-\infty, \infty\}$ we denote $\boldsymbol{x}_{i,j} = (\boldsymbol{x}_i, \dots, \boldsymbol{x}_j) \in (\mathbb{R}^d)^{j-i+1}$, $\boldsymbol{x}^*_{i,j} = \max_{i \le l \le j} |\boldsymbol{x}_l|$ and $\boldsymbol{x}^* = \sup_{j \in \mathbb{Z}} |\boldsymbol{x}_j|$. By $\boldsymbol{0}$ we denote the zero sequence; its dimension can be different in each occurrences.

By $\ell_0(\mathbb{R}^d)$ we denote the set of $\mathbb{R}^d$-valued sequences which tend to zero at infinity. Likewise, $\ell_1(\mathbb{R}^d)$ consists of sequences $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ such that $\sum_{j \in \mathbb{Z}} |\boldsymbol{x}_j| < \infty$.

2.2. Tail and spectral tail process

Let $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying time series with values in $\mathbb{R}^d$ and tail index α. In particular,

$$\lim_{x \to \infty} \frac{\mathbb{P}(|\boldsymbol{X}_0| > tx)}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = t^{-\alpha}$$

for all $t > 0$. Then, there exists a sequence $\boldsymbol{Y} = \{\boldsymbol{Y}_j, j \in \mathbb{Z}\}$ such that

$$\mathbb{P}(x^{-1}(\boldsymbol{X}_i, \dots, \boldsymbol{X}_j) \in \cdot \mid |\boldsymbol{X}_0| > x) \text{ converges weakly to } \mathbb{P}((\boldsymbol{Y}_i, \dots, \boldsymbol{Y}_j) \in \cdot)$$

as $x \to \infty$ for all $i \le j \in \mathbb{Z}$. We call $\boldsymbol{Y}$ the tail process. See [BS09]. We note that, in particular, $|\boldsymbol{Y}_0|$ has Pareto distribution with the density $\alpha x^{-\alpha-1}$, $x > 1$. As such, it follows automatically that $\boldsymbol{Y}^* = \sup_{j \in \mathbb{Z}} |\boldsymbol{Y}_j| > 1$. Equivalently, viewing $\boldsymbol{X}$ and $\boldsymbol{Y}$ as random elements with values in $(\mathbb{R}^d)^{\mathbb{Z}}$, we have for every bounded or non-negative functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, continuous with respect to the product topology,

$$\lim_{x \to \infty} \frac{\mathbb{E}[H(x^{-1}\boldsymbol{X}) \mathbb{1}\{|\boldsymbol{X}_0| > x\}]}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = \mathbb{E}[H(\boldsymbol{Y})] \; .$$

Define $\mathbf{\Theta}_j = \boldsymbol{Y}_j/|\boldsymbol{Y}_0|$, $j \in \mathbb{Z}$. The sequence $\mathbf{\Theta} = \{\mathbf{\Theta}_j, j \in \mathbb{Z}\}$ is called the spectral tail process. The random variable $|\boldsymbol{Y}_0|$ has the Pareto distribution with index α and is independent from $\mathbf{\Theta}$. See [KS20, Chapter 5]. Hence for a non-negative measurable function $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$,

$$\mathbb{E}[H(\boldsymbol{Y})] = \int_1^\infty \mathbb{E}[H(r\mathbf{\Theta})]\alpha r^{-\alpha-1}\mathrm{d}r \; . \tag{2.2.1}$$

Tail process will be na important object in the remainder of the thesis. In particular, the limiting variances in the central limit theorems presented in this thesis will be expressed in terms of an infinite series of the tail process.

2.3. Classes of functions

Functionals H are defined on $\ell_0(\mathbb{R}^d)$ with the convention $H(\boldsymbol{x}_{i,j}) = H((\mathbf{0}, \boldsymbol{x}_{i,j}, \mathbf{0}))$. We consider the following classes:

- $\mathcal{L}$ is the class of bounded real-valued functions H defined on $(\mathbb{R}^d)^{\mathbb{Z}}$ that are either Lipschitz continuous with respect to the uniform norm, that is

$$|H(\boldsymbol{x}) - H(\boldsymbol{y})| \le L_H \sum_{j\in\mathbb{Z}} \left|\boldsymbol{x}_j - \boldsymbol{y}_j\right| \;, \quad \boldsymbol{x}, \boldsymbol{y} \in \ell_1(\mathbb{R}^d) \;, \tag{2.3.1}$$

  for some constant L_H that depends on H, or almost surely continuous with respect to the distribution of the tail process $\boldsymbol{Y}$. This class includes functions like $\mathbb{1}\{\boldsymbol{x}^* > 1\}$ and

$$\mathcal{E}(\boldsymbol{x}) := \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} |\boldsymbol{x}_j| > 1\right\} .$$

  See Remark 6.1.6 in [KS20].
- $\mathcal{A} \subset \mathcal{L}$ is the class of shift-invariant functionals with support separated from $\mathbf{0}$. In particular, for $H \in \mathcal{A}$, $H(\mathbf{0}) = 0$. The class $\mathcal{A}$ includes $\mathbb{1}\{\boldsymbol{x}^* > 1\}$.
- $\mathcal{K}$ is the class of shift-invariant functionals $K : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ defined on $\ell_1(\mathbb{R}^d)$ such that $K(\mathbf{0}) = 0$ and which are Lipschitz continuous with respect to the uniform norm.
- $\mathcal{B} \subset \mathcal{L}$ is the class of functionals H of the form $H = \mathbb{1}\{K > 1\}$, where $K \in \mathcal{K}$. Functionals in $\mathcal{B}$ may have support which is not separated from $\mathbf{0}$. The typical example is $H(\boldsymbol{x}) = \mathbb{1}\left\{\sum_j |\boldsymbol{x}_j| > 1\right\}$; note that $H \notin \mathcal{A}$. At the same time $\mathbb{1}\{\boldsymbol{x}^* > 1\}$ belongs to both $\mathcal{A}$ and $\mathcal{B}$.

The cluster indices of interest are, among others:

- the candidate extremal index obtained with $H_1(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\}$, $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$;
- the cluster size distribution obtained with

$$H_{2,m}(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\} = m\right\} , \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}} \;, \quad m \in \mathbb{N} \;; \tag{2.3.2}$$

- the stop-loss index of a univariate time series obtained with

$$H_{3,\eta}(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} (x_j - 1)_+ > \eta\right\} , \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \;, \quad \eta > 0 \;; \tag{2.3.3}$$

- the large deviation index of a univariate time series obtained with

$$H_4(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\} \;, \quad K(\boldsymbol{x}) = \left(\sum_{j\in\mathbb{Z}} x_j\right)_+ , \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \;; \tag{2.3.4}$$

- the ruin index of a univariate time series obtained with

$$H_5(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\} \ , \quad K(\boldsymbol{x}) = \sup_{i\in\mathbb{Z}} \left(\sum_{j\leq i} x_j \right)_+ \ , \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \ . \tag{2.3.5}$$

As mentioned above, the candidate extremal index is the classical quantity that arises in the extreme value theory for dependent sequences. The cluster size distribution is again a well-known object and was studied in [Hsi91] and [DR10]. The large deviation index was studied under the name *cluster index* in [MW13, MW14]. It quantifies the effect of dependence in large deviations results. The remaining cluster indices seem to be new.

2.4. Anchoring maps

Definition 2.4.1 (Anchoring map). *A measurable map $\mathcal{C} : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{Z} \cup \{-\infty, \infty\}$ is called an anchoring map if the following two properties hold:*

An(i)*: $\mathcal{C}(\boldsymbol{x}) = j$ implies $|\boldsymbol{x}_j| \geq |\boldsymbol{x}_0| \wedge 1$;*
An(ii)*: $\mathcal{C}(B\boldsymbol{x}) = \mathcal{C}(\boldsymbol{x}) + 1$, where B is a backsift operator : $(B\boldsymbol{x})_j = \boldsymbol{x}_{j-1}$.*

Three basic examples of anchoring maps are:

- The infargmax functional: $\mathcal{C}^{(0)}(\boldsymbol{y}) = \inf\{j : \boldsymbol{y}^*_{-\infty,j} = \boldsymbol{y}^*\}$;
- The first exceedence above one: $\mathcal{C}^{(1)}(\boldsymbol{x}) = \inf\{j : |\boldsymbol{x}_j| > 1\}$;
- The last exceedence above one: $\mathcal{C}^{(2)}(\boldsymbol{x}) = \sup\{j : |\boldsymbol{x}_j| > 1\}$.

To the best of our knowledge, anchoring maps were introduced in [BP18]. They also appear in [Has18] under the name 0-pinning functionals.

2.5. Cluster measure and cluster indices

Consider the infargmax functional $\mathcal{C}^{(0)}$ defined on $(\mathbb{R}^d)^{\mathbb{Z}}$ by $\mathcal{C}^{(0)}(\boldsymbol{y}) = \inf\{j : \boldsymbol{y}^*_{-\infty,j} = \boldsymbol{y}^*\}$, with the convention that $\inf \emptyset = +\infty$. If $\mathbb{P}(\mathcal{C}^{(0)}(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$, then we can define

$$\vartheta = \mathbb{P}\Big(\mathcal{C}^{(0)}(\boldsymbol{Y}) = 0\Big) \ . \tag{2.5.1}$$

In fact, $\mathcal{C}^{(0)}$ can be replaced with any anchoring map (see Lemma 2.5.6).

Example 2.5.1. For $\mathcal{C}^{(1)}(\boldsymbol{x}) = \inf\{j : |\boldsymbol{x}_j| > 1\}$; we have

$$\mathbb{P}\Big(\mathcal{C}^{(1)}(\boldsymbol{Y}) = 0\Big) = \mathbb{P}\Big(\sup_{j\leq -1} |\boldsymbol{Y}_j| \leq 1, |\boldsymbol{Y}_0| > 1\Big) = \mathbb{P}\Big(\sup_{j\leq -1} |\boldsymbol{Y}_j| \leq 1\Big).$$

Similarly for $\mathcal{C}^{(2)}(\boldsymbol{x}) = \sup\{j : |\boldsymbol{x}_j| > 1\}$, we have

$$\mathbb{P}\Big(\mathcal{C}^{(2)}(\boldsymbol{Y}) = 0\Big) = \mathbb{P}\Big(\sup_{j\geq 1} |\boldsymbol{Y}_j| \leq 1, |\boldsymbol{Y}_0| > 1\Big) = \mathbb{P}\Big(\sup_{j\geq 1} |\boldsymbol{Y}_j| \leq 1\Big).$$

We have then

$$\vartheta = \mathbb{P}\left(\sup_{j\leq -1} |\boldsymbol{Y}_j| \leq 1\right) = \mathbb{P}\Big(\sup_{j\geq 1} |\boldsymbol{Y}_j| \leq 1\Big) \ . \tag{2.5.2}$$

The relationship between (2.5.1) and (2.5.2) is certainly not obvious. It will be a consequence of Lemma 2.5.6. Equation (2.5.2) emphasizes a special role of the event $\{\sup_{j\le -1}|\boldsymbol{Y}_j|\le 1\}$ and with its help ϑ can be recognized as the (candidate) extremal index. It becomes the usual extremal index under additional mixing and anticlustering conditions.

Definition 2.5.2. *Let $\boldsymbol{Y}$ and $\boldsymbol{\Theta}$ be the tail process and the spectral tail process, respectively, such that $\mathbb{P}(\lim_{|j|\to\infty}\boldsymbol{Y}_j=\boldsymbol{0})=1$. The **cluster measure** is the measure $\boldsymbol{\nu}^*$ on $\ell_0(\mathbb{R}^d)$ defined by*

$$\boldsymbol{\nu}^* = \vartheta\int_0^\infty \mathbb{E}\left[\delta_{r\boldsymbol{\Theta}}\mathbb{1}\left\{\mathcal{C}^{(0)}(\boldsymbol{\Theta})=0\right\}\right]\alpha r^{-\alpha-1}\mathrm{d}r\ . \tag{2.5.3}$$

The measure $\boldsymbol{\nu}^*$ is boundedly finite on $(\mathbb{R}^d)^{\mathbb{Z}}\setminus\{\boldsymbol{0}\}$, puts no mass at $\boldsymbol{0}$ and is α-homogeneous. Furthermore, the cluster measure can be expressed in terms of another sequence.

Definition 2.5.3. *Assume that $\mathbb{P}(\mathcal{C}^{(0)}(\boldsymbol{Y})\notin\mathbb{Z})=0$. The conditional spectral tail process $\boldsymbol{Q}$ is a random sequence with the distribution of $(\boldsymbol{Y}^*)^{-1}\boldsymbol{Y}$ conditionally on $\mathcal{C}^{(0)}(\boldsymbol{Y})=0$.*

The sequence $\boldsymbol{Q}$ appeared implicitly in the seminal paper [DH95]. See also [BS09], [PS18, Definition 3.5] and [KS20, Chapter 5]. An abstract setting is considered in [DHS18].

Note that $\mathcal{C}^{(0)}(\boldsymbol{Y})=0$ if and only if $\mathcal{C}^{(0)}(\boldsymbol{\Theta})=0$. Then also $\boldsymbol{Y}^*=|\boldsymbol{Y}_0|$. Thus, (2.5.3) and the definition of $\boldsymbol{Q}$ give for a bounded or non-negative measurable function H on $\ell_0(\mathbb{R}^d)$,

$$\boldsymbol{\nu}^*(H)=\vartheta\int_0^\infty\mathbb{E}\Big[H(r\boldsymbol{Q})\Big]\alpha r^{-\alpha-1}\mathrm{d}r=\vartheta\int_0^\infty\mathbb{E}\Big[H(r\boldsymbol{\Theta})\mathbb{1}\Big\{\mathcal{C}^{(0)}(\boldsymbol{\Theta})=0\Big\}\Big]\alpha r^{-\alpha-1}\mathrm{d}r\ . \tag{2.5.4}$$

If moreover H is such that $H(\boldsymbol{y})=0$ if $\boldsymbol{y}^*\le\epsilon$ for one $\epsilon>0$, then

$$\boldsymbol{\nu}^*(H)=\epsilon^{-\alpha}\mathbb{E}\Big[H(\epsilon\boldsymbol{Y})\mathbb{1}\Big\{\mathcal{C}^{(0)}(\boldsymbol{Y})=0\Big\}\Big]=\epsilon^{-\alpha}\mathbb{E}\Big[H(\epsilon\boldsymbol{Y})\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1}\le 1\big\}\Big]\ . \tag{2.5.5}$$

Note that with $H(\boldsymbol{x})=\mathbb{1}\{\boldsymbol{x}^*>1\}$ and recalling that $\boldsymbol{Y}^*>1$, (2.5.5) reduces to (2.5.2). As such, functionals from $\mathcal{A}$ will have typically the representation given in (2.5.5). On the other hand, for functionals from $\mathcal{B}$ we are not able to conclude the representation (2.5.5), however, the general form (2.5.4) is still valid, possibly under additional conditions. Comparing (2.5.3) or (2.5.5) with (2.2.1), we can see that the $\boldsymbol{\nu}^*(H)$ does not agree with $\mathbb{E}[H(\boldsymbol{Y})]$. The additional indicator comes essentially from the conditioning on the location of the maximum of the sequence $\boldsymbol{Y}$.

Definition 2.5.4 (Cluster index)**.** *We will call $\boldsymbol{\nu}^*(H)$ the cluster index associated to the functional H.*

2.5.1. Some computations for tail processes

We now state an important identity called the time-change formula, which will be used quite a lot in computations related to the tail process.

Theorem 2.5.5 (Time-change formula)**.** *Let $\boldsymbol{Y}$ be the tail process. Let H be bounded or non-negative measurable functional defined on $(\mathbb{R}^d)^{\mathbb{Z}}$. Then for $j\in\mathbb{Z}$ and $t>0$,*

$$\mathbb{E}\Big[H(tB\boldsymbol{Y})\mathbb{1}\{t|\boldsymbol{Y}_{-j}|>1\}\Big]=t^\alpha\mathbb{E}\Big[H(t\boldsymbol{Y})\mathbb{1}\{|\boldsymbol{Y}_j|>t\}\Big]\ ,$$

where B is the backshift operator.

In the context of the thesis, the following lemma is the most important application of the time-change formula. It indicates that the definition of the cluster measure does not depend on the choice of the anchoring map.

Lemma 2.5.6. *Assume that C and $\tilde{C}$ are anchoring maps. Then*

$$\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=0\Big)=\mathbb{P}\Big(\widetilde{\mathcal{C}}(\boldsymbol{Y})=0\Big).$$

Proof. By the time-change formula, we have

$$\begin{aligned}
\vartheta &= \mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=0\Big)=\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=0,\widetilde{\mathcal{C}}(\boldsymbol{Y})\in\mathbb{Z}\Big)=\sum_{j\in\mathbb{Z}}\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=0,\widetilde{\mathcal{C}}(\boldsymbol{Y})=j,|\boldsymbol{Y}_j|>1\Big)\\
&=\sum_{j\in\mathbb{Z}}\mathbb{P}\Big(\mathcal{C}(B^j\boldsymbol{Y})=0,\widetilde{\mathcal{C}}(B^j\boldsymbol{Y})=j,|\boldsymbol{Y}_{-j}|>1\Big)=\sum_{j\in\mathbb{Z}}\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=-j,\widetilde{\mathcal{C}}(\boldsymbol{Y})=0,|\boldsymbol{Y}_{-j}|>1\Big)\\
&=\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})\in\mathbb{Z},\widetilde{\mathcal{C}}(\boldsymbol{Y})=0\Big)=\mathbb{P}\Big(\widetilde{\mathcal{C}}(\boldsymbol{Y})=0\Big).
\end{aligned}$$

Hence

$$\mathbb{P}\Big(\mathcal{C}(\boldsymbol{Y})=0\Big)=\mathbb{P}\Big(\widetilde{\mathcal{C}}(\boldsymbol{Y})=0\Big).$$

□

2.6. Convergence of cluster measure

Let $|\cdot|$ be an arbitrary norm on $\mathbb{R}^d$ and $\{u_n\}$, $\{r_n\}$ be such that

$$\begin{aligned}
\lim_{n\to\infty}u_n=\lim_{n\to\infty}r_n=\lim_{n\to\infty}n\mathbb{P}(|\boldsymbol{X}_0|>u_n)=\infty\ ,\\
\lim_{n\to\infty}r_n/n=\lim_{n\to\infty}r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)=0\ .
\end{aligned}\tag{$\mathcal{R}(r_n,u_n)$}$$

We will call $\{r_n\}$ the intermediate sequence (of integers) and $\{u_n\}$ the scaling sequence.

Define the measures $\boldsymbol{\nu}^*_{n,r_n}$, $n\geq 1$, on $\ell_0(\mathbb{R}^d)$ as follows:

$$\boldsymbol{\nu}^*_{n,r_n}=\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[\delta_{u_n^{-1}\boldsymbol{X}_{1,r_n}}\right]\ .$$

We are interested in convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$. The results of this section are extracted from [KS20, Chapter 6]. See also [PS18] and [BPS18].

2.6.1. Anticlustering condition

For each fixed $r\in\mathbb{N}$, the distribution of $u_n^{-1}\boldsymbol{X}_{-r,r}$ conditionally on $|\boldsymbol{X}_0|>u_n$ converges weakly to the distribution of $\boldsymbol{Y}_{-r,r}$. In order to let r tend to infinity, we must embed all these finite vectors into one space of sequences. By adding zeroes on each side of the vectors $u_n^{-1}\boldsymbol{X}_{-r,r}$ and $\boldsymbol{Y}_{-r,r}$ we identify them with elements of the space $\ell_0(\mathbb{R}^d)$. Then $\boldsymbol{Y}_{-r,r}$ converges (as $r\to\infty$) to $\boldsymbol{Y}$ in $\ell_0(\mathbb{R}^d)$ if (and only if) $\boldsymbol{Y}\in\ell_0(\mathbb{R}^d)$ almost surely.

However, this is not enough for statistical purposes and we consider the following definition.

Definition 2.6.1 ([DH95], Condition 2.8)**.** *Condition $\mathcal{AC}(r_n,u_n)$ holds if for all $x,y>0$,*

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\max_{k\leq|j|\leq r_n}|\boldsymbol{X}_j|>u_nx\mid|\boldsymbol{X}_0|>u_ny\right)=0\ .\tag{$\mathcal{AC}(r_n,u_n)$}$$

Condition $\mathcal{AC}(r_n,u_n)$ is referred to as the anticlustering condition. It is fulfilled by many models, including geometrically ergodic Markov chains, short-memory linear or max-stable processes (see Section 2.7). $\mathcal{AC}(r_n,u_n)$ implies that $\boldsymbol{Y}\in\ell_0(\mathbb{R}^d)$. Its main consequence is the following result. Condition $\mathcal{AC}(r_n,u_n)$ holds for sequence of i.i.d. random variables whenever $\lim_{n\to\infty}r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)=0$, which can be recognized as on the restrictions imposed in $\mathcal{R}(r_n,u_n)$.

Proposition 2.6.2 ([BS09], Proposition 4.2; [KS20], Theorem 6.1.4)**.** *Let $H \in \mathcal{L}$. If Condition $\mathcal{AC}(r_n, u_n)$ holds, then*

$$\lim_{n\to\infty} \mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{-r_n,r_n}) \mid |\boldsymbol{X}_0| > u_n] = \mathbb{E}[H(\boldsymbol{Y})] \,.$$

2.6.2. Vague convergence

Vague convergence is nowadays a standard tool that is used in extreme value analysis in case of regular variation. There are several notions of vague convergence. We follow here, without providing too many technical details, Appendix B of [KS20]. Below, we present an *incomplete* definition of vague convergence. Recall that for a measure μ on a Polish space E, the notation $\mu(f)$ means $\mu(f) = \int_E f d\mu$.

Definition 2.6.3. *Let $\mu_n, n \geq 1$ be a sequence of boundedly fintie measure on a Polish space. We say that μ_n converges vaguely# to μ if $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ for all Lipschitz continuous functions with bounded support.*

For this definition to be fully understand, one needs to introduced a notion of boundedness and localization on a Polish space. This is rather technical and is thoroughly presented in Appendix B of [KS20]. For the purpose of the thesis we note that the test functions used in the above definition have the property that $f(x) = 0$ whenever $|x| < \epsilon$ for some $\epsilon > 0$, where $|\cdot|$ is an appropriate metric on the Polish space.

2.6.3. Vague convergence of cluster measure

We now investigate the unconditional convergence of $u_n^{-1}\boldsymbol{X}_{1,r_n}$. Contrary to Proposition 2.6.2, where an extreme value was imposed at time 0, a large value in the cluster can happen at any time. Moreover, the convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ to $\boldsymbol{\nu}^*(H)$ may hold only for shift-invariant functionals H. Therefore, we need the following definition.

Definition 2.6.4. *The space $\tilde{\ell}_0(\mathbb{R}^d)$ is the space of equivalence classes of $\ell_0(\mathbb{R}^d)$ endowed with the equivalence relation $\sim$ defined by*

$$\boldsymbol{x} \sim \boldsymbol{y} \Longleftrightarrow \exists j \in \mathbb{Z}\,,\ B^j\boldsymbol{x} = \boldsymbol{y}\,,$$

where B is the backshift operator.

Proposition 2.6.5. *Let condition $\mathcal{AC}(r_n, u_n)$ hold. The sequence of measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \geq 1$ converges vaguely$^\#$ on $\tilde{\ell}_0(\mathbb{R}^d) \setminus \{\boldsymbol{0}\}$ to $\boldsymbol{\nu}^*$, that is, for all $H \in \mathcal{A}$,*

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \boldsymbol{\nu}^*(H)\,. \tag{2.6.1}$$

The immediate consequence is the following limit:

$$\lim_{n\to\infty} \frac{\mathbb{P}(\boldsymbol{X}^*_{1,r_n} > u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \vartheta\,,$$

where ϑ is defined in (2.5.1). Since $H_{2,m}, H_{3,\eta} \in \mathcal{A}$, we can introduce the following cluster indices.

Example 2.6.6 (Cluster size distribution)*.* If $\mathcal{AC}(r_n, u_n)$ holds, Proposition 2.6.5 yields

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{j=1}^{r_n} \mathbb{1}\{|\boldsymbol{X}_j| > u_n\} = m \mid \boldsymbol{X}^*_{1,r_n} > u_n\right)$$

$$= \mathbb{P}\left(\sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{Y}_j| > 1\} = m \mid \boldsymbol{Y}^*_{-\infty,-1} \leq 1\right) =: \pi(m)\,.$$

⊞

Example 2.6.7 (Stop-loss index). Consider a univariate time series. Define the stop-loss index:

$$\theta_{\text{stoploss}}(\eta) = \lim_{n\to\infty} \frac{\mathbb{P}\left(\sum_{j=1}^{r_n}(X_j - u_n)_+ > \eta u_n\right)}{r_n\mathbb{P}(X_0 > u_n)} = \mathbb{P}\left(\sum_{j=0}^{\infty}(Y_j - 1)_+ > \eta, \boldsymbol{Y}^*_{-\infty,-1} \le 1\right) .$$

This index seems to be new. ⊞

2.6.4. Indicator functionals not vanishing around zero

Proposition 2.6.5 entails convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ for $H \in \mathcal{A}$. For functionals which are not defined on the whole space $\ell_0(\mathbb{R}^d)$, such as H_4 and H_5, we need an additional assumption on Asymptotic Negligibility of Small Jumps.

Definition 2.6.8. *Condition* ANSJB(r_n, u_n) *holds if for all* $\eta > 0$,

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} \frac{\mathbb{P}(\sum_{j=1}^{r_n} |\boldsymbol{X}_j|\mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0 . \tag{ANSJB(r_n,u_n)}$$

With this condition, we have the following cluster convergence. See Definition 2.5.3 for the $\boldsymbol{Q}$ sequence.

Lemma 2.6.9. *If* $\mathcal{AC}(r_n, u_n)$ *and* ANSJB(r_n, u_n) *hold, then*

$$\lim_{n\to\infty} \frac{\mathbb{P}(\sum_{i=1}^{r_n} |\boldsymbol{X}_j| > u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \mathbb{E}\left[\left(\sum_{j\in\mathbb{Z}} |\boldsymbol{Q}_j|\right)^{\alpha}\right] < \infty .$$

Proposition 2.6.10. *Assume that* $\mathcal{AC}(r_n, u_n)$ *and* ANSJB(r_n, u_n) *hold. Then for* $K \in \mathcal{K}$,

$$\boldsymbol{\nu}^*(\mathbb{1}\{K > 1\}) = \lim_{n\to\infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \vartheta \int_0^\infty \mathbb{P}(K(z\boldsymbol{Q}) > 1)\alpha z^{-\alpha-1}\mathrm{d}z < \infty .$$

If K is a 1-homogeneous satisfying the assumptions of Proposition 2.6.10, then

$$\boldsymbol{\nu}^*(\mathbb{1}\{K > 1\}) = \vartheta\mathbb{E}[K_+^\alpha(\boldsymbol{Q})] = \mathbb{E}[K_+^\alpha(\boldsymbol{\Theta}_{0,\infty}) - K_+^\alpha(\boldsymbol{\Theta}_{1,\infty})] .$$

Example 2.6.11 (Large deviations index). Let $\{X_j, j \in \mathbb{Z}\}$ be a univariate time series. The functional H_4 defined in (2.3.4) yields the large deviations index:

$$\theta_{\text{largedev}} = \lim_{n\to\infty} \frac{\mathbb{P}\left(\left(\sum_{j=1}^{r_n} X_j\right)_+ > u_n\right)}{r_n\mathbb{P}(|X_0| > u_n)} = \mathbb{E}\left[\left(\sum_{j=0}^{\infty}\Theta_j\right)_+^{\alpha} - \left(\sum_{j=1}^{\infty}\Theta_j\right)_+^{\alpha}\right] .$$

The index θ_{largedev}, under the name *cluster index*, was introduced in [MW16]. ⊞

Example 2.6.12 (Ruin index). Take H_5 defined in (2.3.5). Proposition 2.6.10 gives

$$\theta_{\text{ruin}} = \lim_{n\to\infty} \frac{\mathbb{P}(\max_{1\le j\le r_n}\sum_{i=1}^{j} X_i > u_n)}{r_n\mathbb{P}(|X_0| > u_n)} = \vartheta\mathbb{E}\left[\sup_{i\in\mathbb{Z}}\left(\sum_{j\le i} Q_j\right)_+^{\alpha}\right] .$$

⊞

Remark 2.6.13. At this moment we would like to point out the following. Consider $H \in \mathcal{A}$ to be an indicator functional. Moreover, if H is such that $H(\boldsymbol{y}) = 0$ if $\boldsymbol{y}^* \le 1$, then from (2.5.5),

$$\boldsymbol{\nu}^*(H) = \mathbb{E}[H(\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \in (0,1]\ .$$

This is the situation for the extremal index and the functionals from Examples 2.6.6 and 2.6.7. On the other hand, if H does not vanish around zero, then at the first place we need additional conditions to guarantee that $\boldsymbol{\nu}^*(H) < \infty$ (e.g. ANSJB(r_n, u_n)). Second, there is no restriction on the values of the cluster index.

2.6.5. Unbounded functionals

This section extends the previous results to the convergence of unbounded functionals satisfying certain moment conditions. However, different scenarios can happen if we consider general unbounded functionals. In particular, consider the summation functional $G(\boldsymbol{x}) = \sum_{j\in Z} \phi\Big(\boldsymbol{x}_j/u_n\Big)$ with ϕ vanishing around zero. Then the cluster index will simplify.

Proposition 2.6.14. *Let condition $\mathcal{AC}(r_n,u_n)$ hold. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be such that $\phi(\boldsymbol{x}_0) = 0$ if $|\boldsymbol{x}_0| \le \epsilon$ for some $\epsilon > 0$ and there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} \frac{\mathbb{E}[|\phi|^{1+\delta}(\boldsymbol{X}_0/u_n)]}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} < \infty. \tag{2.6.2}$$

Let H be defined on $\ell_1(\mathbb{R}^d)$ such that $|H(\boldsymbol{x})| \le cst \sum_{j\in\mathbb{Z}} \phi(\boldsymbol{x}_j)$, where cst is a constant. Then

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \boldsymbol{\nu}^*(H)\ . \tag{2.6.3}$$

Proof of Proposition 2.6.14. Note that $G(\boldsymbol{x}) = \sum_{j\in Z} \phi\Big(\boldsymbol{x}_j/u_n\Big)$ is not bounded. The idea is to use the weak convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$ to conclude (2.6.3). However, this is not possible directly since H is not bounded, so we need the truncation argument. Let $M \ge 1$. Then $\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H \wedge M) = \boldsymbol{\nu}^*(H \wedge M)$ and $\lim_{M\to\infty} \boldsymbol{\nu}^*(H\wedge M) = \boldsymbol{\nu}^*(H)$ by monotone convergence . Therefore, the convergence of (2.6.3) will be, by the triangular argument, a consequence of

$$\lim_{M\to\infty}\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H\mathbb{1}\{H > M\}) = 0.$$

Assume without loss of generality that $\phi \ge 0$, we have by stationarity and shift-invariance,

$$\frac{\mathbb{E}\Big[H(u_n^{-1}\boldsymbol{X}_{1,r_n})\mathbb{1}\{H(u_n^{-1}\boldsymbol{X}_{1,r_n}) > M\}\Big]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \le \text{cst}\frac{\mathbb{E}\Big[\phi(u_n^{-1}\boldsymbol{X}_0)\mathbb{1}\Big\{\sum_{j=-r_n}^{r_n}\phi(u_n^{-1}\boldsymbol{X}_j) > M\Big\}\Big]}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\ .$$

By Hölder inequality and since ϕ vanishes around zero, there exists $\epsilon > 0$ such that

$$\begin{aligned}
&\frac{\mathbb{E}\Big[\phi(u_n^{-1}\boldsymbol{X}_0)\mathbb{1}\Big\{\sum_{j=-r_n}^{r_n}\phi(u_n^{-1}\boldsymbol{X}_j) > M\Big\}\Big]}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\
&\le \frac{\mathbb{E}\Big[\phi(u_n^{-1}\boldsymbol{X}_0)\mathbb{1}\{|\boldsymbol{X}_0| > u_n\epsilon\}\mathbb{1}\Big\{\sum_{j=-r_n}^{r_n}\phi(u_n^{-1}\boldsymbol{X}_j) > M\Big\}\Big]}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\
&\le \left(\frac{\mathbb{E}\Big[\phi^{1+\delta}(u_n^{-1}\boldsymbol{X}_0)\Big]}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\right)^{\frac{1}{1+\delta}} \left(\frac{\mathbb{P}\Big(|\boldsymbol{X}_0| > u_n\epsilon, \sum_{j=-r_n}^{r_n}\phi(u_n^{-1}\boldsymbol{X}_j) > M\Big)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\right)^{\frac{\delta}{1+\delta}}.
\end{aligned}$$

Since $\mathcal{AC}(r_n, u_n)$ holds, we have by Proposition 2.6.2,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\boldsymbol{X}_0 > u_n\epsilon, \sum_{j=-r_n}^{r_n} \phi(u_n^{-1}\boldsymbol{X}_j) > M)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \epsilon^{-\alpha}\mathbb{P}\left(\sum_{j\in\mathbb{Z}} \phi(\epsilon\boldsymbol{Y}_j) > M\right).$$

The last term tends to zero as $M \to \infty$ by the finiteness of the sum $\sum_{j\in\mathbb{Z}} \phi(\epsilon\boldsymbol{Y}_j)$. Hence, the proof is concluded on account of condition (2.6.2). $\square$

Lemma 2.6.15. *Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be non-negativesuch that $\mathbb{E}[\phi(\boldsymbol{Y}_0)] < \infty$. Then we have*

$$\sum_{j\in\mathbb{Z}} \mathbb{E}\Big[\phi(\boldsymbol{Y}_j)\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\mathbb{1}\{|\boldsymbol{Y}_j| > 1\}\Big] = \mathbb{E}[\phi(\boldsymbol{Y}_0)]\,.$$

Proof. Recall the time-change formula

$$\mathbb{E}[G(\boldsymbol{Y})\mathbb{1}\big\{|\boldsymbol{Y}_j| > 1\big\}] = \mathbb{E}[G(B^j\boldsymbol{Y})\mathbb{1}\big\{|\boldsymbol{Y}_{-j}| > 1\big\}],$$

for a bounded or non-negative measurable functional G on $(\mathbb{R}^d)^{\mathbb{Z}}$. Application of the time-change formula with $G(\boldsymbol{x}) = \phi(\boldsymbol{x}_j)\mathbb{1}\big\{\boldsymbol{x}^*_{-\infty,-1} \le 1\big\}$ yields,

$$\begin{aligned}
&\sum_{j\in\mathbb{Z}} \mathbb{E}\Big[\phi(\boldsymbol{Y}_j)\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\mathbb{1}\big\{|\boldsymbol{Y}_j| > 1\big\}\Big] \\
&= \sum_{j\in\mathbb{Z}} \mathbb{E}\Big[\phi(B^j\boldsymbol{Y}_j)\mathbb{1}\big\{B^j\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\mathbb{1}\big\{|\boldsymbol{Y}_{-j}| > 1\big\}\Big] \\
&= \sum_{j\in\mathbb{Z}} \mathbb{E}\Big[\phi(\boldsymbol{Y}_0)\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1-j} \le 1\big\}\mathbb{1}\big\{|\boldsymbol{Y}_{-j}| > 1\big\}\Big] = \mathbb{E}\Big[\phi(\boldsymbol{Y}_0)\mathbb{1}\{\boldsymbol{Y}^* > 1\}\Big] = \mathbb{E}[\phi(\boldsymbol{Y}_0)].
\end{aligned}$$

$\square$

Examples

If $\mathcal{AC}(r_n, u_n)$ holds, Proposition 2.6.14 yields the followings:

Example 2.6.16. Consider a univariate time series. Then

- for $H_5(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}}(x_j - 1)_+\mathbb{1}\{\boldsymbol{x}^* > 1\}$, we have

$$\lim_{n\to\infty} \frac{\mathbb{E}\left(\sum_{j=1}^{r_n}(X_j - u_n)_+\mathbb{1}\big\{\boldsymbol{X}^*_{1,r_n} > u_n\big\}\right)}{u_n r_n \mathbb{P}(|X_0| > u_n)} = \mathbb{E}\left(\sum_{j=0}^{\infty}(Y_j - 1)_+\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\right). \quad (2.6.4)$$

  We can further simplify the expression for H_5 thanks to Lemma 2.6.15

$$\begin{aligned}
&\mathbb{E}\left(\sum_{j=0}^{\infty}(Y_j - 1)_+\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\right) \\
&= \mathbb{E}\left(\sum_{j\in\mathbb{Z}}(Y_j - 1)_+\mathbb{1}\big\{|Y_j| > 1\big\}\mathbb{1}\big\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\big\}\right) \\
&= \mathbb{E}[(Y_0 - 1)_+] = \mathbb{E}[Y_0] - 1 = \frac{\alpha}{\alpha - 1} - 1 = \frac{1}{\alpha - 1}\,.
\end{aligned}$$

  provided that

$$\lim_{n\to\infty} \frac{\mathbb{E}\left((X_0 - u_n)_+^{1+\delta}\right)}{u_n^{1+\delta}\mathbb{P}(|X_0| > u_n)} < \infty\,,$$

  The latter holds if $\alpha > 1$.

- for $H_6(\boldsymbol{x}) = \max_{j\in\mathbb{Z}}(x_j - 1)_+ \mathbb{1}\{\boldsymbol{x}^* > 1\}$, we have

$$\lim_{n\to\infty} \frac{\mathbb{E}\left(\max_{1\le j\le r_n}(X_j - u_n)_+ \mathbb{1}\{\boldsymbol{X}^*_{1,r_n} > u_n\}\right)}{u_n r_n \mathbb{P}(|X_0| > u_n)} = \mathbb{E}\left(\max_{j\ge 0}(Y_j - 1)_+ \mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}\right).$$

⊞

Example 2.6.17. If we take $H(\boldsymbol{x}) = \sum_j |\boldsymbol{x}_j|^\eta \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$, $\eta < \alpha$, then by direct computation, we have

$$\begin{aligned}\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) &= \lim_{n\to\infty} \frac{\mathbb{E}[\sum_{j=1}^{r_n} |\boldsymbol{X}_j|^\eta \, \mathbb{1}\{|\boldsymbol{X}_j| > u_n\}]}{u_n^\eta r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\ &= \lim_{n\to\infty} \frac{\mathbb{E}[|\boldsymbol{X}_0|^\eta \, \mathbb{1}\{\boldsymbol{X}_0 > u_n\}]}{u_n^\eta \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \frac{\alpha}{\alpha - \eta} = \mathbb{E}[|\boldsymbol{Y}_0|^\eta] \,. \end{aligned} \tag{2.6.5}$$

We can also obtain the same result by application of Proposition 2.6.14. Indeed, it implies that

$$\begin{aligned}\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) &= \lim_{n\to\infty} \frac{\mathbb{E}[\sum_{j=1}^{r_n} |\boldsymbol{X}_j|^\eta \, \mathbb{1}\{|\boldsymbol{X}_j| > u_n\}]}{u_n^\eta r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\ &= \mathbb{E}[\sum_{j\in\mathbb{Z}} |\boldsymbol{Y}_j|^\eta \mathbb{1}\{|\boldsymbol{Y}_j| > 1\} \mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \,. \end{aligned} \tag{2.6.6}$$

It is not immediately obvious that the expressions in (2.6.5)-(2.6.6) agree. See Lemma 2.6.15 for the proof.

Hence, application of Lemma 2.6.15 to Example 2.6.17 for $H(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} |\boldsymbol{x}_j|^\eta \, \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$ with $\eta < \alpha$ gives

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \mathbb{E}\left(\sum_{j=0}^{\infty} |\boldsymbol{Y}_j|^\eta \, \mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\} \mathbb{1}\{|\boldsymbol{Y}_j| > 1\}\right) = \mathbb{E}[|\boldsymbol{Y}_0|^\eta] = \frac{\alpha}{\alpha - \eta} \,.$$

The discussion above deals with functionals dominated by $\sum_{j\in Z} \phi(\boldsymbol{x}_j)$ with ϕ vanishing around zero. However, for general unbounded functionals, we have different scenarios. Below, some unbounded functionals are considered.

Example 2.6.18. If we take $H(\boldsymbol{x}) = \sum_j |\boldsymbol{x}_j|^\eta$, $\eta < \alpha$, then $\mathbb{E}[|\boldsymbol{X}_0|^\eta] < \infty$. Note that Proposition 2.6.14 is not directly applicable, because $\phi(\boldsymbol{x}_i) = |\boldsymbol{x}_i|^\eta$ does not vanish around zero. We have:

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[\sum_{j=1}^{r_n} |\boldsymbol{X}_j|^\eta]}{u_n^\eta r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \lim_{n\to\infty} \frac{\mathbb{E}[|\boldsymbol{X}_0|^\eta]}{u_n^\eta \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \infty.$$

⊞

For the following example, the computations are done for the i.i.d. case to illustrate that the limit does not depend entirely on the process $\boldsymbol{Y}$.

Example 2.6.19. Take $H(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\} \sum_j |\boldsymbol{x}_j|^\eta$. It is important to note that H vanishes around $\boldsymbol{0} \in (\mathbb{R}^d)^\mathbb{Z}$ (because of the indicator $\mathbb{1}\{\boldsymbol{x}^* > 1\}$), but ϕ does not vanish around $\boldsymbol{0} \in \mathbb{R}^d$, hence Proposition 2.6.14 is not applicable.

By the stationarity and shift-invariance we have

$$
\begin{aligned}
\lim_{n\to\infty}\boldsymbol{\nu}^*_{n,r_n}(H) &= \lim_{n\to\infty}\frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} = \lim_{n\to\infty}\frac{\mathbb{E}[\sum_{i=1}^{r_n}|\boldsymbol{X}_i|^\eta u_n^{-\eta}\mathbb{1}\{\boldsymbol{X}^*_{1,r_n}>u_n\}]}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\\
&= \lim_{n\to\infty}\frac{\sum_{j=1}^{r_n}\mathbb{E}[\sum_{i=1}^{r_n}|\boldsymbol{X}_i|^\eta u_n^{-\eta}\mathbb{1}\{\boldsymbol{X}^*_{1,j-1}\le u_n\}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}]}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\\
&= \frac{1}{r_n}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1}\le u_n\}\sum_{i=1-j}^{r_n-j}|\boldsymbol{X}_i|^\eta\mid|\boldsymbol{X}_0|>u_n\right]\\
&= \frac{1}{r_n}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1}\le u_n\}\sum_{i=1-j}^{-1}|\boldsymbol{X}_i|^\eta\mid|\boldsymbol{X}_0|>u_n\right]\\
&+ \frac{1}{r_n}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1}\le u_n\}\sum_{i=0}^{r_n-j}|\boldsymbol{X}_i|^\eta\mid|\boldsymbol{X}_0|>u_n\right]\\
&= I+II
\end{aligned}
$$

The expression above are rather hard to evaluate in general. If the time series is i.i.d., then II implies

$$
\begin{aligned}
&\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1}\le u_n\}\sum_{i=0}^{r_n-j}|\boldsymbol{X}_i|^\eta\,\mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right]\\
&= \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{P}\left(\boldsymbol{X}^*_{1-j,-1}\le u_n\right)\mathbb{E}[|\boldsymbol{X}_0|^\eta\,\mathbb{1}\{|\boldsymbol{X}_0|>u_n\}]\\
&\quad+\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{r_n}u_n^{-\eta}\mathbb{P}\left(\boldsymbol{X}^*_{1-j,-1}\le u_n\right)\sum_{i=1}^{r_n-j}\mathbb{E}[|\boldsymbol{X}_i|^\eta]\mathbb{P}(|\boldsymbol{X}_0|>u_n)\\
&= \frac{1}{r_n}\sum_{j=1}^{r_n}\mathbb{P}\left(\boldsymbol{X}^*_{1-j,-1}\le u_n\right)\frac{1}{u_n^\eta\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}[|\boldsymbol{X}_0|^\eta\,\mathbb{1}\{|\boldsymbol{X}_0|>u_n\}]\\
&\quad+\mathbb{E}[|\boldsymbol{X}_0|^\eta]\frac{1}{u_n^\eta r_n}\sum_{j=1}^{r_n}(r_n-j)\mathbb{P}\left(\boldsymbol{X}^*_{1-j,-1}\le u_n\right)
\end{aligned}
$$

In the last part set for a moment $u_n^\eta=r_n$. The last term becomes

$$
\mathbb{E}[|\boldsymbol{X}_0|^\eta]\frac{1}{r_n}\sum_{j=1}^{r_n}(1-j/r_n)\mathbb{P}\left(\boldsymbol{X}^*_{1-j,-1}\le u_n\right)
$$

Therefore, for $\eta<\alpha$ and $u_n^\eta=r_n$, we have

$$
\begin{aligned}
\lim_{n\to\infty}II &= \lim_{n\to\infty}\frac{\mathbb{E}[|\boldsymbol{X}_0|^\eta\,\mathbb{1}\{|\boldsymbol{X}_0|>u_n\}]}{u_n^\eta\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\int_0^1\mathbb{P}\left(\boldsymbol{X}^*_{1-[r_ns],-1}\le u_n\right)ds\\
&+\lim_{n\to\infty}\mathbb{E}[|\boldsymbol{X}_0|^\eta]\int_0^1(1-s)\mathbb{P}\left(\boldsymbol{X}^*_{1-[r_ns],-1}\le u_n\right)ds\\
&= \frac{\alpha}{\alpha-\eta}\vartheta+\frac{\vartheta}{2}\mathbb{E}[|\boldsymbol{X}_0|^\eta].
\end{aligned}
$$

Note that under i.i.d. assumption $\vartheta = 1$. The first part (again i.i.d. assumption) and $u_n^\eta = r_n$:

$$\begin{aligned}
\lim_{n\to\infty} I &= \lim_{n\to\infty} \frac{1}{r_n}\sum_{j=1}^{r_n} u_n^{-\eta}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1} \le u_n\} \sum_{i=1-j}^{-1} |\boldsymbol{X}_i|^\eta \mid |\boldsymbol{X}_0| > u_n\right] \\
&= \lim_{n\to\infty} \frac{1}{r_n}\sum_{j=1}^{r_n} u_n^{-\eta}\mathbb{E}\left[\sum_{i=1-j}^{-1} |\boldsymbol{X}_i|^\eta\, \mathbb{1}\{|\boldsymbol{X}_i| \le u_n\} \mid |\boldsymbol{X}_0| > u_n\right] \\
&= \lim_{n\to\infty} \frac{1}{r_n}\sum_{j=1}^{r_n} u_n^{-\eta}\mathbb{E}\left[\sum_{i=1-j}^{-1} |\boldsymbol{X}_i|^\eta\, \mathbb{1}\{|\boldsymbol{X}_i| \le u_n\}\right] \\
&= \lim_{n\to\infty} u_n^{-\eta}\mathbb{E}\left[|\boldsymbol{X}_0|^\eta\, \mathbb{1}\{|\boldsymbol{X}_0| \le u_n\}\right] \frac{1}{r_n}\sum_{j=1}^{r_n}(j-1) \\
&= \lim_{n\to\infty} \mathbb{E}[|\boldsymbol{X}_0|^\eta]\frac{r_n}{2u_n^\eta} = \frac{1}{2}\mathbb{E}[|\boldsymbol{X}_0|^\eta], \quad \text{as}\,.
\end{aligned}$$

Therefore, if $\eta < \alpha$ and $u_n^\eta = r_n$, under i.i.d. condition, we have

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \frac{\alpha}{\alpha-\eta} + \frac{1}{2}\mathbb{E}[|\boldsymbol{X}_0|^\eta] + \frac{1}{2}\mathbb{E}[|\boldsymbol{X}_0|^\eta] = \frac{\alpha}{\alpha-\eta} + \mathbb{E}[|\boldsymbol{X}_0|^\eta].$$

⊞

2.7. Examples of time series

In what follows, we present several examples of time series, which fulfill the anticlustering as well as beta-mixing conditions required for asymptotic theory in the following chapters. The precise statements in terms of each model parameters can be found in [KS20].

2.7.1. Moving averages

Consider the following linear process [KS20, Chapter 15]

$$X_j = \sum_{i\in\mathbb{Z}} \Psi_i Z_{j-i}, \quad j \in \mathbb{Z}, \tag{2.7.1}$$

where $\{Z_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. regularly varying random variables with tail index α. If $\alpha > 1$ then we assume that $\mathbb{E}[Z_0] = 0$. If $\Psi_i = 0$ for $i \le -1$, then the linear process is called one sided or causal. This process is well defined under the summability condition i.e. there exist $\delta \in (0,\alpha) \cap (0,2]$ such that

$$\sum_{i\in\mathbb{Z}} |\Psi_i|^\delta < \infty\ , \tag{2.7.2}$$

which implies that

$$\sum_{i\in\mathbb{Z}} |\Psi_i|^\alpha < \infty. \tag{2.7.3}$$

Furthermore, if $\alpha \in (0,1]$, then the above condition implies that

$$\sum_{i\in\mathbb{Z}} |\Psi_i| < \infty.$$

If $\alpha > 1$, it may happen that there exists $\delta \in (1, 2]$ such that (2.7.2) holds but $\sum_{i\in\mathbb{Z}} |\Psi_i| = \infty$. For e.g. if there exists $c > 0$ such that $\Psi_j = (1 + |j|)^{-c}$, then (2.7.2) holds if $c\alpha > 1$, but (2.7.3) holds only if $c > 1$.

Under (2.7.2), we have thanks to [KS20, Collorary 4.2.1],

$$\mathbb{P}(|X_0| > x) \sim \|\Psi\|_\alpha^\alpha \mathbb{P}(|Z_0| > x),$$

where

$$\|\Psi\|_\alpha^\alpha = \sum_{i\in\mathbb{Z}} \Big(|\Psi_i|^\alpha\Big)^{\frac{1}{\alpha}}.$$

The candidate extremal index is

$$\vartheta = \frac{\sup_{j\in\mathbb{Z}} |\Psi_j|^\alpha}{\sum_{i\in\mathbb{Z}} |\Psi_i|^\alpha}.$$

Under the mild conditions on the coefficients Ψ_j the anticlustering conditions holds. Additionally, if the density of the innovations Z_j exists and satisfies the appropriate integrability condition, the moving average is beta-mixing. See [KS20, Section 15.3].

In particular, for AR(1) process, defined recursively by $X_{j+1} = \rho X_j + Z_{j+1}$, $\rho \in (0, 1)$, we have

$$\mathbf{\Theta}_j = \rho^j (sign(\rho))^N \mathbf{\Theta}_0 \mathbb{1}\{j \geq -N\}, \quad j \in \mathbb{Z},$$

where $\mathbb{P}(\mathbf{\Theta}_0 = 1) = 1 - \mathbb{P}(\mathbf{\Theta}_0 = -1) = p$, the extremal skewness of X_0 and N is an integer valued random variables, independent of $\mathbf{\Theta}_0$, such that

$$\mathbb{P}(N = j) = (1 - |j|^\alpha)^{|j|\alpha}.$$

The extremal index of the series $\{X_j, j \in \mathbb{Z}\}$ is $\theta = 1 - \rho^\alpha$.

2.7.2. Markov chains

Let $\{(\boldsymbol{A}_j, \boldsymbol{B}_j),\ j \in \mathbb{Z}\}$ be a sequence of i.i.d. random pairs in $\mathbb{R}^{d\times d} \times \mathbb{R}^d$, independent of d-dimensional random vector $\boldsymbol{X}_0$. Define $\mathbb{R}^d$-valued process $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ recursively by $\boldsymbol{X}_0$ and

$$\boldsymbol{X}_{j+1} = \boldsymbol{A}_{j+1}\boldsymbol{X}_j + \boldsymbol{B}_{j+1},\ \ j \geq 0. \tag{2.7.4}$$

We say that $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ satisfies a stochastic recurrence equation (SRE). It is a Markov chain with transition kernel $\mathit{\Pi}$ given by

$$\mathit{\Pi}(\boldsymbol{x}, \cdot) = \mathbb{P}(A_0\boldsymbol{x} + \boldsymbol{B}_0 \in \cdot).$$

Stochastic recurrence equation with heavy tailed innovation

Assume that $\boldsymbol{X}_0 \stackrel{d}{=} \boldsymbol{A}_1\boldsymbol{X}_0 + \boldsymbol{B}_1$. By iteration, the stationary solution (if exists) can be written as

$$\boldsymbol{X}_0 = \boldsymbol{B}_0 + \sum_{i=1}^{\infty} \Big(\prod_{k=0}^{i-1} \boldsymbol{A}_{-k}\Big)\boldsymbol{B}_{-i}\ . \tag{2.7.5}$$

We consider that $\boldsymbol{B}_0$ is heavier tailed than $\boldsymbol{A}_0$, meaning that $\boldsymbol{B}_0$ is regularly varying with tail index $\alpha > 0$, $\mathbb{E}(\|\boldsymbol{A}_0\|^\alpha) < 1$ and there exists $\epsilon > 0$ such that $\mathbb{E}(\|\boldsymbol{A}_0\|^{\alpha+\epsilon}) < \infty$. Under these conditions, $\boldsymbol{X}_0$ is regularly varying thanks to [KS20, Theorem 4.1.2].

Tail process: The forward spectral tail process is the multiplicative random walk $\mathbf{\Theta}$ given by $\mathbf{\Theta}_0$ and

$$\mathbf{\Theta}_j = \boldsymbol{A}_j \cdots \boldsymbol{A}_1 \mathbf{\Theta}_0,\ j \geq 1.$$

Note that $\lim_{j\to\infty} \mathbf{\Theta}_j = \mathbf{0}$ by the law of large number since $\mathbb{E}[\log \|\boldsymbol{A}_0\|] < 0$. For simplicity, we describe the backward tail process when $d = 1$ and $A_0 \geq 0$. In that case, the forward tail process never vanishes if $\mathbb{P}(A_0 = 0) = 0$, whereas

$$\mathbb{P}(\Theta_{-j} = 0) = 1 - \mathbb{E}[|\Theta_j|^\alpha] = 1 - (\mathbb{E}[A_0^\alpha])^j.$$

Still when $d = 1$ and $A_0 \geq 0,\ B_0 \geq 0$, the candidate extremal index is given by

$$\vartheta = \mathbb{E}\Big[\left(1 - \max_{j\geq 1} \prod_{i=1}^{j} A_i^\alpha \right)_+ \Big].$$

Stochastic recurrence equation with light tailed innovation

Consider the equation (2.7.4) for $d = 1$, Assume that $\{(A_j, B_j),\ j \in \mathbb{Z}\}$ is a sequence of i.i.d. random vectors, $\mathbb{P}(A_0 \geq 0) = 1$, the law of $\log(A_0)$ conditionally on $A_0 \geq 0$ is non-arithmetic and

$$\mathbb{E}[A_0^\alpha] = 1,\ \mathbb{E}[A_0^\alpha \log_+(A_0)] < \infty,\ \mathbb{E}[|B_0|^\alpha] < \infty.$$

Note that $\mathbb{E}[A_0^\alpha] = 1$ implies (unless $A_0 \equiv 0$) that $\mathbb{P}(A_0 < 1) > 0$ and $\mathbb{P}(A_0 > 1) > 0$. The first condition guarantees the existence of the stationary solution, while the second one is necessary for heavy tails of that stationary solution when $\mathbb{E}[|B_0|^\alpha] < \infty$ holds. Then the series in (2.7.5) is almost surely absolutely summable and

$$\begin{aligned}
\lim_{x\to\infty} x^\alpha \mathbb{P}(X > x) &= \frac{\mathbb{E}[(A_0X_0 + B_0)_+^\alpha - (A_0X_0)_+^\alpha]}{\alpha\mathbb{E}[A_0^\alpha \log(A_0)]} \\
\lim_{x\to\infty} x^\alpha \mathbb{P}(X < -x) &= \frac{\mathbb{E}[(A_0X_0 + B_0)_-^\alpha - (A_0X_0)_-^\alpha]}{\alpha\mathbb{E}[A_0^\alpha \log(A_0)]}
\end{aligned}$$

Under these conditions $\{X_j, j \in \mathbb{Z}\}$ is regularly varying and its forward tail stationary process is given by

$$\Theta_j = A_1 \cdots A_j \Theta_0,\ j \geq 1.$$

The candidate extremal index is the same as in the previous section, that is

$$\vartheta = \mathbb{E}\left[\left(1 - \max_{j\geq 1} \prod_{i=1}^{j} A_i^\alpha \right)_+ \right].$$

In the present framework, the most known examples are ARCH(1) and GARCH(1,1) processes. Let $a_0 > 0$ and $a_1 \geq 0$ and assume that $\{\epsilon_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables. An ARCH(1) process $\{X_j, j \in \mathbb{Z}\}$ satisfies the recursion

$$\begin{aligned}
X_j &= \sigma_j \epsilon_j, \\
\sigma_j^2 &= a_0 + a_1 X_{j-1}^2.
\end{aligned}$$

Then

$$\sigma_j^2 = a_0 + \epsilon_{j-1}^2 X_{j-1}^2. \tag{2.7.6}$$

This is the recursion (2.7.4) with

$$A_j = a_1 \epsilon_{j-1}^2, \quad B_0 = a_0.$$

Under the assumptions of [KS20, Theorem 14.1.1], the stationary solution to (2.7.6) is regularly varying with index $\alpha/2$ given by

$$\mathbb{E}[(a_1 \epsilon_0^\alpha)^{\alpha/2}] = 1.$$

If $\mathbb{E}[|\epsilon_0|^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, then X_0 is regularly varying with index α by Breiman's lemma (see [KS20, Lemma 1.4.3]).

A GARCH(1,1) process $\{X_j, j \in \mathbb{Z}\}$ satisfies the recursion

$$\begin{aligned} X_j &= \sigma_j \epsilon_j, \\ \sigma_j^2 &= a_0 + a_1 X_{j-1}^2 + b_1 \sigma_{j-1}^2 = a_0 + (b_1 + a_1 \epsilon_{j-1}^2)\sigma_{j-1}^2, \end{aligned}$$

where $a_0, b_1 > 0$ and $a_1 \geq 0$. Define

$$A_j = b_1 + a_1 \epsilon_{j-1}^2, \quad B_0 = a_0.$$

Then under the assumptions of [KS20, Theorem 14.1.1], the stationary solution to (2.7.6) is regularly varying with index $\alpha/2 > 0$ if

$$\mathbb{E}[(b_1 + a_1 \epsilon_0^\alpha)^{\alpha/2}] = 1.$$

We can conclude that X_0 is regularly varying with index α provided that if $\mathbb{E}[|\epsilon_0|^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$ thanks to Breiman's lemma.

Under very general conditions, Markov chains fulfill the anticlustering condition and are beta-mixing with rates decaying at the geometric speed.

Chapter 3

Consistency of disjoint blocks estimators

This section summarizes consistency for the disjoint blocks estimators. The theory is established in [KS20, Chapter 10]. All references can be found at the end of the thesis, on page 131. Notation and preliminary background can be found in Chapter 2.

3.1. Disjoint block estimators

Several methods of estimation of the limit $\boldsymbol{\nu}^*(H)$ in (2.6.1) may be employed. The natural one is to consider a statistics based on disjoint blocks of size r_n:

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/u_n) \ ,$$

where $m_n = [n/r_n]$. The following section summaries results on the convergence of the disjoint blocks estimators under mild weak dependence conditions.

3.2. Consistency of disjoint blocks estimators

Let r_n be an intermediate sequence and u_n be a scaling sequence. Define the measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \geq 1$, on $\ell_0(\mathbb{R}^d)$ as follows:

$$\boldsymbol{\nu}^*_{n,r_n} = \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[\delta_{u_n^{-1}\boldsymbol{X}_{1,r_n}}\right] \ ,$$

with its empirical version given by

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n} := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} \delta_{(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/u_n)} \ .$$

We call $\widetilde{\boldsymbol{\nu}}^*_{n,r_n}$ the empirical cluster measure. Of interest is the convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$. The first result establishes consistency of the empirical cluster measure $\widetilde{\boldsymbol{\nu}}^*_{n,r_n}$. In what follows, $\xrightarrow{v^{\#}}$ denotes the vague convergence (see Proposition 2.6.5), while $\xrightarrow{w}$ denotes weak convergence of probability measures.

Lemma 3.2.1. *Assume that* $\mathcal{R}(r_n, u_n)$ *and* $\widetilde{\boldsymbol{\nu}}^*_{n,r_n} \xrightarrow{v^\#} \boldsymbol{\nu}^*$ *hold. Then*

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n} \overset{w}{\Longrightarrow} \boldsymbol{\nu}^* \quad \text{in } \mathcal{M}(\tilde{\ell}_0(\mathbb{R}^d)) \tag{3.2.1}$$

if and only if

$$\lim_{n\to\infty} \left| \mathbb{E}[e^{-\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H)}] - \left(\mathbb{E}[e^{-\boldsymbol{\nu}^*_{n,r_n}(H)}]\right)^{m_n} \right| = 0, \tag{3.2.2}$$

for all shift-invariant, bounded Lipschitz continuous functions H *defined on* $\ell_0(\mathbb{R}^d)$ *with support separated form zero.*

The following two theorems are important in a sense that provide weak convergence of the empirical measures under two different approaches. The first Theorem 3.2.2 is established using the beta-mixing condition

Theorem 3.2.2. *Let* $\{\boldsymbol{X}_j,\, j \in \mathbb{Z}\}$ *be a stationary, regularly varying* $\mathbb{R}^d$*-valued time series. Assume that* $\mathcal{R}(r_n, u_n)$*,* $\mathcal{AC}(r_n, u_n)$ *hold and that the sequence is* β*-mixing with the rates*

$$\lim_{n\to\infty} \frac{n}{r_n} \beta_{\ell_n} = 0\ , \tag{3.2.3}$$

where $\ell_n \to \infty$ *such that* $\ell_n / r_n \to 0$*. Then* (3.2.1) *holds.*

The second theorem relies on the m-dependent approximation.

Theorem 3.2.3. *Let* $\{\boldsymbol{X}_j,\, j \in \mathbb{Z}\}$ *be a stationary, regularly varying* $\mathbb{R}^d$*-valued time series with tail process* $\boldsymbol{Y}$ *such that* $\mathbb{P}(\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)) = 1$*. Assume that* $\mathcal{R}(r_n, u_n)$ *holds. Let* $\{\boldsymbol{X}_j^{(m)},\, j \in \mathbb{Z}\}$*,* $m \geq 1$*, be regularly varying* m*-dependent time series such that* $\{(\boldsymbol{X}_j, \boldsymbol{X}_j^{(m)}),\, j \in \mathbb{Z}\}$ *is stationary for all* $m \geq 1$ *and for all* $\epsilon > 0$*,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{\mathbb{P}(|\boldsymbol{X}_0 - \boldsymbol{X}_0^{(m)}| > u_n \epsilon)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0. \tag{3.2.4}$$

Then (3.2.1) *holds.*

Chapter 4

Consistency of sliding blocks estimators

This chapter consists of the paper titled *Consistency of sliding blocks estimators*. This paper is available in the preprint form. For technical reasons, we put it in the format of the *Electronic Journal of Statistics*, matching the format of the other two papers discussed later. The paper is self-contained, with all the required definitions. The references can be found at the end of the paper, on page 33.

Electronic Journal of Statistics

Consistency of sliding blocks estimators

Youssouph Cissokho*

Abstract: Cluster indices describe extremal behaviour of stationary time series. We consider their sliding blocks estimators and we prove that they are consistent under conditions that can be easily verified for a large class of models, using a modern theory of multivariate, regularly varying time series.

Keywords and phrases: Regularly varying time series, Extremes, Cluster index, Extremal index.

4.1. Introduction

Consider a stationary, regularly varying $\mathbb{R}^d$-real valued time series $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$. We are interested in a thorough understanding of its extremal behaviour. A classical approach to this problem is to calculate the *extremal index*. If $|\cdot|$ is an arbitrary norm on $\mathbb{R}^d$, then the extremal index θ (if exists) of $\{|\boldsymbol{X}_j|, j \in \mathbb{Z}\}$ is defined as a parameter in the limiting distribution of the maxima. With Q being the quantile function of $|\boldsymbol{X}_0|$ and $a_n = Q(1-1/n)$ we have

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1} \max\{|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_n|\} \le x) = \exp(-\theta x^{-\alpha}) \;, \quad x > 0 \;.$$

The parameter $\theta \in (0,1]$ indicates the amount of clustering, with $\theta = 1$ (the case of extremal independence) meaning no-clustering of large values.

The extremal index is just one parameter that describes clustering of extremes. A related object is the cluster size distribution. It is the probability mass function of the number of exceedences over a large threshold within a given cluster. Both the extremal index and the cluster size distribution stem from an application of a suitable functional to a cluster. This leads to a more general concept of cluster indices.

Informally speaking, a cluster is a triangular array $(\boldsymbol{X}_1/u_n, \ldots, \boldsymbol{X}_{r_n}/u_n)$ with $r_n, u_n \to \infty$ that converges in distribution in a certain sense. Cluster indices are obtained by applying the appropriate functional H to the cluster. The functionals are defined on $(\mathbb{R}^d)^{\mathbb{Z}}$, the space of $\mathbb{R}^d$-valued sequences, and are such that their values do not depend on coordinates that are equal to zero. More precisely, for $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z}$, we denote $\boldsymbol{X}_{i,j} = (\boldsymbol{X}_i, \ldots, \boldsymbol{X}_j) \in (\mathbb{R}^d)^{(j-i+1)}$. Then, we identify $H(\boldsymbol{X}_{i,j})$ with $H((\boldsymbol{0}, \boldsymbol{X}_{i,j}, \boldsymbol{0}))$, where $\boldsymbol{0} \in (\mathbb{R}^d)^{\mathbb{Z}}$ is the zero sequence. Such functionals H will be called *cluster functionals*. See [KS20, Chapter 6] for more details.

Let $|\cdot|$ be an arbitrary norm on $\mathbb{R}^d$ and $\{u_n\}$, $\{r_n\}$ be such that

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} r_n = \lim_{n\to\infty} n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = \infty \;, \quad \lim_{n\to\infty} r_n/n = \lim_{n\to\infty} r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0 \;. \tag{$\mathcal{R}(r_n,u_n)$}$$

Given a cluster functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, we want to estimate the limiting quantity

$$\boldsymbol{\nu}^*(H) = \lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \;. \tag{4.1.1}$$

*Department of Mathematics and Statistics, University of Ottawa, 150 Louis Pasteur Private, Ottawa ON K1N 6N5

We call this quantity *the cluster index associated with H*. Additionally to the aforementioned extremal index and cluster size distribution, the other cluster indices of interest are the large deviation index (introduced in [MW13, MW14] as *cluster index*), ruin index or stop-loss index. See again [KS20, Chapter] and a recent work of [BMW21].

To guarantee existence of the limit we will require additional anticlustering assumptions on the time series $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$; see $\mathcal{AC}(r_n, u_n)$.

Several methods of estimation of the limit $\boldsymbol{\nu}^*(H)$ in (4.1.1) may be employed. The natural one is to consider a statistics based on disjoint blocks of size r_n, cf. [DR10] and [KS20],

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/u_n) ,$$

where $m_n = [n/r_n]$. It is proven in [KS20, Chapter 10] that, under mild weak dependence conditions, the disjoint blocks statistic converges in probability to $\boldsymbol{\nu}^*(H)$. Under stronger β-mixing assumptions, [DR10] and [KS20] prove that that the appropriately scaled and centered estimator is asymptotically normal with the limiting variance given by $\boldsymbol{\nu}^*(H^2)$ (see [KS20, Chapter 10] for the expression for the limiting variance).

In this paper we consider the sliding blocks statistics

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) := \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) ,$$

where $q_n = n - r_n$.

The sliding blocks estimators have been studied for some specific functionals H, however there has been no unified theory available. Recently, [DN20] used the framework of [DR10] and showed that the limiting variance of the sliding blocks estimator never exceeds that of the disjoint blocks estimator. In the special case of the extremal index, they showed that the limiting variance of both disjoint and sliding blocks estimators is the same. [CK21] derived the asymptotic normality of the sliding blocks estimators and showed the equality of variance for all cluster functionals. It has to be pointed out that these results hold in the Peak-over-Threshold framework.

The aforementioned central limit theorems were obtained under somehow restrictive mixing conditions. As such, the main goal of this paper is to obtain consistency of the sliding blocks statistics under minimal assumptions. In particular, we prove the consistency for long memory moving averages, which has never been done in the literature before. In order to do so, we adapt the proofs for disjoint blocks in [KS20] in Chapter 10.

In order to proceed, in section 4.2 we fix the notation, recall the notion of the tail process associated to a stationary regularly varying time series; and introduce the cluster indices. Detailed discussion can be found in [KS20, Chapter 6]. See also [PS18] and [BPS18].

The main results of this paper are Theorems 4.3.3 and 4.3.4. We prove the weak convergence of the appropriately normalized empirical cluster measure. Two types of conditions are used: beta-mixing and the m-approximation. Those conditions can be verified for a variety of models: regularly varying functions of Markov chains, infinite order moving averages and max-stable processes. The m-approximation allows for long range dependence. See [CK21], [KSW19] and [KS20, Part III]. In particular, we obtain consistency of the sliding blocks estimator of the extremal index for long memory moving averages. Surprisingly, such result does not seem to exist in the literature.

All proofs are included in Section 4.4. The proofs are similar to that of [KS20] in the context of disjoint block estimators. However, some crucial technical steps have to be significantly modified due to sliding blocks considered in the present paper.

4.2. Preliminaries

Define the measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \geq 1$, on $\ell_0(\mathbb{R}^d)$ as follows:

$$\boldsymbol{\nu}^*_{n,r_n} = \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[\delta_{u_n^{-1}\boldsymbol{X}_{1,r_n}}\right] .$$

We are interested in convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$. The results of this section are extracted from [KS20, Chapter 6]. See also [PS18] and [BPS18].

4.2.1. Anticlustering condition

For each fixed $r \in \mathbb{N}$, the distribution of $u_n^{-1}\boldsymbol{X}_{-r,r}$ conditionally on $|\boldsymbol{X}_0| > u_n$ converges weakly to the distribution of $\boldsymbol{Y}_{-r,r}$. In order to let r tend to infinity, we must embed all these finite vectors into one space of sequences. By adding zeroes on each side of the vectors $u_n^{-1}\boldsymbol{X}_{-r,r}$ and $\boldsymbol{Y}_{-r,r}$ to identify them with elements of the space $\ell_0(\mathbb{R}^d)$. Then $\boldsymbol{Y}_{-r,r}$ converges (as $r \to \infty$) to $\boldsymbol{Y}$ in $\ell_0(\mathbb{R}^d)$ if (and only if) $\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)$ almost surely. However, this is not enough for statistical purposes and we consider the following definition.

Definition 4.2.1 ([DH95], Condition 2.8)**.** *Condition $\mathcal{AC}(r_n, u_n)$ holds if for all $x, y > 0$,*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{k\leq |j|\leq r_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) = 0 . \qquad (\mathcal{AC}(r_n, u_n))$$

Condition $\mathcal{AC}(r_n, u_n)$ is referred to as the anticlustering condition. It holds for sequence of i.i.d. random variables whenever $\lim_{n\to\infty} r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$, which can be recognized as in the restrictions imposed in $\mathcal{R}(r_n, u_n)$. It is also fulfilled by many models, including geometrically ergodic Markov chains, short-memory linear or max-stable processes.

4.2.2. Vague convergence of cluster measure

Definition 4.2.2. *The space $\tilde{\ell}_0(\mathbb{R}^d)$ is the space of equivalence classes of $\ell_0(\mathbb{R}^d)$ endowed with the equivalence relation $\sim$ defined by*

$$\boldsymbol{x} \sim \boldsymbol{y} \Longleftrightarrow \exists j \in \mathbb{Z}\ ,\ B^j \boldsymbol{x} = \boldsymbol{y}\ ,$$

where B is the backshift operator.

The proof of the following result can be found in [KS20], Theorem 6.2.5. Let $\mathcal{A}$ be the class of all bounded continuous shift-invariant functions H with support separated from $\mathbf{0}$.

Proposition 4.2.3. *Let condition $\mathcal{AC}(r_n, u_n)$ hold. The sequence of measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \geq 1$ converges vaguely$^{\#}$ on $\tilde{\ell}_0(\mathbb{R}^d) \setminus \{\mathbf{0}\}$ to $\boldsymbol{\nu}^*$, that is, for all $H \in \mathcal{A}$,*

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \boldsymbol{\nu}^*(H) .$$

4.3. Consistency of sliding blocks statistic

Let $q_n = n - r_n$. Consider the sliding blocks empirical cluster measure defined as

$$\tilde{\boldsymbol{\mu}}^*_{n,r_n} = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} \delta_{u_n^{-1}\boldsymbol{X}_{i+1,i+r_n}} .$$

Let $z_n \to \infty$ be the sequence of integers such that

$$\lim_{n\to\infty} \frac{z_n r_n}{n} = 0 \tag{4.3.1}$$

and

$$\lim_{n\to\infty} \frac{z_n}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0 \ . \tag{4.3.2}$$

Remark 4.3.1. We note that the assumptions (4.3.1) and (4.3.2) are very mild. They are not related to a time series model nor to construction of the sliding blocks estimator. It is only used to decompose the estimator into large and small blocks. Since $\mathcal{R}(r_n, u_n)$ holds, such choice of the sequence z_n is always possible.

Set

$$\widetilde{m}_n = \left[\frac{q_n}{(z_n+2)r_n}\right] \tag{4.3.3}$$

and assume for simplicity that $\widetilde{m}_n$ is an integer.

We are going to use a decomposition into large and small blocks. First, consider disjoint blocks of size r_n:

$$J_j = \{(j-1)r_n, \dots, jr_n - 1\} \ , \quad j = 1, \dots, m_n = [q_n/r_n] \ .$$

For $j = 1, \dots, \widetilde{m}_n$, define large and small blocks as follows:

$$\begin{aligned}
&L_1 = \{0, \dots, z_n r_n - 1\} \ , \quad S_1 = \{z_n r_n \dots, z_n r_n + 2r_n - 1\} \ , \\
&L_2 = \{z_n r_n + 2r_n, \dots, 2z_n r_n + 2r_n - 1\} \ , \quad S_2 = \{2z_n r_n + 2r_n \dots, 2z_n r_n + 4r_n - 1\} \ , \\
&L_j = \{(j-1)z_n r_n + 2(j-1)r_n, \dots, jz_n r_n + 2(j-1)r_n - 1\} \ , \\
&S_j = \{jz_n r_n + 2(j-1)r_n, \dots, jz_n r_n + 2jr_n - 1\} \ .
\end{aligned}$$

The block L_1 is obtained by merging z_n consecutive blocks $J_1, \dots, J_{z_n}$ of size r_n. Likewise, $S_1 = J_{z_n+1} \cup J_{z_n+2}$. Therefore, the large block of size $z_n r_n$ is followed by the small block of size $2r_n$, which in turn is followed by the large block of size $z_n r_n$ and so on. All together,

$$\bigcup_{j=1}^{\widetilde{m}_n} (L_j \cup S_j) = \{0, \dots, n - r_n\} \ .$$

We can decompose $\widetilde{\boldsymbol{\mu}}^*_{n,r_n}$ into the large and small blocks as

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n} = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \left(\sum_{i\in L_j} \delta_{(\boldsymbol{X}_{i+1,i+r_n}/u_n)} + \sum_{i \in S_j} \delta_{(\boldsymbol{X}_{i+1,i+r_n}/u_n)} \right) \ .$$

We will also need the empirical cluster measures based on one large and small block, respectively:

$$\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n} = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \left(\sum_{i\in L_1} \delta_{u_n^{-1}\boldsymbol{X}_{i+1,i+r_n}} \right) = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{z_n r_n - 1} \delta_{u_n^{-1}\boldsymbol{X}_{i+1,i+r_n}} \ ,$$

$$\widetilde{\boldsymbol{\mu}}^{(s)}_{n,r_n} = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \left(\sum_{i\in S_1} \delta_{u_n^{-1}\boldsymbol{X}_{i+1,i+r_n}} \right) = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{r_n - 1} \delta_{u_n^{-1}(\boldsymbol{X}_{i+1,i+r_n})} \ .$$

The first result establishes consistency of the empirical cluster measure $\widetilde{\boldsymbol{\mu}}^*_{n,r_n}$, if it can be approximated by $\widetilde{m}_n$ independent copies of $\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}$. It can be viewed as a counterpart to a disjoint-blocks result in Lemma 10.1.1 in [KS20].

Lemma 4.3.2. *Assume that $\mathcal{R}(r_n, u_n)$, $\mathcal{AC}(r_n, u_n)$ and (4.3.1)-(4.3.2) hold. Then*

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n} \stackrel{w}{\Longrightarrow} \boldsymbol{\nu}^* \quad \text{in } \mathcal{M}(\tilde{\ell}_0(\mathbb{R}^d)) \tag{4.3.4}$$

if and only if

$$\lim_{n\to\infty} \left| \mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H)}] - \left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}]\right)^{\widetilde{m}_n} \right| = 0, \tag{4.3.5}$$

for all shift-invariant, bounded Lipschitz continuous functions H defined on $\ell_0(\mathbb{R}^d)$ with support separated form zero.

The main results of this paper are Theorems 4.3.3 and 4.3.4, the weak convergence of the appropriately normalized empirical cluster measure. Two types of conditions are used. In Theorem 4.3.3 a beta-mixing assumption is used, while in Theorem 4.3.4 the m-approximation assumption is used instead. They can be viewed as the counterparts of Theorems 10.1.2 and 10.1.3 in [KS20].

Theorem 4.3.3. *Let $\{\boldsymbol{X}_j,\, j \in \mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n, u_n)$ and $\mathcal{AC}(r_n, u_n)$ hold and that the sequence is β-mixing with the rates*

$$\lim_{n\to\infty} \frac{n}{r_n}\beta_{\ell_n} = 0\ , \tag{4.3.6}$$

where $\ell_n \to \infty$ such that $\ell_n/r_n \to 0$. Then (4.3.4) holds.

Theorem 4.3.4. *Let $\{\boldsymbol{X}_j,\, j \in \mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series with tail process $\boldsymbol{Y}$ such that $\mathbb{P}(\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)) = 1$. Assume that $\mathcal{R}(r_n, u_n)$ holds. Let $\{\boldsymbol{X}^{(m)}_j,\, j \in \mathbb{Z}\}$, $m \geq 1$, be regularly varying m-dependent time series such that $\{(\boldsymbol{X}_j, \boldsymbol{X}^{(m)}_j),\, j \in \mathbb{Z}\}$ is stationary for all $m \geq 1$ and for all $\epsilon > 0$,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{\mathbb{P}(|\boldsymbol{X}_0 - \boldsymbol{X}^{(m)}_0| > u_n\epsilon)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0. \tag{4.3.7}$$

Then (4.3.4) holds.

Example 4.3.5. The conditions of Theorem 4.3.3 are much weaker than those required for the central limit theorem; see Theorem 4.3 in [CK21]. They are satisfied by e.g. geometrically ergodic Markov chains.

In the context of Theorem 4.3.4, no central limit theorem is available (for both disjoint and sliding blocks estimators). In particular, the conditions of Theorem 4.3.4 are fulfilled by infinite order moving averages with regularly varying noise, including moving averages with long range dependence. Taking $H = \mathbb{1}\{\boldsymbol{x}^* > 1\}$, we obtain consistency of the sliding blocks estimator of the extremal index. We are not aware of any result that proves such consistency for long memory moving averages.

4.4. Proofs

We start with the following technical lemma.

Lemma 4.4.1. *Assume that $\mathcal{AC}(r_n, u_n)$ and $\mathcal{R}(r_n, u_n)$ hold. Then*

$$\lim_{n\to\infty} \mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(s)}_{n,r_n}(H)}] = 1 \tag{4.4.1}$$

for all shift-invariant, bounded Lipschitz continuous functions H defined on $\ell_0(\mathbb{R}^d)$ with support separated form zero. Moreover, if (4.3.1) holds, then

$$\lim_{n\to\infty} \mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}] = 1\ . \tag{4.4.2}$$

Proof. Assume without loss of generality that $H \geq 0$. Then, to prove (4.4.2) it suffices to prove $\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H) \xrightarrow{\mathbb{P}} 0$ (since $\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H) \geq 0$). In order to prove the latter statement, we prove that $\mathbb{E}[\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)] \to 0$ and apply Markov inequality. Set $v_n := q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Using the definition of $\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)$, we have

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)] &= \frac{1}{v_n}\mathbb{E}\left[\sum_{i=0}^{z_n r_n - 1} H(\boldsymbol{X}_{i+1,i+r_n}/u_n)\right] = \frac{z_n r_n}{v_n}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\right] \\ &= \frac{z_n r_n}{n}\frac{\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\right]}{r_n\mathbb{P}(\boldsymbol{X}_0 > u_n)} = \frac{z_n r_n}{n}\boldsymbol{\nu}^*_{n,r_n}(H)\ .\end{aligned}$$

Using the $\mathcal{AC}(r_n, u_n)$ and Proposition 4.2.3, $\boldsymbol{\nu}^*_{n,r_n}(H) \to \boldsymbol{\nu}^*(H)$. We conclude the proof by applying (4.3.1).

Similarly, for $\widetilde{\boldsymbol{\mu}}^{(s)}_{n,r_n}(H)$, we have

$$\begin{aligned}\mathbb{E}[\widetilde{\boldsymbol{\mu}}^{(s)}_{n,r_n}(H)] &= \frac{1}{v_n}\mathbb{E}\left[\sum_{i=0}^{r_n - 1} H(\boldsymbol{X}_{i+1,i+r_n}/u_n)\right] = \frac{r_n}{v_n}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\right] \\ &= \frac{r_n}{n}\frac{\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\right]}{r_n\mathbb{P}(\boldsymbol{X}_0 > u_n)} = \frac{r_n}{n}\boldsymbol{\nu}^*_{n,r_n}(H) = o(1)\ ,\end{aligned}$$

by $\mathcal{R}(r_n, u_n)$ and Proposition 4.2.3. ☐

Proof of Lemma 4.3.2. The proof is similar to that of Lemma 10.1.1 in [KS20], with some crucial modifications to incorporate the sliding blocks.

We will prove that under the assumptions $\mathcal{AC}(r_n, u_n)$ and (4.3.1),

$$\lim_{n\to\infty}\left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}]\right)^{\widetilde{m}_n} = e^{-\boldsymbol{\nu}^*(H)}, \tag{4.4.3}$$

for all shift-invariant bounded Lipschitz continuous functions H defined on $\ell_0(\mathbb{R}^d)$ with support separated from zero. By Theorem 7.1.17 in [KS20], the weak convergence in (4.3.4) is equivalent to

$$\lim_{n\to\infty}\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H)}] = e^{-\boldsymbol{\nu}^*(H)}, \tag{4.4.4}$$

for the same class of functions.

If (4.4.3) holds, then (4.4.4) is equivalent to (4.3.5). We proceed to prove (4.4.3) which, by Lemma 4.4.1 and Taylor expansion of $\log(1+x)$ at $x = 0$, is also equivalent to

$$\lim_{n\to\infty}\widetilde{m}_n\mathbb{E}[1 - e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}] = \boldsymbol{\nu}^*(H).$$

Recall that $v_n = q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Assume again that $H \geq 0$. We are using the bound $|1 - e^{-x} - x| \leq x^2 e^{x_+}$ to get for $\epsilon > 0$,

$$\begin{aligned}&\lim_{n\to\infty}\widetilde{m}_n\mathbb{E}[1 - e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)} - \widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)] \leq \widetilde{m}_n\mathbb{E}\left[\left(\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)\right)^2 \mathrm{e}^{\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}\right] \\ &= \widetilde{m}_n\mathbb{E}\left[\left(\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)\right)^2 \mathrm{e}^{\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}\mathbb{1}\left\{\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H) \leq \epsilon\right\}\right] \\ &\quad + \widetilde{m}_n\mathbb{E}\left[\left(\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)\right)^2 \mathrm{e}^{\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}\mathbb{1}\left\{\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H) > \epsilon\right\}\right] \\ &=: I + J\ .\end{aligned}$$

For the term I we have

$$
\begin{aligned}
I &\leq \mathrm{e}^{\epsilon} \frac{\widetilde{m}_n}{v_n^2} \mathbb{E}\left[\left(\sum_{i=0}^{z_n r_n - 1} H(u_n^{-1} \boldsymbol{X}_{i+1,i+r_n})\right)^2\right] \\
&\leq \mathrm{cst} \frac{\widetilde{m}_n z_n r_n^3 \mathbb{P}(|\boldsymbol{X}_0| > u_n)}{v_n^2} = \mathrm{cst} \frac{O(1)}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = o(1),
\end{aligned}
$$

by Eq. (7.28) in [CK21] and $\mathcal{R}(r_n, u_n)$.

For J we note that since H is bounded, we have

$$
\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H) \leq \|H\| \frac{z_n r_n}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = O\left(\frac{z_n}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\right) = o(1)
$$

by (4.3.2). Thus, J becomes zero for n large enough.

Therefore,

$$
\lim_{n\to\infty} \widetilde{m}_n \mathbb{E}[1 - e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}] = \lim_{n\to\infty} \widetilde{m}_n \mathbb{E}[\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)] = \boldsymbol{\nu}^*(H).
$$

Hence (4.4.3) holds. □

Proof of Theorem 4.3.3. The proof is somehow similar to that of Theorem 10.1.2 in [KS20]. However, due to the sliding blocks situation in the present paper, we have to modify completely the blocking method. We only need to prove that (4.3.6) implies (4.3.5). Recall that $q_n = n - r_n$. Set

$$
\Psi_j^{(l)}(H) = \sum_{i \in L_j} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)\ , \quad \Psi_j^{(s)}(H) = \sum_{i \in S_j} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)\ .
$$

With such the decomposition, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{z_n r_n + r_n - 1}$ used in the definition of $\Psi_1^{(l)}(H)$ are separated by $r_n + 2$ from the random variables that define $\Psi_2^{(l)}(H)$. The mixing condition (4.3.6) allows us to replace $\boldsymbol{X}$ with the independent blocks process, that is, we can treat the random variables $\Psi_j^{(l)}(H)$, $j = 1, \ldots, \widetilde{m}_n$, as independent. The same applies to $\Psi_j^{(s)}(H)$.

We consider the large-small block decomposition:

$$
\begin{aligned}
\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) &= \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n - 1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \\
&= \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(l)}(H) + \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(s)}(H) \qquad (4.4.5) \\
&=: \widetilde{\boldsymbol{\mu}}^{*(\ell)}_{n,r_n}(H) + \widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)\ .
\end{aligned}
$$

We define also the counterparts of $\widetilde{\boldsymbol{\mu}}^{*(\ell)}_{n,r_n}(H)$ and $\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)$, based on the independent blocks process:

$$
\begin{aligned}
\widetilde{\boldsymbol{\mu}}^{*(\ell)\dagger}_{n,r_n}(H) &:= \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \sum_{i \in L_j} H\left(\boldsymbol{X}^\dagger_{i+1,i+r_n}/u_n\right)\ , \\
\widetilde{\boldsymbol{\mu}}^{*(s)\dagger}_{n,r_n}(H) &:= \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \sum_{i \in S_j} H\left(\boldsymbol{X}^\dagger_{i+1,i+r_n}/u_n\right)\ .
\end{aligned}
$$

Let H be shift-invariant bounded Lipschitz continuous function defined on $\ell_0(\mathbb{R}^d)$ with support separated from zero. Lemma E.3.4 in [CK21] implies

$$
\left|\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*(\ell)}_{n,r_n}(H)}] - \mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*(\ell)\dagger}_{n,r_n}(H)}]\right| \leq \widetilde{m}_n \beta_{r_n} \leq m_n \beta_{r_n}\ .
$$

We also have

$$\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*(\ell)\dagger}_{n,r_n}(H)}] = \left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}}]\right)^{\widetilde{m}_n}$$

by independence of the blocks $\Psi^{(l)}_j(H)$, $j = 1, \ldots, \widetilde{m}_n$. Therefore, recalling that $H \geq 0$

$$\begin{aligned}
&\left|\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*}_{n,r_n}(H)}] - \left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}]\right)^{\widetilde{m}_n}\right| = \\
&= \left|\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*(\ell)}_{n,r_n}(H)} e^{-\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)}] - \left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}]\right)^{\widetilde{m}_n}\right| \\
&\leq \left|\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{*(\ell)}_{n,r_n}(H)}] - \left(\mathbb{E}[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}]\right)^{\widetilde{m}_n}\right| + \mathbb{E}\left[e^{-\widetilde{\boldsymbol{\mu}}^{(\ell)}_{n,r_n}(H)}\left\{e^{-\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)} - 1\right\}\right] \\
&\leq m_n \beta_{r_n} + \mathbb{E}\left[\left\{e^{-\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)} - 1\right\}\right] .
\end{aligned}$$

By (4.3.6) and the assumed $H \geq 0$ it suffices to prove that

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)}\right] = 1 ,$$

which in turn (again by $H \geq 0$) follows from $\widetilde{\boldsymbol{\mu}}^{(s)}_{n,r_n}(H) \xrightarrow{\mathbb{P}} 0$. Note that this is more delicate situation as compared to Lemma 4.4.1, where one small block is considered. In order to prove the latter statement, we evaluate (cf. (4.4.5), (4.4.5) and recall that the small blocks S_j are of size $2r_n$)

$$\begin{aligned}
\mathbb{E}[\widetilde{\boldsymbol{\mu}}^{*(s)}_{n,r_n}(H)] &= \frac{\widetilde{m}_n}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}[\Psi^{(s)}_1(H)] = \frac{2r_n \widetilde{m}_n}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)] \\
&= \frac{2r_n \widetilde{m}_n}{q_n} \frac{\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \frac{2r_n \widetilde{m}_n}{q_n} \boldsymbol{\nu}^*_{n,r_n}(H) = O(1/z_n)\boldsymbol{\nu}^*(H)
\end{aligned}$$

by the definition (4.3.3) of $\widetilde{m}_n$ and $z_n \to \infty$.

Thus (4.3.5) holds. □

Proof of Theorem 4.3.4. The proof is very similar to that of Theorem 10.1.3 in [KS20].

Recall that $v_n = q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)$ and set $v^{(m)}_n := q_n r_n \mathbb{P}\left(\left|\boldsymbol{X}^{(m)}_0\right| > u_n\right)$. By Proposition 5.2.5 in [KS20], (4.3.7) implies

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left|\frac{v_n}{v^{(m)}_n} - 1\right| = 0 . \tag{4.4.6}$$

Define the sliding blocks empirical cluster measure based on the sequence $\boldsymbol{X}^{(m)}$:

$$\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(H) = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\widetilde{m}_n} \left(\sum_{i \in L_j} \delta_{\left(\boldsymbol{X}^{(m)}_{i+1,i+r_n}/u_n\right)} + \sum_{i \in S_j} \delta_{\left(\boldsymbol{X}^{(m)}_{i+1,i+r_n}/u_n\right)} \right) ,$$

Since the time series $\boldsymbol{X}^{(m)}$ is m-dependent, $\mathcal{AC}(r_n, u_n)$ holds for all sequences r_n, u_n satisfying $\mathcal{R}(r_n, u_n)$ by Lemma 6.1.3 in [KS20]. Since m-dependent sequence is also β-mixing with geometric rate, we also have by Theorem 4.3.3 that for every $m \geq 1$,

$$\frac{v_n}{v^{(m)}_n} \widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n} \xrightarrow{w} \boldsymbol{\nu}^{*(m)}, \quad n \to \infty. \tag{4.4.7}$$

We will prove that the empirical cluster measures for the sequences $\boldsymbol{X}$ and $\boldsymbol{X}^{(m)}$ are closed to each other:

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(H)-\widetilde{\boldsymbol{\mu}}^{*}_{n,r_n}(H)|>\eta)=0\,,$$

for all $\eta>0$ and all shift-invariant bounded Lipschitz continuous function defined with support separated from zero. By Theorem 7.1.17 and the triangular argument Lemma A.2.10 in [KS20], this ensures that (4.3.4) holds.

Let $\epsilon>0$ be such that $H(\boldsymbol{x})=0$ if $\boldsymbol{x}^*\le\epsilon$. Without loss of generality, suppose that H is 1-Lipschitz. Fix $\xi\in(0,\epsilon/2)$. Write $\boldsymbol{X}_{n,i}=u_n^{-1}\boldsymbol{X}_{i+1,i+r_n}$, then

$$\begin{aligned}
&|\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(H)-\widetilde{\boldsymbol{\mu}}^{*}_{n,r_n}(H)|\\
&=\left|\frac{1}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\left(H\left(\boldsymbol{X}^{(m)}_{n,i}\right)-H\left(\boldsymbol{X}_{n,i}\right)\right)+\frac{1}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in S_j}\left(H\left(\boldsymbol{X}^{(m)}_{n,i}\right)-H\left(\boldsymbol{X}_{n,i}\right)\right)\right|\\
&=:|I+II|\le|I|+|II|\,.
\end{aligned}$$

We have

$$\begin{aligned}
|I|&\le\frac{1}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\left|H\left(\boldsymbol{X}^{(m)}_{n,i}\right)-H\left(\boldsymbol{X}_{n,i}\right)\right|\mathbb{1}\left\{\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\le\xi u_n\right\}\\
&\quad+\frac{1}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\left|H\left(\boldsymbol{X}^{(m)}_{n,i}\right)-H\left(\boldsymbol{X}_{n,i}\right)\right|\mathbb{1}\left\{\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\ge\xi u_n\right\}\\
&\le\frac{\xi}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\mathbb{1}\left\{|\boldsymbol{X}^{(m)*}_{i+1,i+r_n}|\ge\epsilon u_n/2\right\}\\
&\quad+\frac{\mathrm{cst}}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\mathbb{1}\left\{\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\ge\xi u_n\right\}\\
&\le\xi\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(\bar{\boldsymbol{B}}^c(\boldsymbol{0},\epsilon/2))+\frac{\mathrm{cst}}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\mathbb{1}\left\{\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\ge\xi u_n\right\}\,.
\end{aligned}$$

In the second last inequality we used, first, the Lipschitz and the vanishing property of H, second the boundedness of H.

By (4.4.6), for arbitrary $\delta>0$ and m,n large enough (depending on δ), $v_n^{(m)}/v_n\le(1+\delta)$. Hence,

$$|I|\le\xi\frac{v_n}{v_n^{(m)}}(1+\delta)\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(\bar{\boldsymbol{B}}^c(\boldsymbol{0},\epsilon/2))+\frac{\mathrm{cst}}{v_n}\sum_{j=1}^{\widetilde{m}_n}\sum_{i\in L_j}\mathbb{1}\left\{\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\ge\xi u_n\right\}\,.$$

Thus, for $\eta>0$, applying Markov inequality to the second term, we obtain by stationarity

$$\begin{aligned}
&\mathbb{P}(|\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(H)-\widetilde{\boldsymbol{\mu}}^{*}_{n,r_n}(H)|>\eta)\\
&\le\mathbb{P}\left(\frac{v_n}{v_n^{(m)}}\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(\bar{\boldsymbol{B}}^c(\boldsymbol{0},\epsilon/2))>\eta/(2(1+\delta)\xi)\right)\\
&\quad+\frac{1}{\eta}\frac{\mathrm{cst}}{v_n}\sum_{k=1}^{q_n}\mathbb{P}\left(\max_{i+1\le k\le i+r_n}\left|\boldsymbol{X}^{(m)}_k-\boldsymbol{X}_k\right|\ge\xi u_n\right)\\
&=:I_1+\mathrm{cst}\frac{q_nr_n}{v_n}\mathbb{P}\left(\left|\boldsymbol{X}^{(m)}_0-\boldsymbol{X}_0\right|\ge\xi u_n\right)\,.
\end{aligned}$$

Applying (4.3.7), (4.4.7) and the convergence of $\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n} \xrightarrow{v^{\#}} \boldsymbol{\nu}^*$, $(m \to \infty)$, which follows from Lemma 6.2.7 in [KS20]), we obtain

$$
\begin{aligned}
&\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}(|\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(H) - \widetilde{\boldsymbol{\mu}}^{*}_{n,r_n}(H)| > \eta) \\
&\le \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\widetilde{\boldsymbol{\mu}}^{*(m)}_{n,r_n}(\bar{\boldsymbol{B}}^c(\mathbf{0}, \epsilon/2)) > \eta/(2(1+\delta)\xi)\right) \\
&= \mathbb{P}\left(\boldsymbol{\nu}^*(\bar{\boldsymbol{B}}^c(\mathbf{0}, \epsilon/2)) > \eta/(2(1+\delta)\xi)\right) = 0 \,.
\end{aligned}
$$

This conclude the proof, since ξ can be taken small enough. $\square$

Bibliography

[BB18] Betina Berghaus and Axel Bücher. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Annals of Statistics*, 46(5):2307–2335, 2018.

[BBKS20] Clemonell Bilayi-Biakana, Rafał Kulik, and Philippe Soulier. Statistical inference for heavy tailed series with extremal independence. *Extremes*, 23(1):1–33, 2020.

[BPS18] Bojan Basrak, Hrvoje Planinić, and Philippe Soulier. An invariance principle for sums and record times of regularly varying stationary sequences. *Probability Theory and Related Fields*, 172(3-4):869–914, 2018.

[BS09] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[BS18a] Axel Bücher and Johan Segers. Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125, 2018.

[BS18b] Axel Bücher and Johan Segers. Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462, 2018.

[BMW21] Gloria Buriticá Thomas Mikosch and Olivier Wintenberger. Threshold selection for cluster inference based on large deviation principles. *Preprint*, arXiv:2106.12822.

[DM09] Richard A. Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 38A:977–1009, 2009. Probability, statistics and seismology.

[DH95] Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability*, 23(2):879–917, 1995.

[DHS18] Clément Dombry, Enkelejd Hashorva, and Philippe Soulier. Tail measure and spectral tail process of regularly varying time series. *Annals of Applied Probability*, 28(6):3884–3921, 2018.

[DK20] Holger Drees and Miran Knezevic. Peak-over-threshold estimators for spectral tail processes: Random vs deterministic thresholds. *Extremes*, 2020. DOI: https://doi.org/10.1007/s10687-019-00367-x.

[DN20] Holger Drees and Sebastian Neblung. Asymptotics for sliding blocks estimators of rare events. arXiv:2003.01016, 2020.

[DR10] Holger Drees and Holger Rootzén. Limit theorems for empirical processes of cluster functionals. *Annals of Statistics*, 38(4):2145–2186, 2010.

[DSW15] Holger Drees, Johan Segers, and Michał Warchoł. Statistics for tail processes of Markov chains. *Extremes*, 18(3):369–402, 2015.

[EKM97] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch . *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, 1997.

[Hsi91] Tailen Hsing. Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37(1):117–139, 1991.

[Ibr62] I. A. Ibragimov. Some limit theorems for stationary processes. *Theor. Probab. Appl.*, 7:349–382, 1962.

[KS20] Rafał Kulik and Philippe Soulier. *Heavy tailed time series.* Springer, 2020.

[KSW19] Rafał Kulik, Philippe Soulier, and Olivier Wintenberger. The tail empirical process of regularly varying functions of geometrically ergodic markov chains. *Stochastic Processes* *and their Applications*, 129(1):4209–4238, 2019.

[MW13] Thomas Mikosch and Olivier Wintenberger. Precise large deviations for dependent regularly varying sequences. *Probability Theory and Related Fields*, 156(3-4):851–887, 2013.

[MW14] Thomas Mikosch and Olivier Wintenberger. 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, 2014.

[MW16] Thomas Mikosch and Olivier Wintenberger. A large deviations approach to limit theorem for heavy-tailed time series. *Probability Theory and Related Fields*, 166(1-2):233–269, 2016.

[PS18] Hrvoje Planinić and Philippe Soulier. The tail process revisited. *Extremes*, 21(4):551–579, 2018.

[RLdH98] Holger Rootzén, Ross M. Leadbetter, and Laurens de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8(3):868–885, 1998.

[RSF09] Christian Y. Robert, Johan Segers, and Christopher A. T. Ferro. A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3:993–1020, 2009.

[SW94] Richard L. Smith and Ishay Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society. Series B. Methodological*, 56(3):515–528, 1994.

[vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer, New York, 1996.

[CK21] Youssouph Cissokho and Rafal Kulik. Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electronic Journal of Statistics*, 15(1): 2777–2831, 2021.

[WN98] Ishay Weissman and S. Yu. Novak. On blocks and runs estimators of the extremal index. *Journal of Stat. Planning and Inference*, 66(2):281–288, 1998.

[ZVB20] Nan Zou, Stanislav Volgushev, and Axel Bücher. Multiple block sizes and overlapping blocks for multivariate time series extremes. *Annals of Statistics*, 2020.

Chapter 5

Central limit theorem for disjoint blocks estimators

This section summarizes the results of the theory for the disjoint blocks estimators that can be found in [KS20, Chapter 10], where the central limit theorem is established.

Recall that the disjoint block statistics based on the blocks of size r_n can de defined as

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/u_n) ,$$

where $m_n = [n/r_n]$. The following section summaries results on the central limit theorems of the disjoint blocks statistics $\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H)$. The data-based estimator is constructed as follows. Let $k_n \to \infty$ be a sequence of integers and define u_n by $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Let $|\boldsymbol{X}|_{(n:1)} \leq \cdots \leq |\boldsymbol{X}|_{(n:n)}$ be order statistics from $|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_n|$. Define

$$\widehat{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{k_n} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/|\boldsymbol{X}|_{(n:n-k_n)}) . \tag{5.0.1}$$

Moreover, define the empirical cluster process

$$\widehat{\mathbb{G}}_n(H) = \sqrt{k_n}\{\widehat{\boldsymbol{\nu}}^*_{n,r_n}(H) - \boldsymbol{\nu}^*(H)\}.$$

We are interested in the asymptotic distribution of $\widehat{\mathbb{G}}_n$.

Let $\mathbb{G}$ be a Gaussian random measure on $L^2(\boldsymbol{\nu}^*)$ with mean measure $\boldsymbol{\nu}^*$, that is a random process indexed by $L^2(\boldsymbol{\nu}^*)$ whose finite-dimensional distributions are Gaussian with mean zero and covariance

$$\mathrm{cov}\Big(\mathbb{G}(H),\, \mathbb{G}(H')\Big) = \boldsymbol{\nu}^*(HH').$$

For asymptotic normality, we need to strengthen the anticlustering condition $\mathcal{AC}(r_n, u_n)$.

Definition 5.0.1. *Condition $\mathcal{S}(r_n, u_n)$ holds if for all $s, t > 0$*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=m}^{r_n} \mathbb{P}(|\boldsymbol{X}_0| > u_n s, |\boldsymbol{X}_j| > u_n t) = 0 . \tag{$\mathcal{S}(r_n,u_n)$}$$

This condition implies that $\sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) < \infty$. The latter series appears explicitly in the statement for the limiting variance. Dependence in $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ will be controlled by the β-mixing rates $\{\beta_n\}$. Recall $\mathcal{R}(r_n,u_n)$. Let $\{\ell_n\}$ be a sequence of integers such that $\lim_{n\to\infty}\ell_n=\infty$ and $\lim_{n\to\infty}\ell_n/r_n=0$.

Recall that the map $\mathcal{E}$ is defined on $\ell_0(\mathbb{R}^d)$ by $\mathcal{E}(\boldsymbol{x})=\sum_{j\in\mathbb{Z}}\mathbb{1}\{|\boldsymbol{x}_j|>1\}$. For $s>0$, the function $H_s:(\mathbb{R}^d)^{\mathbb{Z}}\to\mathbb{R}$ is defined by $H_s(\boldsymbol{x})=H(\boldsymbol{x}/s)$.

Theorem 5.0.2. *Let $\{\boldsymbol{X}_j,j\in\mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n,u_n)$, $\mathcal{S}(r_n,u_n)$ hold and that the sequence is β-mixing with the rates*

$$\lim_{n\to\infty}\frac{n}{r_n}\beta_{\ell_n}=0\;, \tag{5.0.2}$$

where $\ell_n\to\infty$ such that $\ell_n/r_n\to 0$.. Fix $0<s_0<1<t_0<\infty$. Let $H:(\mathbb{R}^d)^{\mathbb{Z}}\to\mathbb{R}$ be a shift-invariant measurable map such that the class $\mathcal{H}=\{H_s:s\in[s_0,t_0]\}$ is linearly ordered and satisfies

(T1) For all $H\in\mathcal{H}$, $\lim_{n\to\infty}\boldsymbol{\nu}^*_{n,r_n}(H)=\boldsymbol{\nu}^*(H)$.
(T2) For all $H\in\mathcal{H}$ and $\eta>0$,

$$\lim_{n\to\infty}\boldsymbol{\nu}^*_{n,r_n}\left(H^2\mathbb{1}\left\{|H|>\eta\sqrt{n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\right\}\right)=0.$$

(T3) Assume that $\exists\ \ell_n\to\infty$ such that $\ell_n/r_n\to 0$, there exist functions $K_n:(\mathbb{R}^d)^{\ell_n}\to\mathbb{R}_+$

$$\left|H\left(\frac{\boldsymbol{X}_{1,r_n}}{u_n}\right)-H\left(\frac{\boldsymbol{X}_{1,r_n-\ell_n}}{u_n}\right)\right|\le K_n(\boldsymbol{X}_{r_n-\ell_n+1,r_n})\;,\quad \lim_{n\to\infty}\frac{\mathbb{E}\left[K_n^2(\boldsymbol{X}_{1,\ell_n})\right]}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=0\;.$$

Assume moreover that

$$\lim_{n\to\infty}\sqrt{k_n}\sup_{s\in[s_0,t_0]}|\boldsymbol{\nu}^*_{n,r_n}(\mathcal{E}_s)-\boldsymbol{\nu}^*(\mathcal{E}_s)|=0 \tag{5.0.3a}$$

$$\lim_{n\to\infty}\sqrt{k_n}\sup_{s\in[s_0,t_0]}|\boldsymbol{\nu}^*_{n,r_n}(H_s)-\boldsymbol{\nu}^*(H_s)|=0\;. \tag{5.0.3b}$$

Then $\widehat{\mathbb{G}}_n(H)\xrightarrow{\mathrm{d}}\mathbb{G}(H-\boldsymbol{\nu}^*(H)\mathcal{E})$. The convergence holds jointly for a finite collection of H which satisfy the assumptions.

Remark 5.0.3. We note that the mixing conditions for validity of the central limit theorem in Theorem 5.0.2 are the same as those for consistency in Theorem 3.2.2. On the other hand, the anticlustering condition $\mathcal{S}(r_n,u_n)$ used in Theorem 5.0.2 is stronger than the one needed in Theorem 3.2.2.

Remark 5.0.4.
1. Since the class $\mathcal{H}$ satisfies condition (T1) in Theorem 5.0.2, for every $s\in[s_0,t_0]$, we have

$$\lim_{n\to\infty}\boldsymbol{\nu}^*_{n,r_n}(H_s)=\boldsymbol{\nu}^*(H_s)=s^{-\alpha}\boldsymbol{\nu}^*(H).$$

2. If H is shift-invariant the limiting distribution is centered Gaussian with variance

$$\begin{aligned}\boldsymbol{\nu}^*(\{H-\boldsymbol{\nu}^*(H)\mathcal{E}\}^2)&=\boldsymbol{\nu}^*(H^2)-2\boldsymbol{\nu}^*(H)\boldsymbol{\nu}^*(H\mathcal{E})+(\boldsymbol{\nu}^*(H))^2\boldsymbol{\nu}^*(\mathcal{E}^2)\\&=\boldsymbol{\nu}^*(H^2)-2\boldsymbol{\nu}^*(H)\mathbb{E}[H(\boldsymbol{Y})]+(\boldsymbol{\nu}^*(H))^2\sum_{j\in\mathbb{Z}}\mathbb{P}(|\boldsymbol{Y}_j|>1)\;.\end{aligned}$$

$\oplus$

Chapter 6

Central limit theorem for sliding blocks estimators

This chapter consists of the paper titled *Estimation of cluster functionals for regularly varying time series: sliding blocks estimators*. The paper is self-contained, with all the required definitions. The references can be found at the end of the paper, on page 75. This paper appeared in *Electronic Journal of Statistics*.

Electronic Journal of Statistics
Vol. 15(1) (2021) 2777–2831

Estimation of cluster functionals for regularly varying time series: sliding blocks estimators

Youssouph Cissokho* **and Rafal Kulik***

Abstract: Cluster indices describe extremal behaviour of stationary time series. We consider their sliding blocks estimators. Using a modern theory of multivariate, regularly varying time series, we obtain central limit theorems under conditions that can be easily verified for a large class of models. In particular, we show that in the Peaks-Over-Threshold framework, sliding and disjoint blocks estimators have the same limiting variance.

Keywords and phrases: Regularly varying time series, Extremes, Cluster index, Extremal index.

6.1. Introduction

Consider a stationary, regularly varying $\mathbb{R}^d$-real valued time series $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$. We are interested in a thorough understanding of its extremal behaviour. A classical approach to this problem is to calculate the *extremal index*. If $|\cdot|$ is an arbitrary norm on $\mathbb{R}^d$, then the extremal index θ (if exists) of $\{|\boldsymbol{X}_j|, j \in \mathbb{Z}\}$ is defined as a parameter in the limiting distribution of the maxima. With Q being the quantile function of $|\boldsymbol{X}_0|$ and $a_n = Q(1-1/n)$ we have

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1} \max_{j=1,\dots,n} \{|\boldsymbol{X}_1|, \dots, |\boldsymbol{X}_n|\} \le x) = \exp(-\theta x^{-\alpha}) , \quad x > 0 .$$

The parameter $\theta \in (0,1]$ indicates the amount of clustering, with $\theta = 1$ (the case of extremal independence) meaning no-clustering of large values.

The extremal index is just one parameter that describes clustering of extremes. A related object is the cluster size distribution. It is the probability mass function of the number of exceedences over a large threshold within a given cluster. Both the extremal index and the cluster size distribution stem from an application of a suitable functional to a cluster. This leads to a more general concept of cluster indices.

Informally speaking, a cluster is a triangular array $(\boldsymbol{X}_1/u_n, \dots, \boldsymbol{X}_{r_n}/u_n)$ with $r_n, u_n \to \infty$ that converges in distribution in a certain sense. Cluster indices are obtained by applying the appropriate functional H to the cluster. The functionals are defined on $(\mathbb{R}^d)^{\mathbb{Z}}$, the space of $\mathbb{R}^d$-valued sequences, and are such that their values do not depend on coordinates that are equal to zero. More precisely, for $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z}$, we denote $\boldsymbol{X}_{i,j} = (\boldsymbol{X}_i, \dots, \boldsymbol{X}_j) \in (\mathbb{R}^d)^{(j-i+1)}$. Then, we identify $H(\boldsymbol{X}_{i,j})$ with $H((\boldsymbol{0}, \boldsymbol{X}_{i,j}, \boldsymbol{0}))$, where $\boldsymbol{0} \in (\mathbb{R}^d)^{\mathbb{Z}}$ is the zero sequence. Such functionals H will be called *cluster functionals*.

Let $|\cdot|$ be an arbitrary norm on $\mathbb{R}^d$ and $\{u_n\}$, $\{r_n\}$ be such that

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} r_n = \lim_{n\to\infty} n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = \infty , \quad \lim_{n\to\infty} r_n/n = \lim_{n\to\infty} r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0 . \tag{$\mathcal{R}(r_n,u_n)$}$$

*Department of Mathematics and Statistics, University of Ottawa, 150 Louis Pasteur Private, Ottawa ON K1N 6N5

Given a cluster functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, we want to estimate the limiting quantity

$$\boldsymbol{\nu}^*(H) = \lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \,. \tag{6.1.1}$$

To guarantee existence of the limit we will require additional anticlustering assumptions on the time series $\{\boldsymbol{X}_j, j\in\mathbb{Z}\}$. For $\boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ define $\boldsymbol{x}^* = \sup_{j\in\mathbb{Z}} |\boldsymbol{x}_j|$. The cluster indices of interest are, among others:

- the extremal index obtained with $H_1(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\}$, $\boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$;
- the cluster size distribution obtained with

$$H_2(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\} = m\right\}, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}, \quad m\in\mathbb{N}\,; \tag{6.1.2}$$

- the stop-loss index of a univariate time series obtained with

$$H_3(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} (x_j - 1)_+ > \eta\right\}, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}}, \quad \eta > 0\,; \tag{6.1.3}$$

- the large deviation index of a univariate time series obtained with

$$H_4(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\}\,, \quad K(\boldsymbol{x}) = \left(\sum_{j\in\mathbb{Z}} x_j\right)_+, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}}\,; \tag{6.1.4}$$

- the ruin index of a univariate time series obtained with

$$H_5(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\}\,, \quad K(\boldsymbol{x}) = \sup_{i\in\mathbb{Z}} \left(\sum_{j\le i} x_j\right)_+, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j\in\mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}}\,. \tag{6.1.5}$$

As mentioned above, the extremal index is the classical quantity that arises in the extreme value theory for dependent sequences. The cluster size distribution is again a well-known object and was studied in [Hsi91] and [DR10]. The large deviation index was studied under the name *cluster index* in [MW13, MW14]. It quantifies the effect of dependence in large deviations results. The remaining cluster indices seem to be new.

It has to be pointed out that the aforementioned cluster indices describe clustering of extremes in blocks of an increasing size r_n. In this general framework, we can also consider summation functionals

$$H_\phi(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \phi(\boldsymbol{x}_j, \ldots, \boldsymbol{x}_{j+h})$$

with $\phi : \mathbb{R}^{d(h+1)} \to \mathbb{R}$. Such summation functionals yield *tail array sums*, which in turn give finite-dimensional extremal characteristics. A suitable choice of ϕ leads to *extremogram* (see [DM09]) or the distribution of the spectral tail process (see [DSW15]). See also Section 10.4.3 in [KS20] for more examples. However, as we will indicate below, the tail array sums are not interesting in the context of the present paper.

Several methods of estimation of the limit $\boldsymbol{\nu}^*(H)$ in (6.1.1) may be employed. The natural one is to consider a statistics based on disjoint blocks of size r_n, cf. [DR10] and [KS20],

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1, ir_n}/u_n)\,,$$

where $m_n = [n/r_n]$. It is proven in the aforementioned references that the appropriately scaled and centered estimator is asymptotically normal with the limiting variance given by $\boldsymbol{\nu}^*(H^2)$ (see [KS20, Chapter 10] for the expression for the limiting variance). The data-based estimator is constructed as follows. Let $k_n \to \infty$ be a sequence of integers and define u_n by $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Let $|\boldsymbol{X}|_{(n:1)} \le \cdots \le |\boldsymbol{X}|_{(n:n)}$ be order statistics from $|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_n|$. Define

$$\widehat{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{k_n}\sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1,ir_n}/|\boldsymbol{X}|_{(n:n-k_n)}) \ . \tag{6.1.6}$$

Although some special cases were considered (estimation of the extremal index in [Hsi91] and [SW94]; tail array sums in [RLdH98]), the general theory was developed in [DR10]. The summary of the theory for the disjoint blocks estimators can be found in [KS20, Chapter 10], where consistency and the central limit theorems are established. The limiting variance of the disjoint blocks estimator can be represented as

$$\boldsymbol{\nu}^*(\{H - \boldsymbol{\nu}^*(H)\mathcal{E}\}^2) \ , \tag{6.1.7}$$

where $\mathcal{E}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$. This result was established (implicitly) in [DR10], but the form of the limiting variance is again given in [KS20, Chapter 10].

In this paper we consider the sliding blocks statistics

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) := \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \ , \tag{6.1.8}$$

where $q_n = n - r_n - 1$ and the corresponding estimator defined in terms of order statistics:

$$\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H) = \frac{1}{r_n k_n} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/|\boldsymbol{X}|_{(n:n-k_n)}\right) \ . \tag{6.1.9}$$

The sliding blocks estimators have been studied for some specific functionals H, however there has been no unified theory available. Recently, [DN20] used the framework of [DR10] and showed that the limiting variance of the sliding blocks estimator never exceeds that of the disjoint blocks estimator. In case of the extremal index, both variances were proven to be equal.

The goal of this paper is to obtain the asymptotic normality of the sliding blocks estimators. Our focus is on providing the conditions that can be easily verified for a variety of time series models. At the same time, we will show that the limiting variance of both disjoint and sliding blocks estimators is the same. To achieve our goal, we combine [DR10] approach with the modern theory of stationary, regularly varying time series. We note in passing that in case of the tail array sums, the sliding blocks estimators reduce to disjoint blocks one. As such, the limiting theory is already known, see [RLdH98] and Chapter 10 in [KS20]. In order to proceed, in Section 6.2 we fix the notation, recall the notion of the tail process associated to a stationary regularly varying time series; and introduce the cluster indices.

Next, we need to answer a non-trivial question: *When does the limit* $\boldsymbol{\nu}^*(H)$ *exist?*. For this, Section 6.3 deals with convergence of cluster measures and cluster indices $\boldsymbol{\nu}^*(H)$ appear as the limit. Existence of the limit requires an anticlustering assumption. In conjunction with a particular choice of functionals, we will be in position to give specific examples of cluster indices. The contents of this section is based on [KS20, Chapter 6]. Some results stem from [MW14, MW16] and [BPS18].

The main result is Theorem 6.4.3. We prove the central limit theorem for the data-based sliding blocks estimator (6.1.9) under easy to verify assumptions. Those conditions can be verified for a variety of models: regularly varying functions of Markov chains, infinite order moving averages, max-stable processes. See [KSW19] and [KS20, Part III].

The most important (and somehow surprising) conclusion is that both sliding (6.1.9) and disjoint (6.1.6) blocks estimators yield the same variance. This is in agreement with the result for the extremal index in [DN20]. On the other hand, it seems to be a contradiction with other available results; see [BS18b], [BS18a] or [ZVB20]. The main difference is that we obtain our asymptotic results in the Peaks-over-Threshold framework, while the latter papers deal with the block maxima framework. We discuss different results in both frameworks in Section 6.5.

In Section 6.6 we illustrate the asymptotic theory by a small simulation study for simple time series models, AR(1) and ARCH(1). Interestingly, although estimators of the extremal index perform better in case of a stronger dependence (which is not surprising, see e.g. simulation studies in [RSF09]), we have the opposite situation for the stop-loss index.

All proofs are included in Section 6.7.

6.2. Preliminaries

In this section we fix the notation and introduce the relevant classes of functions. In Section 6.2.3 we recall the notion of the tail and the spectral tail process (cf. [BS09]). In Section 6.2.4 we define cluster indices; see [KS20, Chapter 5] for a detailed introduction.

6.2.1. Notation

Let $|\cdot|$ be a norm on $\mathbb{R}^d$. For a sequence $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z} \cup \{-\infty, \infty\}$ we denote $\boldsymbol{x}_{i,j} = (\boldsymbol{x}_i, \dots, \boldsymbol{x}_j) \in (\mathbb{R}^d)^{j-i+1}$, $\boldsymbol{x}^*_{i,j} = \max_{i \le l \le j} |\boldsymbol{x}_l|$ and $\boldsymbol{x}^* = \sup_{j \in \mathbb{Z}} |\boldsymbol{x}_j|$. By $\mathbf{0}$ we denote the zero sequence; its dimension can be different in each of its occurrences.

By $\ell_0(\mathbb{R}^d)$ we denote the set of $\mathbb{R}^d$-valued sequences which tend to zero at infinity. Likewise, $\ell_1(\mathbb{R}^d)$ consists of sequences such that $\sum_{j\in\mathbb{Z}} |\boldsymbol{x}_j| < \infty$.

We will use the blocking method. If $\boldsymbol{X}$ is a time series of interest, then $(\boldsymbol{X}^\dagger_1, \dots, \boldsymbol{X}^\dagger_n)$ is a pseudo-sample such that the blocks $(\boldsymbol{X}^\dagger_{(i-1)r_n+1}, \dots, \boldsymbol{X}^\dagger_{ir_n})$, $i = 1, \dots, m_n = [n/r_n]$, are mutually independent with the same distribution as the original block $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r_n})$.

6.2.2. Classes of functions

Functionals H are defined on $\ell_0(\mathbb{R}^d)$ with the convention $H(\boldsymbol{x}_{i,j}) = H((\mathbf{0}, \boldsymbol{x}_{i,j}, \mathbf{0}))$. In particular, the map $\mathcal{E}$ is defined on $\ell_0(\mathbb{R}^d)$ by $\mathcal{E}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$. For $s > 0$, the function $H_s : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ is defined by $H_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)$. We consider the following classes:

- $\mathcal{L}$ is the class of bounded real-valued functions defined on $(\mathbb{R}^d)^{\mathbb{Z}}$ that are either Lipschitz continuous with respect to the uniform norm or almost surely continuous with respect to the distribution of the tail process $\boldsymbol{Y}$. This class includes functions like $\mathbb{1}\{\boldsymbol{x}^* > 1\}$, $\mathbb{1}\left\{\sum_{j\in\mathbb{Z}} |\boldsymbol{x}_j| > 1\right\}$. See Remark 6.1.6 in [KS20].
- $\mathcal{A} \subset \mathcal{L}$ is the class of shift-invariant functionals with support separated from $\mathbf{0}$. In particular, for $H \in \mathcal{A}$, $H(\mathbf{0}) = 0$. The class $\mathcal{A}$ includes $\mathbb{1}\{\boldsymbol{x}^* > 1\}$.
- $\mathcal{K}$ is the class of shift-invariant functionals $K : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ defined on $\ell_1(\mathbb{R}^d)$ such that $K(\mathbf{0}) = 0$ and which are Lipschitz continuous with constant L_K, i.e.

$$|K(\boldsymbol{x}) - K(\boldsymbol{y})| \le L_K \sum_{j\in\mathbb{Z}} \left|\boldsymbol{x}_j - \boldsymbol{y}_j\right| \ , \quad \boldsymbol{x}, \boldsymbol{y} \in \ell_1(\mathbb{R}^d) \ . \tag{6.2.1}$$

- $\mathcal{B} \subset \mathcal{L}$ is the class of functionals H of the form $H = \mathbb{1}\{K > 1\}$, where $K \in \mathcal{K}$. Functionals in $\mathcal{B}$ may have support which is not separated from $\mathbf{0}$. The typical example is $H(\boldsymbol{x}) = \mathbb{1}\left\{\sum_j |\boldsymbol{x}_j| > 1\right\}$; note that $H \notin \mathcal{A}$.

6.2.3. Tail and spectral tail process

Let $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying time series with values in $\mathbb{R}^d$ and tail index α. In particular,

$$\lim_{x\to\infty} \frac{\mathbb{P}(|\boldsymbol{X}_0| > tx)}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = t^{-\alpha}$$

for all $t > 0$. Then, there exists a sequence $\boldsymbol{Y} = \{\boldsymbol{Y}_j, j \in \mathbb{Z}\}$ such that

$$\mathbb{P}(x^{-1}(\boldsymbol{X}_i, \dots, \boldsymbol{X}_j) \in \cdot \mid |\boldsymbol{X}_0| > x) \text{ converges weakly to } \mathbb{P}((\boldsymbol{Y}_i, \dots, \boldsymbol{Y}_j) \in \cdot)$$

as $x \to \infty$ for all $i \le j \in \mathbb{Z}$. We call $\boldsymbol{Y}$ the tail process. See [BS09]. We note that, in particular, $|\boldsymbol{Y}_0|$ has Pareto distribution with the density $\alpha x^{-\alpha-1}$, $x > 1$. As such, it follows automatically that $\boldsymbol{Y}^* = \sup_{j\in\mathbb{Z}} |\boldsymbol{Y}_j| > 1$. Equivalently, viewing $\boldsymbol{X}$ and $\boldsymbol{Y}$ as random elements with values in $(\mathbb{R}^d)^{\mathbb{Z}}$, we have for every bounded or non-negative functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, continuous with respect to the product topology,

$$\lim_{x\to\infty} \frac{\mathbb{E}[H(x^{-1}\boldsymbol{X})\mathbb{1}\{|\boldsymbol{X}_0| > x\}]}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = \mathbb{E}[H(\boldsymbol{Y})] \; .$$

Define $\boldsymbol{\Theta}_j = \boldsymbol{Y}_j/|\boldsymbol{Y}_0|$, $j \in \mathbb{Z}$. The sequence $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_j, j \in \mathbb{Z}\}$ is called the spectral tail process. The random variable $|\boldsymbol{Y}_0|$ has the Pareto distribution with index α and is independent from $\boldsymbol{\Theta}$. Hence for a non-negative measurable function $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$,

$$\mathbb{E}[H(\boldsymbol{Y})] = \int_1^\infty \mathbb{E}[H(r\boldsymbol{\Theta})]\alpha r^{-\alpha-1}\mathrm{d}r \; . \tag{6.2.2}$$

6.2.4. Cluster measure and cluster indices

Consider the infargmax functional $\mathcal{C}_0$ defined on $(\mathbb{R}^d)^{\mathbb{Z}}$ by $\mathcal{C}_0(\boldsymbol{y}) = \inf\{j : \boldsymbol{y}^*_{-\infty,j} = \boldsymbol{y}^*\}$, with the convention that $\inf\emptyset = +\infty$. If $\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$ then we can define

$$\vartheta = \mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) = 0) \; . \tag{6.2.3}$$

In fact, $\mathcal{C}_0$ can be replaced with any anchoring map (see [PS18] and [KS20, Theorem 5.5.3]), but we do not pursue it here. For the purpose of this paper it is sufficient to note that we can write alternatively

$$\vartheta = \mathbb{P}\left(\sup_{j\le -1} |\boldsymbol{Y}_j| \le 1\right) \; . \tag{6.2.4}$$

The relationship between (6.2.3) and (6.2.4) is certainly not obvious. The proof is given in Section 6.7.1. Equation (6.2.4) emphasizes a special role of the event $\{\sup_{j\le-1} |\boldsymbol{Y}_j| \le 1\}$ and with its help ϑ can be recognized as the (candidate) extremal index. It becomes the usual extremal index under additional mixing and anticlustering conditions.

Definition 6.2.1. *Let $\boldsymbol{Y}$ and $\boldsymbol{\Theta}$ be the tail process and the spectral tail process, respectively, such that $\mathbb{P}(\lim_{|j|\to\infty} \boldsymbol{Y}_j = \mathbf{0}) = 1$. The **cluster measure** is the measure $\boldsymbol{\nu}^*$ on $\ell_0(\mathbb{R}^d)$ defined by*

$$\boldsymbol{\nu}^* = \vartheta \int_0^\infty \mathbb{E}[\delta_{r\boldsymbol{\Theta}}\mathbb{1}\{\mathcal{C}_0(\boldsymbol{\Theta}) = 0\}]\alpha r^{-\alpha-1}\mathrm{d}r \; . \tag{6.2.5}$$

The measure $\boldsymbol{\nu}^*$ is boundedly finite on $(\mathbb{R}^d)^{\mathbb{Z}} \setminus \{\mathbf{0}\}$, puts no mass at $\mathbf{0}$ and is α-homogeneous. Furthermore, the cluster measure can be expressed in terms of another sequence.

Definition 6.2.2. *Assume that $\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$. The conditional spectral tail process $\boldsymbol{Q}$ is a random sequence with the distribution of $(\boldsymbol{Y}^*)^{-1}\boldsymbol{Y}$ conditionally on $\mathcal{C}_0(\boldsymbol{Y}) = 0$.*

The sequence $\boldsymbol{Q}$ appeared implicitly in the seminal paper [DH95]. See also [BS09], [PS18, Definition 3.5] and [KS20, Chapter 5]. An abstract setting is considered in [DHS18].

Note that $\mathcal{C}_0(\boldsymbol{Y}) = 0$ if and only if $\mathcal{C}_0(\boldsymbol{\Theta}) = 0$. Then also $\boldsymbol{Y}^* = |\boldsymbol{Y}_0|$. Thus, (6.2.5) and the definition of $\boldsymbol{Q}$ give for a bounded or non-negative measurable function H on $\ell_0(\mathbb{R}^d)$,

$$\boldsymbol{\nu}^*(H) = \vartheta \int_0^\infty \mathbb{E}[H(r\boldsymbol{Q})]\alpha r^{-\alpha-1}\mathrm{d}r = \vartheta \int_0^\infty \mathbb{E}[H(r\boldsymbol{\Theta})\mathbb{1}\{\mathcal{C}_0(\boldsymbol{\Theta}) = 0\}]\alpha r^{-\alpha-1}\mathrm{d}r \,. \tag{6.2.6}$$

If moreover H is such that $H(\boldsymbol{y}) = 0$ if $\boldsymbol{y}^* \le \epsilon$ for one $\epsilon > 0$, then

$$\boldsymbol{\nu}^*(H) = \epsilon^{-\alpha}\mathbb{E}[H(\epsilon\boldsymbol{Y})\mathbb{1}\{\mathcal{C}_0(\boldsymbol{Y}) = 0\}] = \epsilon^{-\alpha}\mathbb{E}[H(\epsilon\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \,. \tag{6.2.7}$$

Note that with $H(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\}$ and recalling that $\boldsymbol{Y}^* > 1$, (6.2.7) reduces to (6.2.4). As such, functionals from $\mathcal{A}$ will have typically the representation given in (6.2.7). On the other hand, for functionals from $\mathcal{B}$ we are not able to conclude the representation (6.2.7), however, the general form (6.2.6) is still valid, possibly under additional conditions. See Section 6.3.3. Comparing (6.2.5) or (6.2.7) with (6.2.2) we can see that the $\boldsymbol{\nu}^*(H)$ does not agree with $\mathbb{E}[H(\boldsymbol{Y})]$. The additional indicator comes essentially from the conditioning on the location of the maximum of the sequence $\boldsymbol{Y}$.

Definition 6.2.3 (Cluster index)**.** *We will call $\boldsymbol{\nu}^*(H)$ the cluster index associated to the functional H.*

6.3. Convergence of cluster measure

Recall $\mathcal{R}(r_n, u_n)$. Define the measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \ge 1$, on $\ell_0(\mathbb{R}^d)$ as follows:

$$\boldsymbol{\nu}^*_{n,r_n} = \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\mathbb{E}\left[\delta_{u_n^{-1}\boldsymbol{X}_{1,r_n}}\right] \,.$$

We are interested in convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$. The results of this section are extracted from [KS20, Chapter 6]. See also [PS18] and [BPS18].

6.3.1. Anticlustering condition

For each fixed $r \in \mathbb{N}$, the distribution of $u_n^{-1}\boldsymbol{X}_{-r,r}$ conditionally on $|\boldsymbol{X}_0| > u_n$ converges weakly to the distribution of $\boldsymbol{Y}_{-r,r}$. In order to let r tend to infinity, we must embed all these finite vectors into one space of sequences. By adding zeroes on each side of the vectors $u_n^{-1}\boldsymbol{X}_{-r,r}$ and $\boldsymbol{Y}_{-r,r}$ we identify them with elements of the space $\ell_0(\mathbb{R}^d)$. Then $\boldsymbol{Y}_{-r,r}$ converges (as $r \to \infty$) to $\boldsymbol{Y}$ in $\ell_0(\mathbb{R}^d)$ if (and only if) $\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)$ almost surely.

However, this is not enough for statistical purposes and we consider the following definition.

Definition 6.3.1 ([DH95], Condition 2.8)**.** *Condition $\mathcal{AC}(r_n, u_n)$ holds if for all $x, y > 0$,*

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\max_{k\le|j|\le r_n}|\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) = 0 \,. \tag{$\mathcal{AC}(r_n,u_n)$}$$

Condition $\mathcal{AC}(r_n, u_n)$ is referred to as the anticlustering condition. It is fulfilled by many models, including geometrically ergodic Markov chains, short-memory linear or max-stable processes. $\mathcal{AC}(r_n, u_n)$ implies that $\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)$. Its main consequence is the following result.

Proposition 6.3.2 ([BS09], Proposition 4.2; [KS20], Theorem 6.1.4)**.** *Let $H \in \mathcal{L}$. If Condition $\mathcal{AC}(r_n, u_n)$ holds, then*

$$\lim_{n\to\infty} \mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{-r_n,r_n}) \mid |\boldsymbol{X}_0| > u_n] = \mathbb{E}[H(\boldsymbol{Y})] \ .$$

Condition $\mathcal{AC}(r_n, u_n)$ holds for sequence of i.i.d. random variables whenever $\lim_{n\to\infty} r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$, which can be recognized as on the restrictions imposed in $\mathcal{R}(r_n, u_n)$

6.3.2. Vague convergence of cluster measure

We now investigate the unconditional convergence of $u_n^{-1}\boldsymbol{X}_{1,r_n}$. Contrary to Proposition 6.3.2, where an extreme value was imposed at time 0, a large value in the cluster can happen at any time. Moreover, the convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ to $\boldsymbol{\nu}^*(H)$ may hold only for shift-invariant functionals H. Therefore, we need the following definition.

Definition 6.3.3. *The space $\tilde{\ell}_0(\mathbb{R}^d)$ is the space of equivalence classes of $\ell_0(\mathbb{R}^d)$ endowed with the equivalence relation $\sim$ defined by*

$$\boldsymbol{x} \sim \boldsymbol{y} \iff \exists j \in \mathbb{Z}\ ,\ B^j\boldsymbol{x} = \boldsymbol{y}\ ,$$

where B is the backshift operator.

The proof of the next result is given in Section 6.7.

Proposition 6.3.4. *Let condition $\mathcal{AC}(r_n, u_n)$ hold. The sequence of measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \geq 1$ converges vaguely$^\#$ on $\tilde{\ell}_0(\mathbb{R}^d) \setminus \{\boldsymbol{0}\}$ to $\boldsymbol{\nu}^*$, that is, for all $H \in \mathcal{A}$,*

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \boldsymbol{\nu}^*(H)\ .$$

The immediate consequence is the following limit (cf. (6.2.3)):

$$\lim_{n\to\infty} \frac{\mathbb{P}(\boldsymbol{X}^*_{1,r_n} > u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \vartheta\ .$$

Since $H_2, H_3 \in \mathcal{A}$ (cf. (6.1.2)-(6.1.3)), we can introduce the following cluster indices.

Example 6.3.5 (Cluster size distribution)*.* If $\mathcal{AC}(r_n, u_n)$ holds, Proposition 6.3.4 yields

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{j=1}^{r_n} \mathbb{1}\{|\boldsymbol{X}_j| > u_n\} = m \mid \boldsymbol{X}^*_{1,r_n} > u_n\right)$$
$$= \mathbb{P}\left(\sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{Y}_j| > 1\} = m \mid \boldsymbol{Y}^*_{-\infty,-1} \leq 1\right) =: \pi(m)\ .$$

⊞

Example 6.3.6 (Stop-loss index)*.* Consider a univariate time series. Define the stop-loss index:

$$\theta_{\text{stoploss}}(\eta) = \lim_{n\to\infty} \frac{\mathbb{P}\left(\sum_{j=1}^{r_n}(X_j - u_n)_+ > \eta u_n\right)}{r_n\mathbb{P}(X_0 > u_n)} = \mathbb{P}\left(\sum_{j=0}^{\infty}(Y_j - 1)_+ > \eta, \boldsymbol{Y}^*_{-\infty,-1} \leq 1\right)\ .$$

This index seems to be new. ⊞

6.3.3. Indicator functionals not vanishing around zero

Proposition 6.3.4 entails convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ for $H \in \mathcal{A}$. For functionals which are not defined on the whole space $\ell_0(\mathbb{R}^d)$, such as H_4 and H_5 from (6.1.4)-(6.1.5), we need an additional assumption on Asymptotic Negligibility of Small Jumps.

Definition 6.3.7. *Condition* ANSJB(r_n, u_n) *holds if for all* $\eta > 0$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\mathbb{P}(\sum_{j=1}^{r_n} |\boldsymbol{X}_j| \mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0 \; . \qquad (\text{ANSJB}(r_n, u_n))$$

The proofs of the next two results are given in Section 6.7.

Lemma 6.3.8. *If* $\mathcal{AC}(r_n, u_n)$ *and* ANSJB(r_n, u_n) *hold, then*

$$\lim_{n \to \infty} \frac{\mathbb{P}(\sum_{i=1}^{r_n} |\boldsymbol{X}_j| > u_n)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \mathbb{E}\left[\left(\sum_{j \in \mathbb{Z}} |\boldsymbol{Q}_j|\right)^{\alpha}\right] < \infty \; .$$

Proposition 6.3.9. *Assume that* $\mathcal{AC}(r_n, u_n)$ *and* ANSJB(r_n, u_n) *hold. Then for* $K \in \mathcal{K}$,

$$\boldsymbol{\nu}^*(\mathbb{1}\{K > 1\}) = \lim_{n \to \infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \vartheta \int_0^{\infty} \mathbb{P}(K(z\boldsymbol{Q}) > 1) \alpha z^{-\alpha-1} \mathrm{d}z < \infty \; .$$

If K is a 1-homogeneous satisfying the assumptions of Proposition 6.3.9, then

$$\boldsymbol{\nu}^*(\mathbb{1}\{K > 1\}) = \vartheta \mathbb{E}[K_+^{\alpha}(\boldsymbol{Q})] = \mathbb{E}[K_+^{\alpha}(\boldsymbol{\Theta}_{0,\infty}) - K_+^{\alpha}(\boldsymbol{\Theta}_{1,\infty})] \; .$$

Example 6.3.10 (Large deviations index). Let $\{X_j, j \in \mathbb{Z}\}$ be an univariate time series. The functional H_4 defined in (6.1.4) yields the large deviations index:

$$\theta_{\text{largedev}} = \lim_{n \to \infty} \frac{\mathbb{P}\left(\left(\sum_{j=1}^{r_n} X_j\right)_+ > u_n\right)}{r_n \mathbb{P}(|X_0| > u_n)} = \mathbb{E}\left[\left(\sum_{j=0}^{\infty} \Theta_j\right)_+^{\alpha} - \left(\sum_{j=1}^{\infty} \Theta_j\right)_+^{\alpha}\right] \; .$$

The index θ_{largedev}, under the name *cluster index*, was introduced in [MW16]. ⊞

Example 6.3.11 (Ruin index). Take H_5 defined in (6.1.5). Proposition 6.3.9 gives

$$\theta_{\text{ruin}} = \lim_{n \to \infty} \frac{\mathbb{P}(\max_{1 \le j \le r_n} \sum_{i=1}^{j} X_i > u_n)}{r_n \mathbb{P}(|X_0| > u_n)} = \vartheta \mathbb{E}\left[\sup_{i \in \mathbb{Z}} \left(\sum_{j \le i} Q_j\right)_+^{\alpha}\right] \; .$$

⊞

Remark 6.3.12. At this point we would like to point out the following. Consider $H \in \mathcal{A}$ to be an indicator functional. If moreover H is such that $H(\boldsymbol{y}) = 0$ if $\boldsymbol{y}^* \le 1$, then thanks to (6.2.7),

$$\boldsymbol{\nu}^*(H) = \mathbb{E}[H(\boldsymbol{Y}) \mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \in (0, 1] \; .$$

This is the situation for the extremal index and the functionals from Examples 6.3.5 and 6.3.6. On the other hand, if H does not vanish around zero, then at the first place we need additional conditions to guarantee that $\boldsymbol{\nu}^*(H) < \infty$ (e.g. ANSJB(r_n, u_n)). Second, there is no restriction on the values of the cluster index.

6.4. Central limit theorem for blocks estimators

6.4.1. Sliding blocks estimators

Let $q_n = n - r_n + 1$. Thanks to Proposition 6.3.4 and Proposition 6.3.9, we have for $H \in \mathcal{A} \cup \mathcal{B}$,

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{\sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)}\right] = \lim_{n\to\infty} \frac{\mathbb{E}\left[H\left(\boldsymbol{X}_{1,r_n}/u_n\right)\right]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \boldsymbol{\nu}^*(H) \ .$$

This indicates that a consistent pseudo-estimator of $\boldsymbol{\nu}^*(H)$ can be defined as

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) := \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \ . \tag{6.4.1}$$

The above estimator is not feasible, since it involves an unspecified sequence $\{u_n\}$ and the tail of $|\boldsymbol{X}_0|$. Thus, in (6.4.1) we replace $q_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)$ with its empirical estimate $\sum_{j=1}^{q_n} \mathbb{1}\{|\boldsymbol{X}_j| > u_n\}$ to obtain a quasi-feasible estimator

$$\widehat{\widetilde{\boldsymbol{\mu}}}^*_{n,r_n}(H) = \frac{1}{r_n} \frac{1}{\sum_{j=1}^{q_n} \mathbb{1}\{|\boldsymbol{X}_j| > u_n\}} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \ .$$

Likewise, let k_n be an intermediate sequence of integers, i.e. $\lim_{n\to\infty} k_n = \infty$, $\lim_{n\to\infty} k_n/n = 0$. Define u_n by $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Replacing u_n with $|\boldsymbol{X}|_{(n:n-k_n)}$ and noting that (assuming for simplicity that there are not ties in the data)

$$\sum_{j=1}^{n} \mathbb{1}\left\{|\boldsymbol{X}_j| > |\boldsymbol{X}|_{(n:n-k_n)}\right\} = k_n \ ,$$

we obtain a feasible estimator of $\boldsymbol{\nu}^*(H)$ given in (6.1.9).

6.4.2. Weak dependence assumptions

For asymptotic normality, we need to strengthen the anticlustering condition $\mathcal{AC}(r_n, u_n)$.

Definition 6.4.1. *Condition $\mathcal{S}(r_n, u_n)$ holds if for all $s, t > 0$*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=m}^{r_n} \mathbb{P}(|\boldsymbol{X}_0| > u_n s, |\boldsymbol{X}_j| > u_n t) = 0 \ . \tag{$\mathcal{S}(r_n, u_n)$}$$

This condition implies that $\sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) < \infty$. The latter series appears explicitly in the statement for the limiting variance.

Dependence in $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ will be controlled by the β-mixing rates $\{\beta_n\}$. Recall $\mathcal{R}(r_n, u_n)$. Let $\{\ell_n\}$ be a sequence of integers such that $\lim_{n\to\infty} \ell_n = \infty$ and $\lim_{n\to\infty} \ell_n/r_n = 0$.

Definition 6.4.2. *Condition $\beta(r_n)$ holds if:*

1. *$\beta_j = O(j^{-\nu})$, $\nu > 1$ and $\lim_{n\to\infty} r_n^{1+\nu}/n = +\infty$; and*
2. *there exists $\delta > 0$ such that $\lim_{n\to\infty} r_n^{\nu-\delta} \mathbb{P}(|\boldsymbol{X}_0| > u_n) = +\infty$.*

From the basic assumptions on the time series, we have $\lim_{n\to\infty} r_n/n = 0$. Thus, ν has to be big enough. The above mixing condition is clearly satisfied for time series with geometric mixing rates since then ν can be chosen arbitrarily large.

6.4.3. Main result

Let $\mathbb{G}$ be the Gaussian process on $L^2(\boldsymbol{\nu}^*)$ with covariance

$$\operatorname{cov}(\mathbb{G}(H), \mathbb{G}(\widetilde{H})) = \boldsymbol{\nu}^*(H\widetilde{H}) .$$

Recall that for a functional $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ and $s > 0$ we define $H_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)$.

The main result of this paper is Theorem 6.4.3, the asymptotic normality of the appropriately normalized estimator $\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H)$. The limiting variance agrees with the one for the disjoint blocks estimator; cf. [DR10] and [KS20, Chapter 10].

Theorem 6.4.3. *Let $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n, u_n)$, $\beta(r_n)$, $\mathcal{S}(r_n, u_n)$ hold. Fix $0 < s_0 < 1 < t_0 < \infty$. Let $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ be a shift-invariant measurable map such that the class $\{H_s : s \in [s_0, t_0]\}$ is linearly ordered. Assume moreover that*

$$\lim_{n\to\infty} \sqrt{k_n} \sup_{s\in[s_0,t_0]} |\mathbb{E}[\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(\mathcal{E}_s)] - \boldsymbol{\nu}^*(\mathcal{E}_s)| = 0 , \tag{6.4.2a}$$

$$\lim_{n\to\infty} \sqrt{k_n} \sup_{s\in[s_0,t_0]} |\mathbb{E}[\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H_s)] - \boldsymbol{\nu}^*(H_s)| = 0 . \tag{6.4.2b}$$

If $H \in \mathcal{A}$, then

$$\sqrt{k_n} \left\{\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H) - \boldsymbol{\nu}^*(H)\right\} \xrightarrow{\mathrm{d}} \mathbb{G}(H - \boldsymbol{\nu}^*(H)\mathcal{E}) . \tag{6.4.3}$$

If moreover ANSJB(r_n, u_n) is satisfied, then (6.4.3) holds for $H \in \mathcal{B}$.

Remark 6.4.4. We note that the mixing conditions for validity of the central limit theorem in Theorem 6.4.3 are stronger than those for consistency in Theorem 4.3.3. Indeed, if $\beta_j = j^{-\nu}$, then

$$\frac{n}{r_n}\beta_{r_n} = \frac{n}{r_n} r_n^{-\nu} = \frac{n}{r_n^{1+\nu}}$$

which converges to zero by the assumption. Since ℓ_n can be chosen as close as possible to r_n, the above convergence is in line with the mixing assumption of Theorem 4.3.3. In Theorem 6.4.3 we need also the second item in $\beta(r_n)$.

Additionally, the anticlustering condition used in Theorem 6.4.3 is stronger than the one needed in Theorem 4.3.3.

Remark 6.4.5. The limiting distribution is centered Gaussian with variance (cf. Lemma 6.7.22):

$$\begin{aligned}\boldsymbol{\nu}^*(\{H - \boldsymbol{\nu}^*(H)\mathcal{E}\}^2) &= \boldsymbol{\nu}^*(H^2) - 2\boldsymbol{\nu}^*(H)\boldsymbol{\nu}^*(H\mathcal{E}) + (\boldsymbol{\nu}^*(H))^2\boldsymbol{\nu}^*(\mathcal{E}^2) \\ &= \boldsymbol{\nu}^*(H^2) - 2\boldsymbol{\nu}^*(H)\mathbb{E}[H(\boldsymbol{Y})] + (\boldsymbol{\nu}^*(H))^2 \sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) .\end{aligned}$$

Thus, in view of (6.1.7), the limiting variance for the sliding blocks estimator agrees with the one for the disjoint blocks estimator. ⊕

Remark 6.4.6. The linear ordering of the class $\{H_s : s \in [s_0, t_0]\}$ may seem to be too restrictive. However, all the cluster functionals that, from the authors perspective, seem to be of interest in the context of sliding blocks estimators, have this property. This includes the functionals H_1, H_2, H_3, H_4, H_5 considered in the Introduction. The linear ordering can be replaced with an assumption that the function class is of VC-type (see [DR10]), or can be approximated by VC-classes. See [BBKS20, Lemma A.3], [DK20] and [KS20, Appendix C.4]. This generalization is obvious, but does not bring anything to the contents of the paper, since the random entropy assumption has to be checked for each case separately.

6.4.4. Examples

Example 6.4.7 (Extremal index). For $H(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\}$ we have $\boldsymbol{\nu}^*(H) = \boldsymbol{\nu}^*(H^2) = \vartheta$. The data-based estimator (6.1.9) is asymptotically normal with mean zero and the limiting variance is

$$\boldsymbol{\nu}^*(\{H - \boldsymbol{\nu}^*(H)\mathcal{E}\}^2) = \vartheta^2 \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1) - \vartheta \; .$$

See Section 6.5 for a discussion on the existing results. ⊞

Example 6.4.8 (Cluster size distribution). Consider the situation from Example 6.3.5. The limiting distribution is centered normal with the variance

$$\pi(m)\,(1 - 2\mathbb{P}(\mathcal{E}(\boldsymbol{Y}) = m)) + \pi^2(m) \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1) \; .$$

⊞

Example 6.4.9 (Stop-loss index). Consider the stop-loss index $\theta_{\mathrm{stoploss}}(\eta)$ introduced in Example 6.3.6. The limiting distribution is centered normal with the variance

$$\theta_{\mathrm{stoploss}}(\eta) \left(1 - 2\mathbb{P}\left(\sum\nolimits_{j\in\mathbb{Z}} (Y_j - 1)_+ > \eta\right)\right) + \theta^2_{\mathrm{stoploss}}(\eta) \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1) \; .$$

⊞

Example 6.4.10 (Large deviations index). We continue with the situation from Example 6.3.10. The limiting distribution is centered Gaussian with variance

$$\theta_{\mathrm{largedev}} - 2\theta_{\mathrm{largedev}} \mathbb{P}\left(\left(\sum_{j\in\mathbb{Z}} Y_j\right)_+ > 1\right) + \theta^2_{\mathrm{largedev}} \sum_{j\in\mathbb{Z}} \mathbb{P}(|Y_j| > 1) \; .$$

⊞

6.5. Comments and extensions

6.5.1. Existing results

We discuss the existing results. For the sake of clarify, we consider univariate, non-negative, regularly varying time series with the marginal distribution F.

PoT approach. In [DN20] the authors study asymptotic normality of the sliding blocks estimators in a general setting. They show that the limiting variance of such estimators does not exceed the one for the disjoint blocks estimators. For the extremal index they found the variances to be equal. As in this paper, they use the threshold u_n such as in $\mathcal{R}(r_n, u_n)$. The results in [DR10] and [KS20, Chapters 9-10] (disjoint blocks) as well as in [DN20] and in the current paper fit into Peaks-Over-Threshold (PoT) framework.

The results in [DN20] warrant a more detailed comparison. The authors consider there a more general framework (beyond regular variation). At the same time (under some monotonicity conditions on functionals H), they show that sliding blocks estimators yield variance that does not exceed the one for disjoint blocks estimators. In our case we have a precise calculation of the limiting variance, which is not accessible using Drees and Neblung method. Also, [DN20] work with some ad-hoc conditions that have to be verified model by model. In our case, we basically need the anticlustering (along with mixing). For example, [DN20] assume the specific behaviour of covariances, see (C) in their paper. We do not assume this - we prove it and provide the form of the limiting covariance. This can be viewed as our the first major contribution. Its proof is long, but covers all the functionals of interest.

In particular, consider the disjoint blocks estimator of the extremal index,

$$\widetilde{\vartheta}_{n,1} = \widetilde{\vartheta}_{n,1}(x) = \frac{\sum_{i=1}^{m_n} \mathbb{1}\left\{\boldsymbol{X}^*_{(i-1)r_n+1,ir_n} > x\right\}}{\sum_{j=1}^{m_n r_n} \mathbb{1}\{X_j > x\}} .$$

In [KS20, Example 10.4.2] we calculated the limiting variance of $\widetilde{\vartheta}_{n,1}(u_n)$ to be $\sigma_1^2 = -\vartheta + \vartheta^2 \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1)$. This is in agreement with Corollary 4.6 in [Hsi91] (where the variance σ_1^2 is given in a complicated form). We can see that σ_1^2 agrees with the one limiting variance for the sliding blocks estimator in Example 6.4.7. The blocks estimator $\widetilde{\vartheta}_{n,1}(u_n)$ is also considered in [SW94] and [WN98].

Block maxima framework. One can also use the threshold c_{r_n} given by

$$r_n \overline{F}(c_{r_n}) \to 1 . \tag{6.5.1}$$

We are not aware of the asymptotic theory for $\widetilde{\vartheta}_{n,1}(c_{r_n})$. However, using [RSF09, Theorem 4.2] and the delta method we can compare the variances of $\widetilde{\vartheta}_{n,1}(u_n)$ and $\widetilde{\vartheta}_{n,1}(c_{r_n})$:

$$\sigma_1^2 = -\vartheta + \vartheta^2 \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1) \ \text{ vs. } \sigma_3^2 := \mathrm{e}^{-\vartheta}(1-\mathrm{e}^{-\vartheta}) - 2\vartheta\mathrm{e}^{-\vartheta} + \vartheta^2 \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_j > 1) .$$

Thus, the estimator $\widetilde{\vartheta}_{n,1}(u_n)$ has a smaller variance than $\widetilde{\vartheta}_{n,1}(c_{r_n})$.

In the following discussion, we will use the threshold (6.5.1). In [RSF09] the authors consider another disjoint blocks estimator of the extremal index, motivated by the approximation $\log(1-x) \sim x$ $(x \to 0)$. Also, the corresponding sliding blocks estimator is considered. It is shown that the sliding blocks one yields a smaller asymptotic variance.

In [BS18b, BS18a] the authors estimate the parameters (α, σ) of the Fréchet distribution stemming from the limiting behaviour of the maxima. Disjoint blocks yield a larger variance than sliding blocks. Similarly, in [BB18] the authors use the blocking method to estimate the extremal index and again the sliding block estimator is more efficient.

The estimator $\widetilde{\vartheta}_{n,1}(c_{r_n})$ as well as the ones in [RSF09] and [BS18b, BS18a] can be thought of as the application of the block maxima method. Indeed, the threshold c_{r_n} is the normalizing sequence for the limiting distribution of maxima. In the context of the latter two papers, $\boldsymbol{X}^*_{1,r_n}/\sigma_{r_n}$ converges in distribution to a standard Fréchet random variable with tail index α (denoted by Z). On the other hand, for $\xi \in (0,1)$, the pair

$$(\boldsymbol{X}^*_{1,r_n}/\sigma_{r_n}, \boldsymbol{X}^*_{1+[\xi r_n],r_n+[\xi r_n]}/\sigma_{r_n})$$

converges in distribution to a dependent random vector (Z_1, Z_2) with Fréchet marginals and parametrized by $\xi \in (0,1)$. See [BS18a, Lemma 5.1]. Consider now for $f : \mathbb{R} \to \mathbb{R}$

$$\mathbb{G}_n^{(BS)}(f) = \sqrt{m_n}\left\{ m_n^{-1} \sum_{j=1}^{m_n} f\left(\frac{\boldsymbol{X}^*_{(j-1)r_n+1,jr_n}}{\sigma_{r_n}}\right) - \mathbb{E}[f(Z)]\right\} ,$$

$$\mathbb{F}_n^{(BS)}(f) = \sqrt{m_n}\left\{ q_n^{-1} \sum_{i=1}^{q_n} f\left(\frac{\boldsymbol{X}^*_{i,i+r_n-1}}{\sigma_{r_n}}\right) - \mathbb{E}[f(Z)]\right\} .$$

The aforementioned convergence gives the limiting variance. For the disjoint blocks empirical process the limiting variance is $\mathrm{var}(f(Z))$, while for the sliding blocks one it becomes (cf. Lemma 5.3 in [BS18a])

$$C(f) := 2\int_0^1 \mathrm{cov}_\xi(f(Z_1), f(Z_2))\mathrm{d}\xi \; .$$

In the context of our paper, if we choose $f(z) = \mathbb{1}\{z > 1\}$, then we can evaluate:

$$\mathrm{var}(f(Z)) = \exp(-1) - \exp(-2) > C(f) = 2\exp(-1) - 4\exp(-2) \; .$$

In the PoT framework considered in our paper, both the disjoint blocks and the sliding blocks empirical processes yield the limiting variance $\boldsymbol{\nu}^*(H)$.

In summary:

- **The PoT method, as proven in this paper, gives the same limiting behaviour for both disjoint and sliding blocks estimators.**
- The situation seems to be different in case of the block maxima method, at least for the inference problems considered up to date.
- One can argue that the blocks maxima method is restricted to estimation of the parameters of the limiting distribution of maxima (the tail index, the extremal index) and is rather hard to see how the method can be employed to other cluster indices.

6.5.2. Bias

The sliding block estimators of cluster functionals are subjected to bias

$$\mathbb{E}[\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H)] - \boldsymbol{\nu}^*(H) \; .$$

The bias vanishes asymptotically thanks to the assumption (6.4.2b). The latter assumption imposes some restrictions on k_n, r_n. Classically, e.g. in case of the Hill estimator of the tail index, the bias is controlled by the second order condition. In the present context we know nothing about how to control bias, except that *there exist sequences k_n and r_n such that* (6.4.2b) *holds.* On a positive side, from a point of comparing disjoint and sliding blocks estimators, the theoretical bias is obviously the same for both.

6.5.3. Open questions

- For the sliding blocks estimators, obtain consistency under minimal conditions (that is, without relying on β-mixing). In [KS20, Chapter 10] we obtain consistency of the disjoint blocks estimators for time series that can be approximated by m-dependent sequences, including long memory ones.
- Extend Theorem 6.4.3 to unbounded functionals H. The method of the proof presented in the paper should be applicable, however, some substantial modifications may be needed. Certainly, more restrictive conditions will need to be implemented.
- In view of the behaviour of $\widetilde{\vartheta}_{n,1}(u_n)$ and $\widetilde{\vartheta}_{n,1}(c_{r_n})$, it would be interesting to know if (whenever possible) the PoT method always gives a smaller variance than the block maxima ones.

6.6. Simulation study

We conducted some simulations in order to study the finite sample performance of the sliding and disjoint blocks estimators for selected cluster indices.

6.6.1. Stationary AR process

We start with a simple AR(1) process. For this process we have explicit formulas for all cluster indices. Samples of size $n = 1000$ are generated from AR(1) with $\alpha = 4$ and $\rho = 0.5, 0.9$. Simulations for the classical extremal index are compared to simulations for the stop-loss index.

Extremal index. For AR(1) with $\rho > 0$ the extremal index is $\theta = 1 - \rho^\alpha$; cf. [KS20, p. 396].

- We start with the Hill plots in Figure 6.1. There, for one simulated data set we compare a performance of both disjoint and sliding blocks estimators. For weak dependence ($\rho = 0.5$), both sliding and disjoint blocks estimators under-estimate the extremal index while for strong dependence ($\rho = 0.9$), the results are stable around the true extremal index for a small number of order statistics. Bigger values of k introduce more bias. In any case, the performance of both estimators is comparable.
- Table 6.1 includes the results for Monte Carlo simulation for the extremal index based on disjoint and sliding blocks, with the block size $r_n = 7,\ 8,\ 9,\ 10$. We used $k = 5\%$ and 10% order statistics. We note that for the strong dependence ($\rho = 0.9$), the estimation is acceptable for the small block sizes and small k for both disjoint and sliding estimators. A larger block size r_n and/or larger number of order statistics k results in a biased estimation. For weak dependence ($\rho = 0.5$), the results are heavily biased for all considered parameters. We note that both disjoint and sliding blocks estimators yield almost the same variances, which is in agreement with the theoretical results obtained in the paper.
- The box plots and histograms in Figure 6.2 and Figure 6.3 are based again on Monte Carlo simulations. The following parameters are used: $\rho = 0.9$, $\alpha = 4$ and the block size $r_n = 7$ along with $k = 5\%$ and 10%. We notice again that $\rho = 0.9$, $r_n = 7$ and $k = 5\%$ yield acceptable results. However, small ρ yields a lot of bias.

In summary, **in case of the extremal index, both disjoint and sliding blocks estimators yield similar results (as suggested by theory). Stronger dependence implies better performance. Typically estimators suffer from a bias.** As such, bias-reduction techniques should be investigated.

We note also that the fact that stronger dependence yields smaller variability of the estimators is not surprising, cf. e.g. Figure 5 in [RSF09].

Stop-loss index. For AR(1) with $\rho > 0$ the formula for the stop-loss index is given in [KS20, p. 619]:

$$\theta_{\mathrm{stop-loss}}(S) = (1 - \rho^\alpha)\mathbb{P}\left(\sum_{j=0}^{\infty}(\rho^j Y_0 - 1)_+ > S\right), \tag{6.6.1}$$

where Y_0 is a Pareto random variable with the parameter α.

- At the first step we use the formula (6.6.1) and performed the Monte-Carlo simulation to obtain the approximate value of the stop-loss index.
- Figure 6.4 displays Hill plots for the stop-loss index. The Hill plots indicate a similar performance of both disjoint and sliding blocks estimators. We note that, unlike in the extremal index case or in the classical case of the Hill estimator of the tail index, one needs to take a much bigger number of the order statistics. We do not have clear explanation for this.
- With this in mind, we performed simulation studies for $k = 50\%$ and $k = 70\%$. We noticed then (see Table 6.2) that, as opposed to the extremal index, the weaker dependence ($\rho = 0.5$) yields a good estimation for any given block size, while for the strong dependence ($\rho = 0.9$), the simulation results are rather poor. This may be quite intuitive, since the

stop-loss functional is based on *sums* of large values. In any case, both sliding and disjoint blocks estimators yield comparable results.

In summary, **in case of the stop-loss index, both disjoint and sliding blocks estimators yield similar results (as suggested by theory). Weaker dependence implies better performance. One should use a high number of order statistics.**

6.6.2. Stationary ARCH process

We consider a stationary ARCH(1) process defined by $X_j^2 = \sqrt{\beta + \lambda X_{j-1}^2} Z_j$, where $\{Z_j, j \in \mathbb{Z}\}$ are i.i.d standard normal random variables. For $\lambda = 0.9$ the extremal index is $\theta = 0.612$ (see [EKM97, p. 480]).

- We start with Hill plots in Figure 6.5. The plots again illustrate little difference between the disjoint and sliding blocks estimators.
- Monte Carlo results are included in Table 6.3 and visualised as boxplots in Figure 6.6.

Again, **both disjoint and sliding blocks estimators yield similar results (as suggested by the theory)**.

6.6.3. Other cluster indices

We conducted simulation studies for other cluster indices, in particular for those from class $\mathcal{B}$. Here the results are rather not very promising. One of the reasons is the following observation. Recall (see Remark 6.3.12) that unlike in class $\mathcal{A}$, the cluster indices for functional from class $\mathcal{B}$ can have arbitrary values. At the same time, the largest possible value of the disjoint blocks estimator is m_n/k_n. In theory, this ratio is

$$\frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)}$$

and hence diverges to infinity thanks to $\mathcal{R}(r_n, u_n)$. However, for finite samples, the ratio stays bigger than one for very few values of k_n. As such, we believe that alternative methods of estimation of e.g. the large deviation index have to be implemented.

6.7. Proofs

In Section 6.7.1 we prove the equivalence between (6.2.3) and (6.2.4). In Section 6.7.3 we show that (6.1.1) holds for $H \in \mathcal{A} \cup \mathcal{B}$. The proofs in that section stem from [KS20]. The results from Section 6.7.3 are extended in Section 6.7.4 to covariance of clusters. In Section 6.7.5 we introduce the empirical process of sliding blocks and state its functional convergence. The proof of the latter is separated into several parts. First, in Section 6.7.6 we derive the limiting covariance of the empirical process of sliding blocks. Next, in Section 6.7.7 we prove the finite-dimensional convergence. Asymptotic continuity is dealt with in Section 6.7.8. We conclude the proof in Section 6.7.9.

6.7.1. Representations of the (candidate) extremal index

We first quote the time-change formula (see [BS09], [KS20, Theorem 5.3.1]). Let B be the backshift operator on $(\mathbb{R}^d)^{\mathbb{Z}}$ defined by $(B\boldsymbol{x})_i = \boldsymbol{x}_{i-1}$, $i \in \mathbb{Z}$.

Lemma 6.7.1. *Let $\boldsymbol{Y}$ be the tail process. Let H be a bounded or non-negative measurable functional on $(\mathbb{R}^d)^{\mathbb{Z}}$. Then for all $i \in \mathbb{Z}$,*

$$\mathbb{E}[H(B^i\boldsymbol{Y})\mathbb{1}\{|\boldsymbol{Y}_{-i}| > 1\}] = \mathbb{E}[H(\boldsymbol{Y})\mathbb{1}\{|\boldsymbol{Y}_i| > 1\}] \ .$$

Lemma 6.7.2. *Assume that $\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$. Then*

$$\vartheta = \mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) = 0) = \mathbb{P}(\boldsymbol{Y}^*_{-\infty,-1} \le 1) \ .$$

Proof. Recall that $\mathcal{C}_0(\boldsymbol{y}) = \inf\{j : \boldsymbol{y}^*_{-\infty,j} = \boldsymbol{y}^*\}$. From the definition of the tail process we have $\boldsymbol{Y}^* > 1$. Thus

$$\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) = 0) = \mathbb{P}\left(\sup_{j\le -1}|\boldsymbol{Y}_j| < |\boldsymbol{Y}_0|, \sup_{j\ge 1}|\boldsymbol{Y}_j| \le |\boldsymbol{Y}_0|, \boldsymbol{Y}^* > 1\right) \ .$$

By the assumption $\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$, the maximum $\boldsymbol{Y}^*$ is achieved at some $|\boldsymbol{Y}_i|$, $i \in \mathbb{Z}$. We split the event $\{\boldsymbol{Y}^* > 1\}$ according to the first exceedence over 1 and then apply the time-change formula:

$$\begin{aligned}
\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) = 0) &= \sum_{i\in\mathbb{Z}} \mathbb{P}\left(\boldsymbol{Y}^*_{-\infty,-1} < |\boldsymbol{Y}_0|, \boldsymbol{Y}^*_{1,\infty} \le |\boldsymbol{Y}_0|, \boldsymbol{Y}^*_{-\infty,i-1} \le 1, |\boldsymbol{Y}_i| > 1\right) \\
&= \sum_{i\in\mathbb{Z}} \mathbb{P}\left(\boldsymbol{Y}^*_{-\infty,-i-1} < |\boldsymbol{Y}_{-i}|, \boldsymbol{Y}^*_{-i+1,\infty} \le |\boldsymbol{Y}_{-i}|, \boldsymbol{Y}^*_{-\infty,-1} \le 1, |\boldsymbol{Y}_{-i}| > 1\right) \\
&= \sum_{i\in\mathbb{Z}} \mathbb{P}(C_i \cap \{\boldsymbol{Y}^*_{-\infty,-1} \le 1\})
\end{aligned}$$

with

$$C_i = \{\boldsymbol{Y}^*_{-\infty,-i-1} < |\boldsymbol{Y}_{-i}|, \boldsymbol{Y}^*_{-i+1,\infty} \le |\boldsymbol{Y}_{-i}|, |\boldsymbol{Y}_{-i}| > 1\} \ .$$

The events C_i are disjoint, their union gives $\{\boldsymbol{Y}^* > 1\}$ and the latter event holds with probability one. Thus

$$\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) = 0) = \mathbb{P}(\{\boldsymbol{Y}^* > 1\} \cap \{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}) = \mathbb{P}(\boldsymbol{Y}^*_{-\infty,-1} \le 1) \ .$$

□

The next lemma shows that there are other possible representations for ϑ.

Lemma 6.7.3. *Assume that $\mathbb{P}(\mathcal{C}_0(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$. Then*

$$\vartheta = \mathbb{P}(\boldsymbol{Y}^*_{1,\infty} \le 1) \ .$$

Proof. We use again $\boldsymbol{Y}^* > 1$ and the fact that the maximum $\boldsymbol{Y}^*$ is achieved at some $|\boldsymbol{Y}_i|$. This time we split the event $\{\boldsymbol{Y}^* > 1\}$ according to the last exceedence over 1, and then apply the time-change formula:

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Y}^*_{-\infty,-1} \le 1) = \mathbb{P}(\boldsymbol{Y}^*_{-\infty,-1} \le 1, \boldsymbol{Y}^* > 1) \\
&= \sum_{i\in\mathbb{Z}} \mathbb{P}(\boldsymbol{Y}^*_{-\infty,-1} \le 1, |\boldsymbol{Y}_i| > 1, \boldsymbol{Y}^*_{i+1,\infty} \le 1) \\
&= \sum_{i\in\mathbb{Z}} \mathbb{P}(\boldsymbol{Y}^*_{-\infty,-i-1} \le 1, |\boldsymbol{Y}_{-i}| > 1, \boldsymbol{Y}^*_{1,\infty} \le 1) \\
&= \mathbb{P}(\boldsymbol{Y}^* > 1, \boldsymbol{Y}^*_{1,\infty} \le 1) = \mathbb{P}(\boldsymbol{Y}^*_{1,\infty} \le 1) \ .
\end{aligned}$$

□

In fact, we can replace $\mathcal{C}_0$ with any anchoring map; see [KS20, Theorem 5.5.3] for more details.

6.7.2. Consequences of the mixing assumption

Since ℓ_n can be chosen as $r_n^{1-\delta}$ ($\delta > 0$) with δ arbitrarily close to zero, $\beta(r_n)$ gives:

$$\lim_{n\to\infty} \frac{n}{r_n}\beta_{r_n} = 0 \;, \tag{6.7.1a}$$

$$\lim_{n\to\infty} \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=\ell_n}^{\infty} \beta_j = 0 \;, \tag{6.7.1b}$$

$$\lim_{n\to\infty} \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\infty} \beta_{jr_n} = 0 \;. \tag{6.7.1c}$$

We recall the covariance inequality for bounded, beta-mixing random variables (in fact, the inequality holds for α-mixing). Let $\beta(\mathcal{F}_1, \mathcal{F}_2)$ be the β-mixing coefficient between two sigma fields. Then ([Ibr62])

$$|\mathrm{cov}(H(Z_1), H(Z_2))| \le \mathrm{cst}\; \|H\|_\infty \|\widetilde{H}\|_\infty \beta(\sigma(Z_1), \sigma(Z_2)) \;. \tag{6.7.2}$$

In (6.7.2) the constant cst does not depend on $H, \widetilde{H}$.

6.7.3. Convergence of cluster measure

Proof of Proposition 6.3.4. Since H has a support separated from zero, there exists $\epsilon > 0$ such that $H(\boldsymbol{x}) = 0$ if $\boldsymbol{x}^* \le \epsilon$. Applying its shift invariance and the stationarity, we obtain

$$\begin{aligned}
\boldsymbol{\nu}^*_{n,r_n}(H) &= \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{r_n} \mathbb{E}\left[H(u_n^{-1}\boldsymbol{X}_{1,r_n})\mathbb{1}\{\boldsymbol{X}^*_{1,j-1} \le u_n\epsilon\}\mathbb{1}\{|\boldsymbol{X}_j| > u_n\epsilon\}\right] \\
&= \frac{\mathbb{P}(|\boldsymbol{X}_0| > u_n\epsilon)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \frac{1}{r_n} \sum_{j=1}^{r_n} \mathbb{E}\left[H(u_n^{-1}\boldsymbol{X}_{1-j,r_n-j})\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1} \le u_n\epsilon\} \mid |\boldsymbol{X}_0| > u_n\epsilon\right] \\
&= \frac{\mathbb{P}(|\boldsymbol{X}_0| > u_n\epsilon)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \int_0^1 g_n(v)\mathrm{d}v \;,
\end{aligned}$$

with

$$g_n(v) = \mathbb{E}\left[H(u_n^{-1}\boldsymbol{X}_{1-[r_nv],r_n-[r_nv]})\mathbb{1}\left\{\boldsymbol{X}^*_{1-[r_nv],-1} \le u_n\epsilon\right\} \mid |\boldsymbol{X}_0| > u_n\epsilon\right] \;.$$

By Proposition 6.3.2, $\lim_{n\to\infty} g_n(v) = \mathbb{E}[H(\epsilon\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}]$ for each $v \in (0,1)$. Moreover, the sequence g_n is uniformly bounded, thus by dominated convergence, regular variation of $|\boldsymbol{X}_0|$ and (6.2.7), we obtain

$$\lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \epsilon^{-\alpha}\mathbb{E}[H(\epsilon\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] = \boldsymbol{\nu}^*(H) \;.$$

□

Proof of Lemma 6.3.8. By Proposition 6.3.4 and (6.2.6), we have

$$\lim_{n\to\infty} \frac{\mathbb{P}(\sum_{j=1}^{r_n} |\boldsymbol{X}_j|\,\mathbb{1}\{|\boldsymbol{X}_j| > \epsilon u_n\} > u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = \vartheta \int_0^\infty \mathbb{P}\left(z\sum_{j\in\mathbb{Z}} |\boldsymbol{Q}_j|\,\mathbb{1}\left\{z\,|\boldsymbol{Q}_j| > \epsilon\right\} > 1\right) \alpha z^{-\alpha-1}\mathrm{d}z \;. \tag{6.7.3}$$

By monotone convergence, the right hand side converges as $\epsilon \to 0$ to $\vartheta\mathbb{E}\left[\left(\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\right)^\alpha\right]$.

Consider the function

$$g(\zeta) = \vartheta \int_0^\infty \mathbb{P}\left(z\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\mathbb{1}\{z|\boldsymbol{Q}_j|>\zeta\}>1\right)\alpha z^{-\alpha-1}\mathrm{d}z \; .$$

It increases when ζ decreases to zero and its limit is $\vartheta\mathbb{E}\left[\left(\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\right)^\alpha\right]$. To prove that this quantity is finite, it suffices to prove that the function g is bounded. Fix $\epsilon>0$ and $\eta\in(0,1)$. By ANSJB(r_n,u_n), there exists ζ such that

$$\limsup_{n\to\infty}\frac{\mathbb{P}(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\,\mathbb{1}\{|\boldsymbol{X}_j|\le\zeta u_n\}>\eta u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\le\epsilon \; .$$

Fix $\zeta'<\zeta$. Starting from (6.7.3) and applying ANSJB(r_n,u_n), we obtain

$$\begin{aligned}
0 \le g(\zeta') &= \vartheta \int_0^\infty \mathbb{P}\left(z\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\mathbb{1}\{z|\boldsymbol{Q}_j|>\zeta'\}>1\right)\alpha z^{-\alpha-1}\mathrm{d}z \\
&= \lim_{n\to\infty}\frac{\mathbb{P}(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\mathbb{1}\{|\boldsymbol{X}_j|>u_n\zeta'\}>u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \\
&= \lim_{n\to\infty}\frac{\mathbb{P}\left(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\mathbb{1}\{|\boldsymbol{X}_j|>u_n\zeta\}+|\boldsymbol{X}_j|\mathbb{1}\{u_n\zeta\ge|\boldsymbol{X}_j|>\epsilon u_n\zeta'\}>u_n\right)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \\
&\le \limsup_{n\to\infty}\frac{\mathbb{P}(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\,\mathbb{1}\{|\boldsymbol{X}_j|\le u_n\zeta\}>\eta u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \\
&\qquad + \lim_{n\to\infty}\frac{\mathbb{P}(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\,\mathbb{1}\{|\boldsymbol{X}_i|>u_n\zeta\}>(1-\eta)u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \\
&\le \epsilon+\vartheta\int_0^\infty \mathbb{P}\left(z\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\,\mathbb{1}\left\{z\,|\boldsymbol{Q}_j|>\zeta\right\}>1-\eta\right)\alpha z^{-\alpha-1}\mathrm{d}z\le\epsilon+\vartheta\zeta^{-\alpha} \; .
\end{aligned}$$

The latter bound holds since the probability inside the integral is zero if $z\le\zeta$ since $|\boldsymbol{Q}_j|\le 1$ for all j. This proves that the function g is bounded in a neighbourhood of zero as claimed.

By Condition ANSJB(r_n,u_n), we finally obtain

$$\begin{aligned}
\lim_{n\to\infty}\frac{\mathbb{P}\left(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|>u_n\right)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} &= \lim_{\epsilon\to0}\lim_{n\to\infty}\frac{\mathbb{P}\left(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\mathbb{1}\{|\boldsymbol{X}_j|>u_n\epsilon\}>u_n\right)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \\
&= \lim_{\epsilon\to0}\vartheta\int_0^\infty\mathbb{P}\left(z\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\mathbb{1}\{z|\boldsymbol{Q}_j|>\epsilon\}>\epsilon\right)\alpha z^{-\alpha-1}\mathrm{d}z=\vartheta\mathbb{E}\left[\left(\sum_{j\in\mathbb{Z}}|\boldsymbol{Q}_j|\right)^\alpha\right] \; .
\end{aligned}$$

□

Proof of Proposition 6.3.9. For $\epsilon>0$, we define the truncation operator T_ϵ by

$$T_\epsilon(\boldsymbol{x}) = \{\boldsymbol{x}_j\mathbb{1}_{\{|\boldsymbol{x}_j|>\epsilon\}}, j\in\mathbb{Z}\} \; . \tag{6.7.4}$$

The operator T_ϵ is continuous with respect to the uniform norm at every $\boldsymbol{x}\in\ell_0$ such that $|\boldsymbol{x}_j|\ne\epsilon$ for all $j\in\mathbb{Z}$.

Fix $\eta \in (0,1)$ and $\zeta > 0$. Let L_K be as in (6.2.1) and choose $\epsilon > 0$ such that

$$\limsup_{n\to\infty} \frac{\mathbb{P}(\sum_{j=1}^{r_n} |\boldsymbol{X}_j| \mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n/L_K)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \le \zeta .$$

Set $K_\epsilon = K \circ T_\epsilon$. Applying assumption (6.2.1), we obtain

$$\begin{aligned}
&\frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\
&\le \frac{\mathbb{P}(K_\epsilon(\boldsymbol{X}_{1,r_n}/u_n) > 1-\eta)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} + \frac{\mathbb{P}(|K(\boldsymbol{X}_{1,r_n}/u_n) - K_\epsilon(\boldsymbol{X}_{1,r_n}/u_n)| > \eta)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\
&\le \frac{\mathbb{P}(K_\epsilon(\boldsymbol{X}_{1,r_n}/u_n) > 1-\eta)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} + \frac{\mathbb{P}(\sum_{i=1}^{r_n} |\boldsymbol{X}_j| \mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n/\mathrm{cst})}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} .
\end{aligned}$$

Applying Proposition 6.3.4 to K_ϵ along with the representation (6.2.6), this yields

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} &\le \limsup_{n\to\infty} \frac{\mathbb{P}(K_\epsilon(\boldsymbol{X}_{1,r_n}/u_n) > 1-\eta)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} + \zeta \\
&= \int_0^\infty \mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > 1-\eta)\alpha z^{-\alpha-1}\mathrm{d}z + \zeta .
\end{aligned}$$

Similarly,

$$\liminf_{n\to\infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \ge \int_0^\infty \mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > 1+\eta)\alpha z^{-\alpha-1}\mathrm{d}z - \zeta .$$

Since $K(\boldsymbol{0}) = 0$, (6.2.1) implies that $|K(\boldsymbol{x})| \le \mathrm{cst} \sum_{j\in\mathbb{Z}} |\boldsymbol{x}_j|$, thus for all $y > 0$,

$$\mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > y) \le \mathbb{P}\left(\sum_{j\in\mathbb{Z}} z\left|\boldsymbol{Q}_j\right| > y/\mathrm{cst}\right)$$

and the latter quantity is integrable (as a function of z) with respect to $\alpha z^{-\alpha-1}\mathrm{d}z$ in view of ANSJB(r_n, u_n) and Lemma 6.3.8. By bounded convergence, this yields

$$\lim_{\epsilon\to 0} \int_0^\infty \mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > y)\alpha z^{-\alpha-1}\mathrm{d}z = \int_0^\infty \mathbb{P}(K(z\boldsymbol{Q}) > y)\alpha z^{-\alpha-1}\mathrm{d}z .$$

Altogether, we obtain

$$\begin{aligned}
\int_0^\infty \mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > 1+\eta)\alpha z^{-\alpha-1}\mathrm{d}z - \zeta &\le \liminf_{n\to\infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \\
&\le \limsup_{n\to\infty} \frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n) > 1)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \le \int_0^\infty \mathbb{P}(K_\epsilon(z\boldsymbol{Q}) > 1-\eta)\alpha z^{-\alpha-1}\mathrm{d}z + \zeta .
\end{aligned}$$

Since ζ and η are arbitrary, this finishes the proof. □

6.7.4. Covariance of clusters

We consider the limit

$$\lim_{n\to\infty} \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathrm{cov}\left(H(\boldsymbol{X}_{1,r_n}/u_n), \widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\right)$$

for different choices of h, possibly depending on n. Under the conditions of Proposition 6.3.4, if moreover $\mathcal{R}(r_n, u_n)$ holds, the above limit is the same as

$$\lim_{n\to\infty} \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\right] .$$

Thus, we impose $\mathcal{R}(r_n, u_n)$ and switch freely between $\mathbb{E}$ and cov whenever suitable.

Uniform convergence of cluster measure

In Propositions 6.3.4 and 6.3.9 we proved (6.1.1) for $H \in \mathcal{A} \cup \mathcal{B}$. We note further that if (6.1.1) holds for H and $\widetilde{H}$, then it also holds for any linear combination of both functions. To deal with asymptotic normality, we need (6.1.1) to hold uniformly over a subclass of functions. With this in mind, we introduce two additional classes of functions. First, we recall that for a class $\mathcal{G}$ of functions $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ its envelope is

$$\mathbf{G}(\boldsymbol{x}) = \sup_{H \in \mathcal{G}} |H(\boldsymbol{x})| \;, \quad \boldsymbol{x} \in (\mathbb{R}^d)^{\mathbb{Z}} \;.$$

Definition 6.7.4. $\widetilde{\mathcal{A}} \subseteq \operatorname{span}(\mathcal{A})$ *(resp.* $\widetilde{\mathcal{B}} \subseteq \operatorname{span}(\mathcal{B})$*) is a class of functions with a finite envelope such that*

$$\lim_{n\to\infty} \sup_{H \in \widetilde{\mathcal{A}}} \boldsymbol{\nu}^*_{n,r_n}(|H|) < \infty \tag{6.7.5}$$

(resp. $\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{B}}} \boldsymbol{\nu}^*_{n,r_n}(|H|) < \infty$*) and that for each H there exist functions* $K_n^H : (\mathbb{R}^d)^{\ell_n} \to \mathbb{R}_+$ *such that*

$$\left| H\left(\frac{\boldsymbol{X}_{1,r_n}}{u_n}\right) - H\left(\frac{\boldsymbol{X}_{1,r_n-\ell_n}}{u_n}\right) \right| \le K_n^H(\boldsymbol{X}_{r_n-\ell_n+1,r_n}) \;, \quad \lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}} \frac{\mathbb{E}\left[K_n^H(\boldsymbol{X}_{1,\ell_n})\right]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0 \;. \tag{6.7.6}$$

Remark 6.7.5. The uniform convergence condition (6.7.5) strengthens the statement of Proposition 6.3.4. Conditions (6.7.5) and (6.7.6) are needed for asymptotic equicontinuity of empirical cluster process to be introduced below. ⊕

Remark 6.7.6. We note that

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[K_n^H(\boldsymbol{X}_{1,\ell_n})\right]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0$$

for each $H \in \widetilde{\mathcal{A}} \cup \widetilde{\mathcal{B}}$. Let us verify it for $H \in \mathcal{B}$. We have

$$\begin{aligned}
&|\mathbb{1}\{K(\boldsymbol{x}_{1,r_n}) > 1\} - \mathbb{1}\{K(\boldsymbol{x}_{1,r_n-\ell_n}) > 1\}| \\
&\qquad = \mathbb{1}\{K(\boldsymbol{x}_{1,r_n}) > 1\}\mathbb{1}\{K(\boldsymbol{x}_{1,r_n-\ell_n}) \le 1\} + \mathbb{1}\{K(\boldsymbol{x}_{1,r_n}) \le 1\}\mathbb{1}\{K(\boldsymbol{x}_{1,r_n-\ell_n}) > 1\} \;.
\end{aligned}$$

We consider the first pair of indicators in the last line. The events $\{K(\boldsymbol{x}_{1,r_n}) > 1\}$ and $\{K(\boldsymbol{x}_{1,r_n-\ell_n}) \le 1\}$ imply that there exists $s > 0$ such that $K(\boldsymbol{x}_{1,r_n}) - K(\boldsymbol{x}_{1,r_n-\ell_n}) > s$. Applying the same reasoning to the second pair of indicators, we have

$$|\mathbb{1}\{K(\boldsymbol{x}_{1,r_n}) > 1\} - \mathbb{1}\{K(\boldsymbol{x}_{1,r_n-\ell_n}) > 1\}| \le 2\mathbb{1}\left\{ \mathrm{cst} \sum_{j=r_n-\ell_n+1}^{r_n} |\boldsymbol{x}_j| > s \right\} .$$

Since $\ell_n = o(r_n)$

$$\mathbb{P}\left(\sum_{j=r_n-\ell_n+1}^{r_n} |\boldsymbol{X}_j| > s u_n \right) = O(\ell_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)) = (r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)) \;.$$

In summary, (6.7.6) holds if the envelope function is in $\widetilde{\mathcal{A}} \cup \widetilde{\mathcal{B}}$. ⊕

Remark 6.7.7. Let $\delta > 0$. If H is bounded then

$$\boldsymbol{\nu}^*_{n,r_n}(|H|^{1+\delta}) \le \|H\|_\infty^\delta \boldsymbol{\nu}^*_{n,r_n}(|H|)$$

and by the assumptions on the classes $\widetilde{\mathcal{A}}$ and $\widetilde{\mathcal{B}}$,

$$\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \boldsymbol{\nu}^*_{n,r_n}(|H|^{1+\delta}) < \infty \;.$$

⊕

Remark 6.7.8. Assume that $\mathcal{AC}(r_n,u_n)$ holds. Fix $0<s_0<t_0<\infty$. Let $H\in\mathcal{A}$ and recall that $H_s(\boldsymbol{x})=H(\boldsymbol{x}/s)$. Assume that $\widetilde{\mathcal{A}}:=\{H_s, s\in[s_0,t_0]\}$ is linearly ordered. Note that $\widetilde{\mathcal{A}}\subset\mathcal{A}$. The envelope is $|H_{s_0}|\vee|H_{t_0}|\in\widetilde{\mathcal{A}}$ hence (6.7.6) holds. Moreover, $\sup_{s\in[s_0,t_0]}\boldsymbol{\nu}^*_{n,r_n}(|H_s|)$ is achieved at s_0 or t_0. Likewise,

$$\lim_{n\to\infty}\sup_{s,t\in[s_0,t_0]}\boldsymbol{\nu}^*_{n,r_n}(|H_s-H_t|)<\infty\;.$$

The same applies to $H\in\mathcal{B}$ if additionally $\mathrm{ANSJB}(r_n,u_n)$ holds. ⊕

Conditional convergence

We consider conditional convergence of functions $H,\widetilde{H}$ acting on overlapping blocks.

Lemma 6.7.9. *Assume that $\mathcal{AC}(r_n,u_n)$ holds. Let $h<r_n$, $H,\widetilde{H}\in\mathcal{L}$ and $\widetilde{H}(\mathbf{0})=0$. Then*

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\mid|\boldsymbol{X}_0|>u_n]\\&=\begin{cases}\mathbb{E}[H(\boldsymbol{Y}_{1,\infty})\widetilde{H}(\boldsymbol{Y}_{1+h,\infty})]\;, & \text{if } h \text{ fixed}\;,\\ 0\;, & \text{if } h=h_n\to\infty\;.\end{cases}\end{aligned}$$

and

$$\lim_{n\to\infty}\mathbb{E}[H(\boldsymbol{X}_{-r_n,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{-r_n+h,r_n+h}/u_n)\mid|\boldsymbol{X}_0|>u_n]=\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})]\;.$$

Proof. Since $H,\widetilde{H}$ are bounded, the first expectation of interest is dominated by

$$\|H\|_\infty\|\widetilde{H}\|_\infty\mathbb{P}(\boldsymbol{X}^*_{1+h,r_n+h}>u_n\mid|\boldsymbol{X}_0|>u_n)\;.$$

Thus, the statement for $h=h_n\to\infty$ follows immediately from $\mathcal{AC}(r_n,u_n)$ (cf. the argument in the proof of [KS20, Theorem 6.1.4]).

Now, let h be fixed. Fix r. Since $H,\widetilde{H}$ are bounded Lipschitz continuous, we have by Proposition 6.3.2,

$$\begin{aligned}&\lim_{n\to\infty}\mathbb{E}[H(\boldsymbol{X}_{1,r}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r+h}/u_n)\mid|\boldsymbol{X}_0|>u_n]=\mathbb{E}[H(\boldsymbol{Y}_{1,r})\widetilde{H}(\boldsymbol{Y}_{1+h,r+h})]\;,\\&\lim_{n\to\infty}\mathbb{E}[H(\boldsymbol{X}_{-r,r}/u_n)\widetilde{H}(\boldsymbol{X}_{-r+h,r+h}/u_n)\mid|\boldsymbol{X}_0|>u_n]=\mathbb{E}[H(\boldsymbol{Y}_{-r,r})\widetilde{H}(\boldsymbol{Y}_{-r+h,r+h})]\;.\end{aligned}$$

Since the tail process tends to zero under condition $\mathcal{AC}(r_n,u_n)$, it also holds that

$$\begin{aligned}&\lim_{r\to\infty}\mathbb{E}[H(\boldsymbol{Y}_{1,r})\widetilde{H}(\boldsymbol{Y}_{1+h,r+h})]=\mathbb{E}[H(\boldsymbol{Y}_{1,\infty})\widetilde{H}(\boldsymbol{Y}_{1+h,\infty})]\;,\\&\lim_{r\to\infty}\mathbb{E}[H(\boldsymbol{Y}_{-r,r})\widetilde{H}(\boldsymbol{Y}_{-r+h,r+h})]=\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})]=\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})]\;.\end{aligned}$$

Indeed, considering the first statement only we have

$$\begin{aligned}&\lim_{r\to\infty}\left|\mathbb{E}[H(\boldsymbol{Y}_{1,r})\widetilde{H}(\boldsymbol{Y}_{1+h,r+h})]-\mathbb{E}[H(\boldsymbol{Y}_{1,\infty})\widetilde{H}(\boldsymbol{Y}_{1+h,\infty})]\right|\\&\le\lim_{r\to\infty}\mathbb{E}[|H(\boldsymbol{Y}_{1,r})-H(\boldsymbol{Y}_{1,\infty})|\,|\widetilde{H}|(\boldsymbol{Y}_{1+h,r+h})]\\&\quad+\lim_{r\to\infty}\mathbb{E}\left[|H|(\boldsymbol{Y}_{1,\infty})\left|\widetilde{H}(\boldsymbol{Y}_{1+h,r+h})-\widetilde{H}(\boldsymbol{Y}_{1+h,\infty})\right|\right]\\&\le\|\widetilde{H}\|_\infty\lim_{r\to\infty}\left\{\mathbb{E}\left[|H(\boldsymbol{Y}_{1,r})-H(\boldsymbol{Y}_{1,\infty})|\right]+\mathbb{E}\left[\left|\widetilde{H}(\boldsymbol{Y}_{1+h,r+h})-\widetilde{H}(\boldsymbol{Y}_{1+h,\infty})\right|\right]\right\}=0\;.\end{aligned}$$

To conclude, we only need to apply the triangular argument, that is to prove that

$$\lim_{r\to\infty}\limsup_{n\to\infty}$$
$$\left|\mathbb{E}[H(\boldsymbol{X}_{1,r}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r+h}/u_n) - H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n) \mid |\boldsymbol{X}_0| > u_n]\right| = 0 .$$

Using again the fact that $H, \widetilde{H}$ are bounded, the conditional expectation is dominated by

$$\begin{aligned}&\|\widetilde{H}\|_\infty \left|\mathbb{E}[H(\boldsymbol{X}_{1,r}/u_n) - H(\boldsymbol{X}_{1,r_n}/u_n) \mid |\boldsymbol{X}_0| > u_n]\right| \\ &+ \|H\|_\infty \left|\mathbb{E}[\widetilde{H}(\boldsymbol{X}_{1+h,r+h}/u_n) - \widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n) \mid |\boldsymbol{X}_0| > u_n]\right| .\end{aligned} \tag{6.7.7}$$

Fix $\epsilon > 0$. Since H is Lipschitz continuous, applying condition $\mathcal{AC}(r_n, u_n)$ yields

$$\begin{aligned}&\lim_{r\to\infty}\limsup_{n\to\infty} \left|\mathbb{E}[H(\boldsymbol{X}_{1,r}/u_n) - H(\boldsymbol{X}_{1,r_n}/u_n) \mid |\boldsymbol{X}_0| > u_n]\right| \\ &\qquad \le \mathrm{cst}\left\{\epsilon + \lim_{r\to\infty}\limsup_{n\to\infty}\mathbb{P}(\boldsymbol{X}^*_{r,r_n} > \epsilon u_n \mid |\boldsymbol{X}_0| > u_n)\right\} = \mathrm{cst}\times\epsilon .\end{aligned}$$

The same argument applies to (6.7.7). Since ϵ is arbitrary, this concludes the proof. □

Covariance of clusters: Disjoint blocks

The first result is straightforward under the beta-mixing conditions.

Lemma 6.7.10 (Disjoint blocks I)**.** *Assume that $\mathcal{AC}(r_n, u_n)$, $\mathcal{R}(r_n, u_n)$, (6.7.1b) hold. Then*

$$\lim_{n\to\infty}\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sup_{\xi'>1}\sup_{H,\widetilde{H}\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+[\xi' r_n],r_n+[\xi' r_n]}/u_n)\right] = 0 .$$

Proof of Lemma 6.7.10. Let $H, \widetilde{H} \in \widetilde{\mathcal{A}} \cup \widetilde{\mathcal{B}}$. Then, using (6.7.6),

$$\begin{aligned}&\left|\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+[\xi' r_n],r_n+[\xi' r_n]}/u_n)\right]\right| \\ &\le \left|\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n-\ell_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+[\xi' r_n],r_n+[\xi' r_n]}/u_n)\right]\right| + \|\widetilde{H}\|_\infty\mathbb{E}\left[K_n^H(\boldsymbol{X}_{1,\ell_n})\right]\end{aligned}$$

and the latter term is $o(r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n))$, uniformly over $\widetilde{\mathcal{A}} \cup \widetilde{\mathcal{B}}$ by the assumption.
Using (6.7.2) and (6.7.1b), we have

$$\frac{\left|\mathrm{cov}\left(H(\boldsymbol{X}_{1,r_n-\ell_n}/u_n), \widetilde{H}(\boldsymbol{X}_{1+[\xi' r_n],r_n+[\xi' r_n]}/u_n)\right)\right|}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \le \|H\|_\infty\|\widetilde{H}\|_\infty\frac{\beta_{\ell_n+[(\xi'-1)r_n]}}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}$$

and the latter is $o(1)$ uniformly over the class of functions. □

We extend the above result to the excess functional $\mathcal{E}_s(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}}\mathbb{1}\{|\boldsymbol{x}_j| > s\}$.

Lemma 6.7.11 (Disjoint blocks II)**.** *Assume that $\mathcal{AC}(r_n, u_n)$, $\mathcal{R}(r_n, u_n)$, (6.7.1b) hold. Then*

$$\lim_{n\to\infty}\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sup_{\xi'>1}\sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}}\sup_{s\in[s_0,t_0]}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\mathcal{E}_s(\boldsymbol{X}_{1+[\xi' r_n],r_n+[\xi' r_n]}/u_n)\right] = 0 .$$

Proof. Recall that $\ell_n = o(r_n)$. Split the sum $\sum_{j=[\xi' r_n]+1}^{r_n+[\xi' r_n]}$ into $\sum_{j=[\xi' r_n]+1}^{r_n+\ell_n}$ and $\sum_{j=r_n+\ell_n+1}^{r_n+[\xi' r_n]}$.

For the first sum we have

$$\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=[\xi' r_n]+1}^{r_n+\ell_n}\left|\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)\mathbb{1}\{|\boldsymbol{X}_j|>u_n s\}]\right|\leq\|H\|_\infty\frac{\ell_n\mathbb{P}(|\boldsymbol{X}_0|>u_n s)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=o(1)$$

uniformly over the class of functions and over s.

Using (6.7.2) we have

$$\begin{aligned}&\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=r_n+\ell_n+1}^{r_n+[\xi' r_n]}\left|\operatorname{cov}(H(\boldsymbol{X}_{1,r_n}/u_n),\mathbb{1}\{|\boldsymbol{X}_j|>u_n s\})\right|\\&\leq\frac{\|H\|_\infty}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=r_n+\ell_n+1}^{\infty}\beta_{j-r_n}\ .\end{aligned}$$

We finish the proof using the mixing assumption (6.7.1b). □

Covariance of clusters: Overlapping blocks

We consider three cases separately: a) $H,\widetilde{H}\in\mathcal{A}$ (Proposition 6.7.12); b) $H,\widetilde{H}\in\mathcal{B}$ (Proposition 6.7.13); c) the excess functional (Proposition 6.7.14).

Proposition 6.7.12 (Overlapping blocks I)**.** *Assume that $\mathcal{AC}(r_n,u_n)$ and $\mathcal{R}(r_n,u_n)$ hold. Let $h<r_n$ and $\xi\in(0,1)$. For $H,\widetilde{H}\in\mathcal{A}$ we have*

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\right]\\&=\begin{cases}\boldsymbol{\nu}^*(H\widetilde{H})\ , & \text{if } h/r_n\to 0\ ,\\(1-\xi)\boldsymbol{\nu}^*(H\widetilde{H})\ , & \text{if } h=h_n=[\xi r_n]\end{cases}\ .\end{aligned}\tag{6.7.8}$$

Proof of Proposition 6.7.12. Note that if $H,\widetilde{H}\in\mathcal{A}$, then $H\widetilde{H}\in\mathcal{A}$ (but it does not mean that we can apply Lemma 6.7.9 since here the functions are applied to different blocks).

Since $H,\widetilde{H}$ vanish around $\mathbf{0}$, there exists $\epsilon>0$ such that $H(\boldsymbol{x}_{1,r_n})\widetilde{H}(\boldsymbol{x}_{1,r_n})=0$ whenever $\boldsymbol{x}^*_{1,r_n}<\epsilon$. Then, splitting the event $\{\boldsymbol{X}^*_{1,r_n}>u_n\}$ and using stationarity we write the expression of interest as

$$\begin{aligned}&\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\mathbb{1}\{\boldsymbol{X}^*_{1,r_n}>u_n\epsilon\}\right]\\&=\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{r_n}\mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h,r_n+h}/u_n)\mathbb{1}\{\boldsymbol{X}^*_{1,j-1}\leq u_n\epsilon\}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\epsilon\}\right]\\&=\frac{1}{r_n}\frac{\mathbb{P}(|\boldsymbol{X}_0|>\epsilon u_n)}{\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\\&\quad\times\sum_{j=1}^{r_n}\mathbb{E}\left[H(\boldsymbol{X}_{1-j,r_n-j}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h-j,r_n+h-j}/u_n)\mathbb{1}\{\boldsymbol{X}^*_{1-j,-1}\leq u_n\epsilon\}\mid|\boldsymbol{X}_0|>u_n\epsilon\right]\ .\end{aligned}$$

We write the last expression as $\int_0^1 g_n(v)\mathrm{d}v$ with

$$g_n(v)=\mathbb{E}[H(\boldsymbol{X}_{1-[r_nv],r_n-[r_nv]}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h-[r_nv],r_n+h-[r_nv]}/u_n)\mathbb{1}\left\{\boldsymbol{X}^*_{1-[r_nv],-1}\leq u_n\epsilon\right\}\mid|\boldsymbol{X}_0|>u_n\epsilon]\ .$$

If $h=o(r_n)$, then using the second part of Lemma 6.7.9 we get

$$\lim_{n\to\infty}g_n(v)=\epsilon^{-\alpha}\mathbb{E}[H(\epsilon\boldsymbol{Y})\widetilde{H}(\epsilon\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1}\leq 1\}]=\boldsymbol{\nu}^*(H\widetilde{H})$$

independently of $v \in (0,1)$. If $h = [\xi r_n]$, $\xi \in (0,1)$, then we split the integral.

If $\xi > v$, then we use boundedness of both $H, \widetilde{H}$ and the fact that $\widetilde{H}$ vanishes around $\mathbf{0}$. Thanks to the anticlustering condition $\mathcal{AC}(r_n, u_n)$, we have as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{E}[H(\boldsymbol{X}_{1-[r_n v], r_n - [r_n v]}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h-[r_n v], r_n+h-[r_n v]}/u_n)\mathbb{1}\left\{\boldsymbol{X}^*_{1-[r_n v],-1} \le u_n\epsilon\right\} \mid |\boldsymbol{X}_0| > u_n\epsilon] \\
&\le \mathrm{cst}\mathbb{P}\left(\boldsymbol{X}^*_{1+[\xi r_n]-[r_n v], r_n+[\xi r_n]-[r_n v]} > u_n\epsilon \mid |\boldsymbol{X}_0| > u_n\epsilon\right) \\
&\le \mathrm{cst}\mathbb{P}\left(\boldsymbol{X}^*_{[r_n(\xi - v)], 3r_n} > u_n\epsilon \mid |\boldsymbol{X}_0| > u_n\epsilon\right) \to 0 \; .
\end{aligned}
$$

If $\xi \le v$, then we apply the second part of Lemma 6.7.9:

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}[H(\boldsymbol{X}_{1-[r_n v], r_n - [r_n v]}/u_n)\widetilde{H}(\boldsymbol{X}_{1+h-[r_n v], r_n+h-[r_n v]}/u_n)\mathbb{1}\left\{\boldsymbol{X}^*_{1-[r_n v],-1} \le u_n\epsilon\right\} \mid |\boldsymbol{X}_0| > u_n\epsilon] \\
&= \epsilon^{-\alpha}\mathbb{E}[H(\epsilon\boldsymbol{Y})\widetilde{H}(\epsilon\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] = \boldsymbol{\nu}^*(H\widetilde{H}) \; .
\end{aligned}
$$

Since the sequence $\{g_n\}$ is uniformly bounded, we have

$$
\lim_{n\to\infty} \int_0^\xi g_n(v)\mathrm{d}v + \lim_{n\to\infty} \int_\xi^1 g_n(v)\mathrm{d}v = 0 + (1-\xi)\boldsymbol{\nu}^*(H\widetilde{H}) \; .
$$

□

Proposition 6.7.13 (Overlapping blocks II)**.** *Assume that* $\mathcal{AC}(r_n, u_n)$, $\mathrm{ANSJB}(r_n, u_n)$ *and* $\mathcal{R}(r_n, u_n)$ *hold. Let* $h < r_n$ *and* $\xi \in (0,1)$*. Then* (6.7.8) *holds for* $H, \widetilde{H} \in \mathcal{B}$*.*

Proof of Proposition 6.7.13. We mimic the proof of Proposition 6.3.9 (refer to that proof for the notation). Set $K^\epsilon = K \circ T^\epsilon$, $\widetilde{K}^\epsilon = \widetilde{K} \circ T^\epsilon$. Let $\eta \in (0,1)$. Note that $H^\epsilon_\pm := \mathbb{1}\{K^\epsilon > 1 \pm \eta\} \in \mathcal{A}$, $\widetilde{H}^\epsilon_\pm := \mathbb{1}\left\{\widetilde{K}^\epsilon > 1 \pm \eta\right\} \in \mathcal{A}$ and hence $H^\epsilon\widetilde{H}^\epsilon \in \mathcal{A}$; see the comment at the beginning of the proof of Proposition 6.7.12.

Fix $\eta \in (0,1)$ and $\zeta > 0$. Let $L_K, L_{\widetilde{K}}$ be as in (6.2.1) and choose $\epsilon > 0$ such that

$$
\limsup_{n\to\infty} \frac{2\mathbb{P}(\sum_{j=1}^{r_n} |\boldsymbol{X}_j| \, \mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n/(L_K \vee L_{\widetilde{K}}))}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \le \zeta \; .
$$

This is allowed thanks to $\mathrm{ANSJB}(r_n, u_n)$. We have

$$
\begin{aligned}
&\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\mathbb{P}\left(K(\boldsymbol{X}_{1,r_n}/u_n) > 1, \widetilde{K}(\boldsymbol{X}_{1+h,r_n+h}/u_n) > 1\right) \\
&\le \frac{2\mathbb{P}(\sum_{i=1}^{r_n} |\boldsymbol{X}_j| \, \mathbb{1}\{|\boldsymbol{X}_j| \le \epsilon u_n\} > \eta u_n/\mathrm{cst})}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} + \frac{\mathbb{E}[H^\epsilon_-(\boldsymbol{X}_{1,r_n}/u_n)\widetilde{H}^\epsilon_-(\boldsymbol{X}_{1+h,r_n+h}/u_n)]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \; .
\end{aligned}
$$

Application of Proposition 6.7.12 gives

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\mathbb{P}\left(K(\boldsymbol{X}_{1,r_n}/u_n) > 1, \widetilde{K}(\boldsymbol{X}_{1+h,r_n+h}/u_n) > 1\right) \\
&\le \zeta + \left\{ \begin{array}{ll} \boldsymbol{\nu}^*(H^\epsilon_-(\boldsymbol{Y})\widetilde{H}^\epsilon_-(\boldsymbol{Y})) \; , & \text{if } h/r_n \to 0 \; , \\ (1-\xi)\boldsymbol{\nu}^*(H^\epsilon_-(\boldsymbol{Y})\widetilde{H}^\epsilon_-(\boldsymbol{Y})) \; , & \text{if } h = h_n = [\xi r_n] \end{array} \right. \; .
\end{aligned}
$$

Similarly, we obtain the lower bound with $H^\epsilon_+, \widetilde{H}^\epsilon_+$ instead of $H^\epsilon_-, \widetilde{H}^\epsilon_-$ and $-\zeta$ instead of $+\zeta$. Since ζ is arbitrary, the proof is concluded by letting $\epsilon \to 0$. This follows the same argument as in the proof of Proposition 6.3.9. □

Proposition 6.7.14 (Overlapping blocks III)**.** *Assume that $\mathcal{AC}(r_n, u_n)$ and $\mathcal{R}(r_n, u_n)$ hold. Let $h < r_n$. For $H \in \mathcal{L}$ we have*

$$\lim_{n\to\infty} \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\mathcal{E}_s(\boldsymbol{X}_{1+h,r_n+h}/u_n)\right]$$
$$= \begin{cases} s^{-\alpha}\mathbb{E}[H(s\boldsymbol{Y})] \,, & \text{if } h/r_n \to 0 \,, \\ s^{-\alpha}(1-\xi)\mathbb{E}[H(s\boldsymbol{Y})] \,, & \text{if } h = h_n = \xi r_n \end{cases} .$$

Proof of Proposition 6.7.14. We have for $h < r_n$,

$$\frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)\mathcal{E}_s(\boldsymbol{X}_{1+h,r_n+h}/u_n)\right]$$
$$= \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=h+1}^{r_n+h} \mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)\mathbb{1}\{|\boldsymbol{X}_j| > u_n s\}]$$
$$= \frac{1}{r_n} \frac{\mathbb{P}(|\boldsymbol{X}_0| > u_n s)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=h+1}^{r_n+h} \mathbb{E}[H(\boldsymbol{X}_{1-j,r_n-j}/u_n) \mid |\boldsymbol{X}_0| > u_n s] \,.$$

We write the last expression as

$$\frac{\mathbb{P}(|\boldsymbol{X}_0| > u_n s)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \int_{h/r_n}^{1+h/r_n} g_n(v)\mathrm{d}v$$

with (omitting the dependence on s)

$$g_n(v) = \mathbb{E}[H(s\boldsymbol{X}_{1-[r_n v],r_n-[r_n v]}/(u_n s)) \mid |\boldsymbol{X}_0| > u_n s] \,.$$

Since H is bounded, $\mathcal{AC}(r_n, u_n)$ and Proposition 6.3.2 give

$$\lim_{n\to\infty} g_n(v) = \begin{cases} \mathbb{E}[H(s\boldsymbol{Y})] & \text{if } v \in (0,1) \,, \\ 0 & \text{if } v > 1 \,. \end{cases}$$

We split

$$\int_{h/r_n}^{1+h/r_n} g_n(v)\mathrm{d}v = \int_{h/r_n}^{1} g_n(v)\mathrm{d}v + \int_{1}^{1+h/r_n} g_n(v)\mathrm{d}v \,.$$

Since the sequence $\{g_n\}$ is uniformly bounded, for any $h < r_n$ the second integral above converges to zero as $n \to \infty$. If $h = o(r_n)$ and since there is no problem at $v = 0$ with $g_n(v)$, then the first integral converges to

$$\int_0^1 \mathbb{E}[H(s\boldsymbol{Y})]\mathrm{d}v = \mathbb{E}[H(s\boldsymbol{Y})] \,.$$

Likewise, when $h = [\xi r_n]$ then the first integral converges to

$$\int_\xi^1 \mathbb{E}[H(s\boldsymbol{Y})]\mathrm{d}v = (1-\xi)\mathbb{E}[H(s\boldsymbol{Y})] \,.$$

□

6.7.5. Empirical cluster process of sliding blocks

Recall that for $s > 0$, $H_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)$. In order to deal with asymptotic normality of sliding blocks estimators, we study the empirical process

$$\mathbb{F}_n(H_s) := \sqrt{k_n}\left\{\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H_s) - \boldsymbol{\nu}^*(H_s)\right\} = \sqrt{k_n}\left\{\frac{\sum_{i=0}^{q_n-1} H_s\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} - s^{-\alpha}\boldsymbol{\nu}^*(H)\right\} .$$

The process $\mathbb{F}_n(H_s)$ is viewed as a random element with values in $\mathbb{D}([s_0, t_0])$.

Theorem 6.7.15. *Let $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n, u_n)$, $\beta(r_n)$ and $\mathcal{AC}(r_n, u_n)$ hold. Let $H \in \mathcal{A}$ be such that the class $\{H_s : s \in [s_0, t_0]\}$ is linearly ordered and (6.4.2b) holds.*

Then $\mathbb{F}_n(H_\cdot)$ converges weakly in $(\mathbb{D}([s_0, t_0]), J_1)$ to a Gaussian process with the covariance $\boldsymbol{\nu}^*(H_s H_t)$.

If moreover ANSJB(r_n, u_n) is satisfied, then the convergence holds for $H \in \mathcal{B}$.

If additionally $\mathcal{S}(r_n, u_n)$ and (6.4.2a) are satisfied, then the processes $\mathbb{F}_n(H_\cdot)$ and $\mathbb{F}_n(\mathcal{E}_\cdot)$ converge jointly.

Tail empirical process

Consider the following tail empirical process:

$$\widetilde{\mathbb{T}}_n(s) = \sqrt{k_n}\left\{T_n(s) - s^{-\alpha}\right\} = \sqrt{k_n}\left\{\frac{\sum_{j=1}^{q_n} \mathbb{1}\{|\boldsymbol{X}_j| > u_n s\}}{q_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} - s^{-\alpha}\right\}\ ,\ s > 0\ .$$

Note that this is the classical tail empirical process based on the random variables $|\boldsymbol{X}_j|$, $j \geq 1$, with the only one difference: q_n replaces n. We argue that this process can be obtained (approximately) as the empirical process of sliding blocks. Indeed,

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(\mathcal{E}_s) = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)}\left\{\sum_{j=1}^{r_n} j + r_n \sum_{j=r_n+1}^{q_n} + \sum_{j=q_n+1}^{n}(n-j)\right\}\mathbb{1}\{|\boldsymbol{X}_j| > u_n s\}\ .$$

The difference between $\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(\mathcal{E}_s)$ and $T_n(s)$ is

$$A := \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)}\left\{\sum_{j=1}^{r_n}(r_n - j) - \sum_{j=q_n+1}^{n}(n-j)\right\}\mathbb{1}\{|\boldsymbol{X}_j| > u_n s\}\ .$$

We have $\sum_{j=1}^{r_n}(r_n - j) \leq r_n^2$ and $\sum_{j=q_n+1}^{n}(n-j) \leq r_n^2$, thus under $\mathcal{R}(r_n, u_n)$:

$$\lim_{n\to\infty} \sqrt{k_n}\mathbb{E}[|A|] \leq \mathrm{cst} \lim_{n\to\infty} \sqrt{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\frac{r_n}{q_n} = \mathrm{cst} \lim_{n\to\infty} \sqrt{\frac{r_n}{n}}\sqrt{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} = 0\ .$$

This implies that $\mathbb{F}_n(\mathcal{E}_s)$ and $\widetilde{\mathbb{T}}_n(s)$ are asymptotically equivalent in the sense that they yield the same process $\mathbb{F}(\mathcal{E}_s)$ as the distributional limit.

6.7.6. Covariance of the empirical process of sliding blocks

Proposition 6.7.16. *Assume that $\mathcal{AC}(r_n, u_n)$ and $\mathcal{R}(r_n, u_n)$ are satisfied. Let*

- *$H, \widetilde{H} \in \widetilde{\mathcal{A}}$, or*
- *$H, \widetilde{H} \in \widetilde{\mathcal{B}}$ and ANSJB(r_n, u_n) holds.*

If (6.7.1b) and (6.7.1c) hold then

$$\lim_{n\to\infty} \mathrm{cov}(\mathbb{F}_n(H), \mathbb{F}_n(\widetilde{H})) = \boldsymbol{\nu}^*(H\widetilde{H})\ . \tag{6.7.9}$$

If (6.7.1b) holds then

$$\lim_{n\to\infty} \mathrm{cov}(\mathbb{F}_n(H), \mathbb{F}_n(\mathcal{E}_s)) = \boldsymbol{\nu}^*(H\mathcal{E}_s) = \mathbb{E}[H(s\boldsymbol{Y})]\ . \tag{6.7.10}$$

Remark 6.7.17.

- The second equality in (6.7.10) follows from Lemma 6.7.22.
- In view of the discussion in Section 6.7.5, (6.7.10) can be re-phrased as

$$\lim_{n\to\infty} \operatorname{cov}(\mathbb{F}_n(H), \widetilde{\mathbb{T}}_n(s)) = \boldsymbol{\nu}^*(H\mathcal{E}_s) = \mathbb{E}[H(s\boldsymbol{Y})] \; .$$

⊕

Bounds for integral representation

Before we proceed with the proof, we define

$$g_n(\xi;H) = \mathbb{E}\left[H(\boldsymbol{X}_{1,r_n}/u_n)H(\boldsymbol{X}_{1+[r_n\xi],[r_n\xi]+r_n}/u_n)\right] \; , \quad \xi > 0$$

and

$$\widetilde{g}_n(\xi;H) = \frac{g_n(\xi;H)}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \; .$$

For $\xi = 0$, using Remark 6.7.7 we immediately obtain under $\mathcal{AC}(r_n, u_n)$:

$$\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \widetilde{g}_n(0;H) = \lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \boldsymbol{\nu}^*_{n,r_n}(H^2) < \infty \; . \tag{6.7.11}$$

Furthermore, for $j = 1, 2, 3, \ldots$,

$$\frac{1}{r_n} \sum_{i=(j-1)r_n}^{jr_n-1} \widetilde{g}_n(i/r_n;H) = \int_{j-1}^{j} \widetilde{g}_n(\xi;H)\mathrm{d}\xi \; .$$

For $j = 1$ we will need the precise behaviour of this integral and we will handle it using Propositions 6.7.12 and 6.7.13. For $j \geq 2$ the integral vanishes with a given rate.

Lemma 6.7.18. *Assume that $\mathcal{AC}(r_n, u_n)$ holds.*

- *If (6.7.1b) holds then for any finite M,*

$$\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \int_1^M \widetilde{g}_n(\xi;H)\mathrm{d}\xi = 0$$

- *For $j \geq 3$,*

$$\sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \int_{j-1}^{j} \widetilde{g}_n(\xi;H)\mathrm{d}\xi \leq \mathrm{cst} \; \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \beta_{(j-2)r_n} \; .$$

Proof. For the first part we apply Lemma 6.7.10 and the dominated convergence:

$$\sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \sup_{\xi\in(1,2)} |\widetilde{g}_n(\xi;H)| \leq \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \|H\|_\infty \boldsymbol{\nu}^*_{n,r_n}(|H|) \leq \mathrm{cst} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} \boldsymbol{\nu}^*_{n,r_n}(|H|) < \infty \; .$$

For the second part, we use (6.7.2) and the fact that $\widetilde{\mathcal{A}} \cup \widetilde{\mathcal{B}}$ has a finite envelope. □

Representation for covariance between blocks

Recall that $q_n = n - r_n + 1$. Evaluation of the covariance of the empirical process of sliding blocks will use consecutive disjoint blocks of indices of size r_n:

$$J_j = \{(j-1)r_n, \ldots, jr_n - 1\} \; , \quad j = 1, \ldots, m_n = [q_n/r_n] \; .$$

Clearly, $\bigcup_{j=1}^{m_n} J_j = \{0,\dots,n-r_n\}$. We will assume for simplicity that q_n/r_n is an integer. Write

$$\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) = \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)} \sum_{j=1}^{m_n} \Psi_j(H)$$

with

$$\Psi_j(H) = \sum_{i\in J_j} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \; . \tag{6.7.12}$$

Note that the indices of the random vectors $\boldsymbol{X}_1,\dots,\boldsymbol{X}_{2r_n-1}$ used in the construction of Ψ_1 overlap with the indices of $\boldsymbol{X}_{r_n+1},\dots,\boldsymbol{X}_{3r_n-1}$ used to define Ψ_2, but do not overlap with the indices used in the definition of Ψ_3. Likewise, the indices used in the definition of Ψ_2 overlap with those in Ψ_3, but not with any other term Ψ_j, $j\geq 4$. This partially explains where does a contribution to the limiting variance come from: from the dependence within each block J_j and cross dependence between J_j and two neighbouring blocks.

For $j\geq 1$ we have

$$\begin{aligned}
&\frac{\mathbb{E}[\Psi_1(H)\Psi_{j+1}(H)]}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)} = \frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[\sum_{h=0}^{r_n-1} H\left(\boldsymbol{X}_{h+1,h+r_n}/u_n\right) \sum_{i=jr_n}^{(j+1)r_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)\right] \\
&= \frac{1}{r_n}\sum_{i=(j-1)r_n}^{jr_n}\left(\frac{i}{r_n}-(j-1)\right)\widetilde{g}_n(i/r_n;H) + \frac{1}{r_n}\sum_{i=jr_n+1}^{(j+1)r_n}\left((j+1)-\frac{i}{r_n}\right)\widetilde{g}_n(i/r_n;H)
\end{aligned}$$

and hence

$$\frac{\mathbb{E}[|\Psi_1(H)\Psi_{j+1}(H)|]}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \leq \frac{1}{r_n}\sum_{i=(j-1)r_n}^{(j+1)r_n}|\widetilde{g}_n(i/r_n;H)| \leq \int_{j-1}^{j+1}|\widetilde{g}_n(\xi;H)|\mathrm{d}\xi \; . \tag{6.7.13}$$

Proof of Proposition 6.7.16, Eq. (6.7.9)

Proof. Note that (since $q_n\sim n$)

$$\frac{k_n m_n}{\left(q_n r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)\right)^2} = \frac{q_n n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}{r_n\left(q_n r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)\right)^2} \sim \frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)} \; . \tag{6.7.14}$$

Write $\mathrm{var}(\mathbb{F}_n(H))$ as

$$\frac{k_n m_n}{\left(q_n r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)\right)^2}\mathrm{cov}(\Psi_2(H),\Psi_1(H)+\Psi_2(H)+\Psi_3(H)) + A_n(H) \tag{6.7.15}$$

with the reminder $A_n(H)$ given by

$$\begin{aligned}
A_n(H) :=& -2\frac{k_n}{\left(q_n r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)\right)^2}\mathrm{cov}(\Psi_1(H),\Psi_2(H)) && (6.7.16)\\
&+2\frac{1+o(1)}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=2}^{m_n-1}\left(1-\frac{j}{m_n}\right)\left\{\mathrm{cov}\left(\Psi_1(H),\Psi_{1+j}(H)\right)\right\} && (6.7.17)\\
&=: A_{n,1}(H) + (1+o(1))B_n(H) \; .
\end{aligned}$$

If we show that the leading term on the right-hand side of (6.7.15) converges to a finite limit, then automatically $\lim_{n\to\infty} A_{n,1}(H) = 0$ (since $m_n \to \infty$). Thus, the reminder $A_n(H)$ will be negligible if we show that

$$\lim_{n\to\infty} B_n(H) = 0 \,. \tag{6.7.18}$$

We will start by analysing the first term in (6.7.15). Set

$$R_n(H) = \frac{1}{r_n} \sum_{i=r_n}^{2r_n-1} (2 - i/r_n)\widetilde{g}_n(i/r_n; H) \,.$$

Since (6.7.1b) holds, the application of the first part of Lemma 6.7.18 gives

$$\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} R_n(H) = 0 \,. \tag{6.7.19}$$

Write the first term in (6.7.15) as (cf. (6.7.14))

$$\begin{aligned}
&\frac{1+o(1)}{r_n^3\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\left\{r_n g_n(0;H) + 2r_n \sum_{i=1}^{r_n-1} g_n(i/r_n;H) + 2r_n R_n\right\} \\
&= (1+o(1))\left\{\frac{1}{r_n}\widetilde{g}_n(0;H) + 2\frac{1}{r_n}\sum_{i=1}^{r_n-1}\widetilde{g}_n(i/r_n;H) + 2R_n(H)\right\} \,.
\end{aligned} \tag{6.7.20}$$

Then, using (6.7.11), (6.7.18) and (6.7.19), we have

$$\lim_{n\to\infty} \mathrm{var}(\mathbb{F}_n(H)) = 2\lim_{n\to\infty}\int_0^1 \widetilde{g}_n(\xi;H)\mathrm{d}\xi \,,$$

Applying Propositions 6.7.12 and 6.7.13 (the case $h = [\xi r_n]$), we have

$$\lim_{n\to\infty} \mathrm{var}(\mathbb{F}_n(H)) = 2\boldsymbol{\nu}^*(H^2)\int_0^1 (1-\xi)\mathrm{d}\xi = \boldsymbol{\nu}^*(H^2) \,.$$

To conclude the proof, we show (6.7.18) in the following lemma. □

Lemma 6.7.19. *Assume that (6.7.1b)-(6.7.1c) hold. Then*

$$\lim_{n\to\infty} \sup_{H\in\widetilde{\mathcal{A}}\cup\widetilde{\mathcal{B}}} B_n(H) = 0 \,. \tag{6.7.21}$$

Proof of Lemma 6.7.19. Using (6.7.13) we have

$$\begin{aligned}
|B_n(H)| &\le \frac{\mathbb{E}[|\Psi_1(H)\Psi_3(H)|]}{r_n^3\mathbb{P}(|\boldsymbol{X}_0| > u_n)} + \frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sum_{j=3}^{m_n-1}|\mathbb{E}[|\Psi_1(H)\Psi_{1+j}(H)|] \\
&\le \int_1^3 |\widetilde{g}_n(\xi;H)|\mathrm{d}\xi + \sum_{j=3}^{m_n-1}\int_{j-1}^{j+1}|\widetilde{g}_n(\xi;H)|\mathrm{d}\xi \,.
\end{aligned}$$

The first term is $o(1)$ uniformly over the class of functions (cf. the first part of Lemma 6.7.18). Using the second part of Lemma 6.7.18 we bound

$$\sum_{j=3}^{m_n-1}\int_{j-1}^{j+1}|\widetilde{g}_n(\xi;H)|\mathrm{d}\xi \le \mathrm{cst}\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sum_{j=1}^{\infty}\beta_{jr_n} \,.$$

We finish the proof by applying the mixing assumption (6.7.1c). □

Proof of Proposition 6.7.16, Eq. (6.7.10)

Proof. We write (recall that $q_n \sim n$)

$$
\begin{aligned}
&k_n \mathrm{cov}\left(\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1} H(\boldsymbol{X}_{i+1,i+r_n}/u_n), \frac{1}{q_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{q_n} \mathbb{1}\{|\boldsymbol{X}_j|>u_n\}\right)\\
&\sim \frac{k_n}{n^2 r_n \mathbb{P}^2(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1}\sum_{j=1}^{q_n} \mathrm{cov}\left(H(\boldsymbol{X}_{i+1,i+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_j|>u_n\}\right)\\
&= \frac{1}{n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1}\sum_{j=1}^{q_n} \mathrm{cov}\left(H(\boldsymbol{X}_{i-j+1,i-j+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right) .
\end{aligned}
$$

Split the inner sum into two pieces, $\sum_{j=1}^{i}$ and $\sum_{j=i+1}^{q_n}$, in the first one replace j with $h=i-j$, in the second one replace j with $h=j-i$ to get

$$
\begin{aligned}
&\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1}\sum_{j=1}^{i} \mathrm{cov}\left(H(\boldsymbol{X}_{i-j+1,i-j+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right)\\
&\quad+\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1}\sum_{j=i+1}^{q_n} \mathrm{cov}\left(H(\boldsymbol{X}_{i-j+1,i-j+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right)\\
&=\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=1}^{q_n-1}\sum_{h=0}^{i-1} \mathrm{cov}\left(H(\boldsymbol{X}_{h+1,h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right)\\
&\quad+\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{i=0}^{q_n-1}\sum_{h=1}^{q_n-i} \mathrm{cov}\left(H(\boldsymbol{X}_{-h+1,-h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right) .
\end{aligned}
$$

This gives further

$$
\begin{aligned}
&\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{h=0}^{q_n-2}\sum_{i=h+1}^{q_n-1} \mathrm{cov}\left(H(\boldsymbol{X}_{h+1,h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right)\\
&\quad+\frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{h=1}^{q_n-2}\sum_{i=1}^{q_n-h} \mathrm{cov}\left(H(\boldsymbol{X}_{-h+1,-h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right)\\
&=\frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{h=0}^{q_n-1}(1-h/q_n) \mathrm{cov}\left(H(\boldsymbol{X}_{h+1,h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right) \qquad (6.7.22)\\
&\quad+\frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{h=0}^{q_n-1}(1-h/q_n) \mathrm{cov}\left(H(\boldsymbol{X}_{-h+1,-h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\right) . \qquad (6.7.23)
\end{aligned}
$$

We show that the term in (6.7.22) is negligible, while the one in (6.7.23) yields the limit. We split the term in (6.7.22) into two pieces, according to $h \le r_n$ and $h > r_n$. Then the first part is bounded by

$$
\frac{1}{r_n}\sum_{h=0}^{r_n} \mathbb{E}\left[|H(\boldsymbol{X}_{h+1,h+r_n}/u_n)| \mid |\boldsymbol{X}_0|>u_n\right] = \int_0^1 g_n(v)\mathrm{d}v
$$

with

$$
g_n(v) = \mathbb{E}\left[|H(\boldsymbol{X}_{[r_n v]+1,[r_n v]+r_n}/u_n)| \mid |\boldsymbol{X}_0|>u_n\right] .
$$

Under $\mathcal{AC}(r_n, u_n)$, $g_n(v) \to 0$ (cf. the first part of Lemma 6.7.9 with $H \equiv 1$ and $\widetilde{H} = |H|$). Likewise, since $r_n/q_n \to 0$,

$$\frac{1}{r_n q_n}\sum_{h=0}^{r_n} h\mathbb{E}\left[|H(\boldsymbol{X}_{h+1,h+r_n}/u_n)|\,|\,|\boldsymbol{X}_0| > u_n\right] = q_n^{-1}\int_0^1 [r_n v] g_n(v)\mathrm{d}v \to 0\ .$$

Furthermore, applying (6.7.2),

$$\begin{aligned}&\frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sum_{h=r_n+1}^{q_n-1}(1-h/q_n)|\mathrm{cov}\left(H(\boldsymbol{X}_{h+1,h+r_n}/u_n), \mathbb{1}\{|\boldsymbol{X}_0| > u_n\}\right)| \\ &\le \frac{\|H\|_\infty}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}\sum_{h=1}^{n}\beta_{h+r_n}\end{aligned}$$

and the latter term vanishes by (6.7.1b). In summary, (6.7.22) is negligible.

For the term in (6.7.23) we write (recall that we can replace cov with $\mathbb{E}$ thanks to $\mathcal{R}(r_n, u_n)$)

$$\frac{1}{r_n}\sum_{h=0}^{r_n}\mathbb{E}\left[H(\boldsymbol{X}_{-h+1,-h+r_n}/u_n)\mid |\boldsymbol{X}_0| > u_n\right] = \int_0^1 g_n(v)\mathrm{d}v$$

with

$$g_n(v) = \mathbb{E}\left[H(\boldsymbol{X}_{-[r_n v]+1,-[r_n v]+r_n}/u_n)\mid |\boldsymbol{X}_0| > u_n\right]\ .$$

By Proposition 6.3.2, $g_n(v) \to \mathbb{E}[H(\boldsymbol{Y})]$ for each v. □

6.7.7. Proof of Theorem 6.7.15 - fidi convergence

Recall that $q_n = n - r_n + 1$ and recall the disjoint blocks of size r_n:

$$J_j = \{(j-1)r_n, \ldots, jr_n - 1\}\ ,\quad j = 1, \ldots, m_n = [q_n/r_n]\ .$$

These blocks were chosen to calculate the limiting covariance of the process $\mathbb{F}_n$. However, they are not appropriate for a proof of the central limit theorem. We need to introduce a large-small blocks decomposition. For this purpose let z_n be a sequence of integers such that $z_n \to \infty$ and

$$\lim_{n\to\infty}\frac{z_n}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}} = 0\ . \tag{6.7.24}$$

Set

$$\widetilde{m}_n = \left[\frac{q_n}{(z_n+2)r_n}\right]$$

and assume for simplicity that $\widetilde{m}_n$ is an integer. Since $z_n \to \infty$, we have $\widetilde{m}_n = o(m_n)$. For $j = 1, \ldots, \widetilde{m}_n$ define now large and small blocks as follows:

$$\begin{aligned}&L_1 = \{0, \ldots, z_n r_n - 1\}\ ,\quad S_1 = \{z_n r_n \ldots, z_n r_n + 2r_n - 1\}\ ,\\ &L_2 = \{z_n r_n + 2r_n, \ldots, 2z_n r_n + 2r_n - 1\}\ ,\quad S_2 = \{2z_n r_n + 2r_n \ldots, 2z_n r_n + 4r_n - 1\}\ ,\\ &L_j = \{(j-1)z_n r_n + 2(j-1)r_n, \ldots, jz_n r_n + 2(j-1)r_n - 1\}\ ,\\ &S_j = \{jz_n r_n + 2(j-1)r_n, \ldots, jz_n r_n + 2jr_n - 1\}\ .\end{aligned}$$

The block L_1 is obtained by merging z_n consecutive blocks $J_1, \ldots, J_{z_n}$ of size r_n. Likewise, $S_1 = J_{z_n+1} \cup J_{z_n+2}$. Therefore, the large block of size $z_n r_n$ is followed by the small block of size $2r_n$, which in turn is followed by the large block of size $z_n r_n$ and so on. All together,

$$\bigcup_{j=1}^{\widetilde{m}_n}(L_j \cup S_j) = \{0, \ldots, n - r_n\}\ .$$

Write

$$\sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) = \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(l)}(H) + \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(s)}(H) , \tag{6.7.25}$$

where now

$$\Psi_j^{(l)}(H) = \sum_{i\in L_j} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) , \quad \Psi_j^{(s)}(H) = \sum_{i\in S_j} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) .$$

With such the decomposition, $\boldsymbol{X}_1,\dots,\boldsymbol{X}_{z_nr_n+r_n-1}$ used in the definition of $\Psi_1^{(l)}(H)$ are separated by r_n+2 from the random variables that define $\Psi_2^{(l)}(H)$. The mixing condition (6.7.1a) allows us to replace $\boldsymbol{X}$ with the independent blocks process, that is, we can treat the random variables $\Psi_j^{(l)}(H)$, $j=1,\dots,\widetilde{m}_n$, as independent. The same applies to $\Psi_j^{(s)}(H)$.

Set

$$\mathbb{Z}_n(H) = \sum_{j=1}^{\widetilde{m}_n} \{Z_{n,j}(H) - \mathbb{E}[Z_{n,j}(H)]\} =: \sum_{j=1}^{\widetilde{m}_n} \bar{Z}_{n,j}(H) \tag{6.7.26}$$

with

$$Z_{n,j}(H) = \frac{\sqrt{k_n}}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\Psi_j^{(l)}(H) . \tag{6.7.27}$$

The next steps are standard.

- First, we show that the limiting variance of the large blocks process $\mathbb{Z}_n$ is the same as that of the process $\mathbb{F}_n$;
- Next, we show that the small blocks process (the second term in (6.7.25)) is negligible;
- Finally, we will verify the Lindeberg condition for the large blocks process.

Variance of the large blocks. We have (using the assumed independence of $\Psi_j^{(l)}(H)$)

$$\begin{aligned}
k_n\mathrm{var}&\left(\frac{1}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{\widetilde{m}_n}\Psi_j^{(l)}(H)\right) = \frac{k_n\widetilde{m}_n}{(q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n))^2}\mathrm{var}(\Psi_1^{(l)}(H)) \\
&\sim \frac{1}{z_nr_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathrm{var}\left(\sum_{i=0}^{z_nr_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)\right) \\
&= \frac{1}{z_nr_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathrm{var}\left(\sum_{j=1}^{z_n}\Psi_j(H)\right) ,
\end{aligned} \tag{6.7.28}$$

where in the last line we decomposed the block $L_1=\{0,\dots,z_nr_n-1\}$ into z_n disjoint blocks $J_1,\dots,J_{z_n}$, used the notation (6.7.12), the asymptotics (6.7.14) and $\widetilde{m}_n\sim m_n/z_n$.

The next steps are a repetition of the proof of Proposition 6.7.16, with the appropriate adjustments. The term in (6.7.28) becomes

$$\frac{\mathrm{var}\left(\Psi_1(H)\right)}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)} + 2\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{z_n-1}\left(1-\frac{j}{z_n}\right)\mathrm{cov}(\Psi_1(H),\Psi_{1+j}(H))$$

and as in (6.7.15) we can write it as

$$\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\{\mathrm{cov}(\Psi_2(H),\Psi_1(H)+\Psi_2(H)+\Psi_3(H))\} + \widetilde{A}_n(H) \tag{6.7.29}$$

with the reminder $\widetilde{A}_n(H)$ given this time by (cf. (6.7.16)-(6.7.17))

$$\widetilde{A}_n(H) := -2\frac{1}{z_n}\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathrm{cov}(\Psi_1(H),\Psi_2(H))$$
$$+\frac{2}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=2}^{z_n-1}\left(1-\frac{j}{z_n}\right)\{\mathrm{cov}\left(\Psi_1(H),\Psi_{1+j}(H)\right)\} =: \widetilde{A}_{n,1}(H)+\widetilde{B}_n(H)\,. \qquad (6.7.30)$$

The reminder is negligible by the same argument as before. Indeed, we note that $\widetilde{B}_n(H)$ is just $B_n(H)$ from (6.7.17) with m_n replaced with z_n. The dependence on m_n vanishes in the final stage of the proof of Lemma 6.7.19. The leading term in (6.7.29) is the same as in the proof of Proposition 6.7.16; cf. (6.7.15).

In summary, the variance of the large block process is

$$\lim_{n\to\infty}\mathrm{var}\left(\mathbb{Z}_n(H)\right)=\boldsymbol{\nu}^*(H^2)\,.$$

Variance of the small blocks. We have (using again the assumed independence of $\Psi_j^{(s)}(H)$ thanks to the beta-mixing)

$$k_n\mathrm{var}\left(\frac{1}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{\widetilde{m}_n}\Psi_j^{(s)}(H)\right)\sim\frac{1}{z_nr_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathrm{var}(\Psi_1^{(s)}(H))\,.$$

Since $\Psi_1^{(s)}(H)$ is just $\Psi_1(H)$ defined in (6.7.12), we have

$$k_n\mathrm{var}\left(\frac{1}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{\widetilde{m}_n}\Psi_j^{(s)}(H)\right)=O(1/z_n)=o(1)\,.$$

Lindeberg condition for $\mathbb{Z}_n(H)$. We need to show that for all $\eta>0$,

$$\lim_{n\to\infty}\widetilde{m}_n\mathbb{E}\left[Z_{n,1}^2(H)\mathbb{1}\{|Z_{n,1}(H)|>\eta\}\right]=0\,. \qquad (6.7.31)$$

Since H is bounded, then by (6.7.24),

$$|Z_{n,1}(H)|\le\frac{\sqrt{k_n}z_nr_n}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\|H\|_\infty\sim\frac{z_n}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}}\|H\|_\infty=o(1)\,.$$

Thus, the indicator in (6.7.31) becomes zero for large n.

Lindeberg condition for $\mathbb{Z}_n(\mathcal{E})$. The functional $\mathcal{E}$ is not bounded and we will prove the Lindeberg condition under $\mathcal{S}(r_n,u_n)$. Write

$$\widetilde{w}_n=\frac{\sqrt{k_n}}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}$$

so that

$$Z_{n,1}(\mathcal{E})=\widetilde{w}_n\sum_{i=0}^{z_nr_n-1}\mathcal{E}\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)=\widetilde{w}_n\sum_{i=0}^{z_nr_n-1}\sum_{j=i+1}^{i+r_n}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}$$
$$=\widetilde{w}_n\left\{\sum_{j=1}^{r_n}\sum_{i=0}^{j-1}+\sum_{j=r_n+1}^{r_n(z_n+1)}\sum_{i=j-r_n}^{j-1}\right\}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}\le\widetilde{w}_nr_n\sum_{j=1}^{r_n(z_n+1)}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}$$
$$\le\frac{\sqrt{k_n}}{q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}r_n\sum_{j=1}^{2r_n}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}=\frac{1+o(1)}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}}\sum_{j=1}^{2r_n}\mathbb{1}\{|\boldsymbol{X}_j|>u_n\}\,.$$

The last term can be recognized as one (scaled) block of size $2r_n$ of the tail empirical process $\widetilde{\mathbb{T}}_n(s)$. [KSW19, Lemma 3.6] (see also [KS20, Lemma 9.2.8]) gives

$$\lim_{n\to\infty} m_n \mathbb{E}\left[Z_{n,1}^2(\mathcal{E})\mathbb{1}\{|Z_{n,1}(\mathcal{E})| > \eta\}\right] = 0 \; .$$

If moreover $\mathcal{R}(r_n, u_n)$ holds then

$$\lim_{n\to\infty} m_n \mathbb{E}\left[\bar{Z}_{n,1}^2(\mathcal{E})\mathbb{1}\{|\bar{Z}_{n,1}(\mathcal{E})| > \eta\}\right] = 0 \; .$$

Since $\widetilde{m}_n = o(m_n)$, we obtain the Lindeberg condition for $\mathbb{Z}_n(\mathcal{E})$.

6.7.8. Proof of Theorem 6.7.15 - asymptotic equicontinuity

We need the following lemma which is an adapted version of Theorem 2.11.1 in [vdVW96]. Let $\mathbb{Z}_n$ be the empirical process indexed by a semi-metric space $(\mathcal{G}, \rho)$, defined by

$$\mathbb{Z}_n(f) = \sum_{j=1}^{\widetilde{m}_n} \{Z_{n,j}(f) - \mathbb{E}[Z_{n,j}(f)]\} \; ,$$

where $\{Z_{n,j}, n \geq 1\}$, $j = 1, \ldots, \widetilde{m}_n$, are i.i.d. separable, stochastic processes and $\widetilde{m}_n$ is a sequence of integers such that $\widetilde{m}_n \to \infty$. Define the random semi-metric d_n on $\mathcal{G}$ by

$$d_n^2(f, g) = \sum_{j=1}^{\widetilde{m}_n} \{Z_{n,j}(f) - Z_{n,j}(g)\}^2 \; , f, g \in \mathcal{G} \; .$$

Lemma 6.7.20. *Assume that $(\mathcal{G}, \rho)$ is totally bounded. Assume moreover that:*

(i) For all $\eta > 0$,

$$\lim_{n\to\infty} \widetilde{m}_n \mathbb{E}[\|Z_{n,1}\|_{\mathcal{G}}^2 \mathbb{1}\{\|Z_{n,1}\|_{\mathcal{G}}^2 > \eta\}] = 0 \; .$$

(ii) For every sequence $\{\delta_n\}$ which decreases to zero,

$$\lim_{n\to\infty} \sup_{\substack{f,g\in\mathcal{G} \\ \rho(f,g)\leq\delta_n}} \mathbb{E}[d_n^2(f, g)] = 0 \; . \tag{6.7.32}$$

(iii) There exists a measurable majorant $N^*(\mathcal{G}, d_n, \epsilon)$ of the covering number $N(\mathcal{G}, d_n, \epsilon)$ such that for every sequence $\{\delta_n\}$ which decreases to zero,

$$\int_0^{\delta_n} \sqrt{\log N^*(\mathcal{G}, d_n, \epsilon)} \mathrm{d}\epsilon \xrightarrow{\mathbb{P}} 0 \; . \tag{6.7.33}$$

Then $\{\mathbb{Z}_n, n \geq 1\}$ is asymptotically ρ-equicontinuous, i.e. for each $\eta > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\substack{f,g\in\mathcal{G} \\ \rho(f,g)<\delta}} |\mathbb{Z}_n(f) - \mathbb{Z}_n(g)| > \eta \right) = 0 \; .$$

Remark 6.7.21. The separability assumption is not in [vdVW96]. It implies measurability of $\|Z_{n,1}\|_{\mathcal{G}}$. Furthermore, the separability also implies that for all $\delta > 0$, $n \in \mathbb{N}$, $(e_j)_{1\leq j\leq \widetilde{m}_n} \in \{-1, 0, 1\}^{\widetilde{m}_n}$ and $i \in \{1, 2\}$, the supremum

$$\sup_{\substack{f,g\in\mathcal{G} \\ \rho(f,g)<\delta}} \left| \sum_{j=1}^{\widetilde{m}_n} e_j \left(Z_{n,j}(f) - Z_{n,j}(g)\right)^i \right| = \sup_{\substack{f,g\in\mathcal{G}_0 \\ \rho(f,g)<\delta}} \left| \sum_{j=1}^{\widetilde{m}_n} e_j \left(Z_{n,j}(f) - Z_{n,j}(g)\right)^i \right|$$

is measurable, which is an assumption of [vdVW96]. ⊕

Asymptotic equicontinuity of the empirical process of sliding blocks

Recall the big-blocks process $\mathbb{Z}_n(H)$ (cf. (6.7.26)-(6.7.27)). Recall also that thanks to β-mixing we can consider random variables $\Psi_j^{(l)}(H)$, $j=1,\ldots,\widetilde{m}_n$ to be independent. We need to prove asymptotic equicontinuity of $\mathbb{Z}_n(H)$ indexed by the class $\mathcal{G}=\{H_s, s\in[s_0,t_0]\}$ equipped with the metric $\boldsymbol{\rho}^*(H,\widetilde{H})=\boldsymbol{\nu}^*(\{H-\widetilde{H}\}^2)$. The same argument can be used to prove asymptotic equicontinuity for the small blocks process. This yields asymptotic equicontinuity of $\mathbb{F}_n(H.)$. We note further that asymptotic continuity of $\mathbb{F}_n(\mathcal{E}.)$ follows from [KSW19].

- The Lindeberg condition (i) of Lemma 6.7.20 holds because the class $\mathcal{G}$ is linearly ordered and by applying (6.7.31).
- Since $\mathcal{G}$ is linearly ordered, the random entropy condition (6.7.33) of Lemma 6.7.20 holds.
- Define the random metric

$$d_n^2(H,\widetilde{H})=\sum_{j=1}^{\widetilde{m}_n}(Z_{n,j}(H)-Z_{n,j}(\widetilde{H}))^2\;.$$

We need to evaluate $\mathbb{E}[d_n^2(H_s,H_t)]$:

$$\begin{aligned}
&\mathbb{E}[d_n^2(H_s,H_t)]\\
&=\frac{k_n\widetilde{m}_n}{(q_nr_n\mathbb{P}(|\boldsymbol{X}_0|>u_n))^2}\mathbb{E}\left[\left(\sum_{i=0}^{z_nr_n-1}\{H_s\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)-H_t\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right)\}\right)^2\right]\\
&\sim\frac{1}{z_nr_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[\left(\sum_{j=1}^{z_n}\{\Psi_j(H_s)-\Psi_j(H_t)\}\right)^2\right]\;,
\end{aligned}\tag{6.7.34}$$

where in the last line we decomposed the block L_1 into z_n disjoint blocks $J_1,\ldots,J_{z_n}$, used the notation (6.7.12), the asymptotics (6.7.14) and $\widetilde{m}_n\sim m_n/z_n$; cf. (6.7.28).
The term in (6.7.34) becomes

$$\begin{aligned}
&\frac{\mathbb{E}[(\Psi_1(H_s)-\Psi_1(H_t))^2]}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\\
&+2\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=1}^{z_n-1}\left(1-\frac{j}{z_n}\right)\mathbb{E}[\{\Psi_1(H_s)-\Psi_1(H_t)\}\{\Psi_{1+j}(H_s)-\Psi_{1+j}(H_t)\}]
\end{aligned}$$

and as in (6.7.15) we can write it as

$$\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}\left[\{\Psi_2(H_s)-\Psi_2(H_t)\}\sum_{j=1}^{2}\{\Psi_j(H_s)-\Psi_j(H_t)\}\right\}+\widetilde{A}_n(H,s,t)\tag{6.7.35}$$

with the reminder (cf. (6.7.30))

$$\begin{aligned}
\widetilde{A}_n(H,s,t)&:=-2\frac{1}{z_n}\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\mathbb{E}[\{\Psi_1(H_s)-\Psi_1(H_t)\}\{\Psi_2(H_s)-\Psi_2(H_t)\}]\\
&+2\frac{1}{r_n^3\mathbb{P}(|\boldsymbol{X}_0|>u_n)}\sum_{j=2}^{z_n-1}\left(1-\frac{j}{z_n}\right)\mathbb{E}\left[\{\Psi_1(H_s)-\Psi_1(H_t)\}\{\Psi_{j+1}(H_s)-\Psi_{j+1}(H_t)\}\right]\\
&=\widetilde{A}_{n,1}(H_s-H_t)+\widetilde{B}_n(H_s-H_t)\;.
\end{aligned}$$

Remark 6.7.8 applies and hence by Lemma 6.7.19,

$$\lim_{n\to\infty}\sup_{s\in[s_0,t_0]}\widetilde{B}_n(H_s-H_t)=0\;.$$

The leading term in (6.7.35) is decomposed as (cf. (6.7.20))

$$\frac{1}{r_n}\widetilde{g}_n(0;H_s-H_t)+2\frac{1}{r_n}\sum_{i=1}^{r_n-1}\widetilde{g}_n(i/r_n;H_s-H_t)+2R_n(H_s-H_t)\;.$$

Again, Remark 6.7.8 applies and (6.7.19) gives

$$\lim_{n\to\infty}\sup_{s\in[s_0,t_0]}R_n(H_s-H_t)=0\;.$$

It remains to show that for every sequence $\{\delta_n\}$ decreasing to zero,

$$\lim_{n\to\infty}\sup_{\substack{s,t\in[s_0,t_0]\\|s-t|\le\delta_n}}\frac{1}{r_n}\sum_{i=1}^{r_n-1}\widetilde{g}_n(i/r_n;H_s-H_t)=\lim_{n\to\infty}\sup_{\substack{s,t\in[s_0,t_0]\\|s-t|\le\delta_n}}\int_0^1\widetilde{g}_n(\xi,H_s-H_t)\mathrm{d}\xi=0\;.$$

Because of the monotonicity

$$|\widetilde{g}_n(\xi,H_s-H_t)|\le 2\sup_{s\in[s_0,t_0]}|H_s|\boldsymbol{\nu}^*_{n,r_n}(|H_s-H_t|)\le 2\max\{|H_{s_0}|,|H_{t_0}|\}|\boldsymbol{\nu}^*_{n,r_n}(H_s)-\boldsymbol{\nu}^*_{n,r_n}(H_t)|\;.$$

The convergence of $\boldsymbol{\nu}^*_{n,r_n}(H_s)$ to $s^{-\alpha}\boldsymbol{\nu}^*(H^2)$ is uniform on $[s_0,t_0]$. Thus, for $s,t\in[s_0,t_0]$,

$$\left|\boldsymbol{\nu}^*_{n,r_n}(H_s)-\boldsymbol{\nu}^*_{n,r_n}(H_t)\right|\quad\le\quad 2\sup_{s_0\le u\le t_0}\left|\boldsymbol{\nu}^*_{n,r_n}(H_u)-\boldsymbol{\nu}^*(H_u)\right|\;+\;\boldsymbol{\nu}^*(H)\{s^{-\alpha}\;-\;t^{-\alpha}\}\quad.$$

Fix $\eta>0$. For large enough n, the uniform convergence yields

$$\sup_{\substack{s_0\le s,t\le t_0\\|s-t|\le\delta_n}}\left|\boldsymbol{\nu}^*_{n,r_n}(H_s)-\boldsymbol{\nu}^*_{n,r_n}(H_t)\right|\le\eta+\boldsymbol{\nu}^*(H)\sup_{\substack{s_0\le s,t\le t_0\\|s-t|\le\delta_n}}\{s^{-\alpha}-t^{-\alpha}\}\le\eta+\alpha s_0^{-\alpha-1}\delta_n\boldsymbol{\nu}^*(H)\;.$$

This proves that (6.7.32) holds.

The conditions of Lemma 6.7.20 hold, thus the sequence $\mathbb{Z}_n$ is asymptotically equicontinuous.

6.7.9. Proof of Theorem 6.4.3

Write $\zeta_n=|\boldsymbol{X}|_{(n:n-k_n)}/u_n$. Since $k_n=n\mathbb{P}(|\boldsymbol{X}_0|>u_n)$, we have the relationship $\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H)=\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H_{\zeta_n})$ (cf. (6.1.8)-(6.1.9)). Therefore,

$$\sqrt{k_n}\left\{\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H)-\boldsymbol{\nu}^*(H)\right\}=\mathbb{F}_n(H_{\zeta_n})+\sqrt{k_n}\left\{\boldsymbol{\nu}^*(H_{\zeta_n})-\boldsymbol{\nu}^*(H)\right\}\;.\tag{6.7.36}$$

Step 1. Theorem 6.7.15 gives local uniform convergence of $\{\mathbb{F}_n(H_s),s\in[s_0,t_0]\}$ to a continuous Gaussian process $\mathbb{G}$. At the same time, convergence of $\{\mathbb{F}_n(\mathcal{E}_s),s\in[s_0,t_0]\}$ yields $\zeta_n\xrightarrow{\mathbb{P}}1$, jointly with $\mathbb{F}_n(H_s)$. Therefore, $\mathbb{F}_n(H_{\zeta_n})\xrightarrow{\mathrm{d}}\mathbb{G}(H)$.
Step 2. Using Vervaat's theorem, we have, jointly with the previous convergence, $\sqrt{k}(\zeta_n^{-\alpha}-1)\xrightarrow{\mathrm{d}}\mathbb{G}(\mathcal{E})$. Therefore, by the homogeneity of $\boldsymbol{\nu}^*$,

$$\sqrt{k}\left\{\boldsymbol{\nu}^*(H_{\zeta_n})-\boldsymbol{\nu}^*(H)\right\}=\boldsymbol{\nu}^*(H)\sqrt{k}(\zeta_n^{-\alpha}-1)\xrightarrow{\mathrm{d}}-\boldsymbol{\nu}^*(H)\mathbb{G}(\mathcal{E})\;.$$

Since the convergences hold jointly, we conclude the result.

6.7.10. Auxiliary results

Lemma 6.7.22 (Problems 5.24 and 5.25 in [KS20]). *Assume that* $\mathbb{P}(\lim_{|j|\to\infty} |\boldsymbol{Y}_j| = 0) = 1$ *and let* H, H' *be bounded functionals on* $(\mathbb{R}^d)^{\mathbb{Z}}$ *such that* $H'(\boldsymbol{x}) = 0$ *if* $\boldsymbol{x}^* \le 1$ *and* $\mathbb{E}[|H(\boldsymbol{Y})||H'(\boldsymbol{Y}_{0,\infty}) - H'(\boldsymbol{Y}_{1,\infty})|] < \infty$. *Then*

$$\begin{aligned}
&\boldsymbol{\nu}^*(HH') = \mathbb{E}[H(\boldsymbol{Y})\{H'(\boldsymbol{Y}_{0,\infty}) - H'(\boldsymbol{Y}_{1,\infty})\}] \;, \\
&\boldsymbol{\nu}^*(H\mathcal{E}) = \mathbb{E}[H(\boldsymbol{Y})] \;, \quad \boldsymbol{\nu}^*(\mathcal{E}) = 1 \;, \quad \boldsymbol{\nu}^*(\mathcal{E}^2) = \sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) \;.
\end{aligned}$$

Proof. Applying (6.2.7) and the time change formula (see [PS18, Lemma 2.2]), we obtain

$$\begin{aligned}
\boldsymbol{\nu}^*(HH') &= \mathbb{E}[H(\boldsymbol{Y})H'(\boldsymbol{Y})\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] = \mathbb{E}[|H(\boldsymbol{Y})||H'(\boldsymbol{Y}_{0,\infty})|\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \\
&\le \sum_{j=0}^{\infty} \mathbb{E}[|H(\boldsymbol{Y})||H'(\boldsymbol{Y}_{j,\infty}) - H'(\boldsymbol{Y}_{j+1,\infty})|\mathbb{1}\{|\boldsymbol{Y}_j| > 1\}\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-1} \le 1\}] \\
&= \sum_{j=0}^{\infty} \mathbb{E}[H(\boldsymbol{Y})|H'(\boldsymbol{Y}_{0,\infty}) - H'(\boldsymbol{Y}_{1,\infty})|\mathbb{1}\{|\boldsymbol{Y}_{-j}| > 1\}\mathbb{1}\{\boldsymbol{Y}^*_{-\infty,-j-1} \le 1\}] \\
&= \mathbb{E}[H(\boldsymbol{Y})|H'(\boldsymbol{Y}_{0,\infty}) - H'(\boldsymbol{Y}_{1,\infty})|] < \infty \;.
\end{aligned}$$

This proves that $\boldsymbol{\nu}^*(HH') < \infty$. Hence, we can switch the expectation with the summation and the first result follows. The second statement follows by noting that $\mathcal{E}(\boldsymbol{Y}_{0,\infty}) - \mathcal{E}(\boldsymbol{Y}_{1,\infty}) = 1$ almost surely. □

Lemma 6.7.23 (Example 6.2.2 and Problem 6.7 in [KS20]). *Assume that* $\mathbb{P}(\lim_{|j|\to\infty} |\boldsymbol{Y}_j| = 0) = 1$ *and let* $\pi(m)$, $m \ge 0$, *be the limiting cluster size distribution. Then*

$$\sum_{m=1}^{\infty} m\pi(m) = \vartheta^{-1} \;, \quad \sum_{m=1}^{\infty} m^2\pi(m) = \vartheta^{-1} \sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) \;.$$

Proof. For the first statement, applying (6.2.7) and Lemma 6.7.22, we have,

$$\sum_{m=1}^{\infty} m\pi(m) = \sum_{m=1}^{\infty} m\mathbb{P}(\mathcal{E}(\boldsymbol{Y}) = m \mid \mathcal{C}_0(\boldsymbol{Y}) = 0) = \mathbb{E}[\mathcal{E}(\boldsymbol{Y}) \mid \mathcal{C}_0(\boldsymbol{Y}) = 0] = \vartheta^{-1}\boldsymbol{\nu}^*(\mathcal{E}) = \vartheta^{-1} \;.$$

Likewise,

$$\sum_{m=1}^{\infty} m^2\pi(m) = \mathbb{E}[\mathcal{E}^2(\boldsymbol{Y}) \mid \mathcal{C}_0(\boldsymbol{Y}) = 0] = \vartheta^{-1}\boldsymbol{\nu}^*(\mathcal{E}^2) = \vartheta^{-1} \sum_{j\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_j| > 1) \;.$$

□

Bibliography

[BB18] Betina Berghaus and Axel Bücher. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Annals of Statistics*, 46(5):2307–2335, 2018.

[BBKS20] Clemonell Bilayi-Biakana, Rafał Kulik, and Philippe Soulier. Statistical inference for heavy tailed series with extremal independence. *Extremes*, 23(1):1–33, 2020.

[BPS18] Bojan Basrak, Hrvoje Planinić, and Philippe Soulier. An invariance principle for sums and record times of regularly varying stationary sequences. *Probability Theory and Related Fields*, 172(3-4):869–914, 2018.

[BS09] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[BS18a] Axel Bücher and Johan Segers. Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125, 2018.

[BS18b] Axel Bücher and Johan Segers. Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462, 2018.

[DM09] Richard A. Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 38A:977–1009, 2009. Probability, statistics and seismology.

[DH95] Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability*, 23(2):879–917, 1995.

[DHS18] Clément Dombry, Enkelejd Hashorva, and Philippe Soulier. Tail measure and spectral tail process of regularly varying time series. *Annals of Applied Probability*, 28(6):3884–3921, 2018.

[DK20] Holger Drees and Miran Knezevic. Peak-over-threshold estimators for spectral tail processes: Random vs deterministic thresholds. *Extremes*, 2020. DOI: https://doi.org/10.1007/s10687-019-00367-x.

[DN20] Holger Drees and Sebastian Neblung. Asymptotics for sliding blocks estimators of rare events. arXiv:2003.01016, 2020.

[DR10] Holger Drees and Holger Rootzén. Limit theorems for empirical processes of cluster functionals. *Annals of Statistics*, 38(4):2145–2186, 2010.

[DSW15] Holger Drees, Johan Segers, and Michał Warchoł. Statistics for tail processes of Markov chains. *Extremes*, 18(3):369–402, 2015.

[EKM97] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, 1997.

[Hsi91] Tailen Hsing. Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37(1):117–139, 1991.

[Ibr62] I. A. Ibragimov. Some limit theorems for stationary processes. *Theor. Probab. Appl.*, 7:349–382, 1962.

[KS20] Rafał Kulik and Philippe Soulier. *Heavy tailed time series.* Springer, 2020.

[KSW19] Rafał Kulik, Philippe Soulier, and Olivier Wintenberger. The tail empirical process of regularly varying functions of geometrically ergodic markov chains. *Stochastic Processes and their Applications*, 129(1):4209–4238, 2019.

[MW13] Thomas Mikosch and Olivier Wintenberger. Precise large deviations for dependent regularly varying sequences. *Probability Theory and Related Fields*, 156(3-4):851–887, 2013.

[MW14] Thomas Mikosch and Olivier Wintenberger. 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, 2014.

[MW16] Thomas Mikosch and Olivier Wintenberger. A large deviations approach to limit theorem for heavy-tailed time series. *Probability Theory and Related Fields*, 166(1-2):233–269, 2016.

[PS18] Hrvoje Planinić and Philippe Soulier. The tail process revisited. *Extremes*, 21(4):551–579, 2018.

[RLdH98] Holger Rootzén, Ross M. Leadbetter, and Laurens de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8(3):868–885, 1998.

[RSF09] Christian Y. Robert, Johan Segers, and Christopher A. T. Ferro. A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3:993–1020, 2009.

[SW94] Richard L. Smith and Ishay Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society. Series B. Methodological*, 56(3):515–528, 1994.

[vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes.* Springer, New York, 1996.

[WN98] Ishay Weissman and S. Yu. Novak. On blocks and runs estimators of the extremal index. *Journal of Stat. Planning and Inference*, 66(2):281–288, 1998.

[ZVB20] Nan Zou, Stanislav Volgushev, and Axel Bücher. Multiple block sizes and overlapping blocks for multivariate time series extremes. arXiv:1907.09477, 2020.

	$\rho = 0.9$, Extremal Index=0.34				$\rho = 0.5$, Extremal Index= 0.94			
(k %)	$k = 5$		$k = 10$		$k = 5$		$k = 10$	
$r_n = 7$								
Disjoint bl	**0.34**	(0.05)	**0.31**	(0.03)	**0.68**	(0.05)	**0.58**	(0.03)
Sliding bl	**0.35**	(0.04)	**0.31**	(0.03)	**0.68**	(0.04)	**0.58**	(0.03)
$r_n = 8$								
Disjoint bl	**0.32**	(0.05)	**0.29**	(0.03)	**0.67**	(0.05)	**0.56**	(0.03)
Sliding bl	**0.33**	(0.04)	**0.29**	(0.03)	**0.67**	(0.04)	**0.56**	(0.03)
$r_n = 9$								
Disjoint bl	**0.32**	(0.05)	**0.28**	(0.03)	**0.66**	(0.05)	**0.53**	(0.03)
Sliding bl	**0.32**	(0.04)	**0.28**	(0.03)	**0.65**	(0.05)	**0.53**	(0.03)
$r_n = 10$								
Disjoint bl	**0.30**	(0.05)	**0.26**	(0.03)	**0.64**	(0.05)	**0.52**	(0.03)
Sliding bl	**0.30**	(0.04)	**0.26**	(0.03)	**0.63**	(0.05)	**0.52**	(0.03)

TABLE 6.1

The median and the variance (in brackets) of disjoint and sliding blocks estimators for the extremal index. Data are simulated from AR(1) with $\alpha = 4$, $\rho = 0.5$ (thus, $\theta = 0.94$), and $\rho = 0.9$ (thus $\theta = 0.34$). Block size $r_n = 7,\ 8,\ 9,\ 10$. The number of order statistics is $k = 5\%$ and 10% for a sample $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

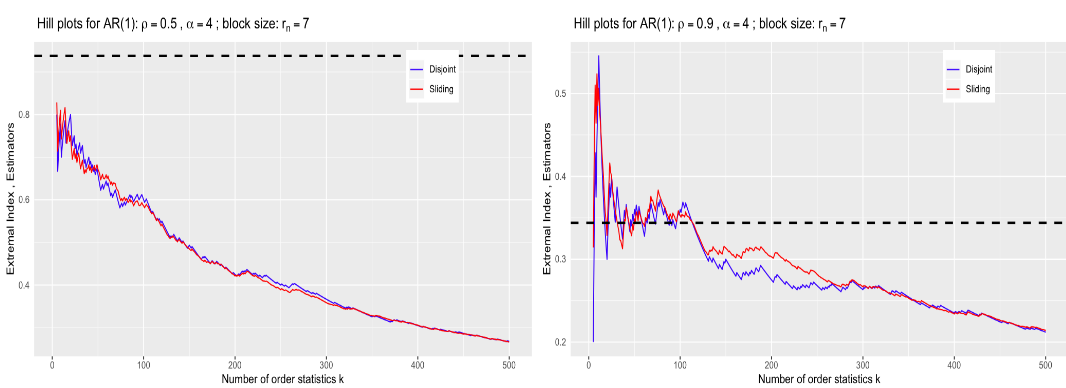

FIG 6.1. *Hill plots of sliding and disjoint blocks estimators for the extremal index plotted against the number of order statistics k. Data are simulated from $AR(1)$ with $\rho = 0.5$ and $\theta = 0.94$ (left panel), $\rho = 0.9$ and $\theta = 0.34$ (right panel), $\alpha = 4$, block size $r_n = 7$. Dotted lines indicated the true value of the extremal index.*

(k %)	$\rho = 0.9$, Stop-loss Index=0.096				$\rho = 0.5$, Stop-loss Index= 0.11			
	$k = 50$		$k = 70$		$k = 50$		$k = 70$	
$r_n = 7$								
Disjoint bl	**0.0320**	(0.009)	**0.0414**	(0.008)	**0.0960**	(0.009)	**0.1071**	(0.007)
Sliding bl	**0.0314**	(0.008)	**0.0418**	(0.008)	**0.0954**	(0.008)	**0.1067**	(0.006)
$r_n = 8$								
Disjoint bl	**0.0340**	(0.008)	**0.0443**	(0.007)	**0.0980**	(0.009)	**0.1071**	(0.006)
Sliding bl	**0.0332**	(0.008)	**0.0436**	(0.007)	**0.0968**	(0.007)	**0.1055**	(0.005)
$r_n = 9$								
Disjoint bl	**0.0360**	(0.008)	**0.0457**	(0.007)	**0.0960**	(0.008)	**0.1043**	(0.006)
Sliding bl	**0.0349**	(0.007)	**0.0460**	(0.007)	**0.0967**	(0.007)	**0.1038**	(0.004)
$r_n = 10$								
Disjoint bl	**0.0320**	(0.007)	**0.0429**	(0.007)	**0.0980**	(0.007)	**0.1014**	(0.005)
Sliding bl	**0.0314**	(0.007)	**0.0424**	(0.006)	**0.0974**	(0.006)	**0.1009**	(0.004)

TABLE 6.2

The median and the variance (in brackets) of disjoint and sliding blocks estimators for stop-loss index with $S = 0.7$. Data are simulated from AR(1) with $\alpha = 4$, $\rho = 0.5$, $\rho = 0.9$. The block size is $r_n = 7, 8, 9, 10$. The number of order statistics is $k = 50\%$ and 70% for a sample $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

(k %)	Extremal Index=0.612			
	$k = 5$		$k = 10$	
$r_n = 7$				
Disjoint bl	**0.680**	(0.06)	**0.620**	(0.04)
Sliding bl	**0.670**	(0.06)	**0.620**	(0.03)
$r_n = 8$				
Disjoint bl	**0.660**	(0.06)	**0.600**	(0.04)
Sliding bl	**0.648**	(0.06)	**0.593**	(0.03)
$r_n = 9$				
Disjoint bl	**0.640**	(0.06)	**0.590**	(0.04)
Sliding bl	**0.631**	(0.05)	**0.567**	(0.03)
$r_n = 10$				
Disjoint bl	**0.620**	(0.06)	**0.550**	(0.04)
Sliding bl	**0.616**	(0.05)	**0.546**	(0.03)

TABLE 6.3

The median and the variance (in brackets) of disjoint and sliding blocks estimators for the extremal index in ARCH(1) model with $\lambda = 0.9$. The block size is $r_n = 7, 8, 9, 10$. The number of order statistics is $k = 5\%$ and 10% for a sample $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

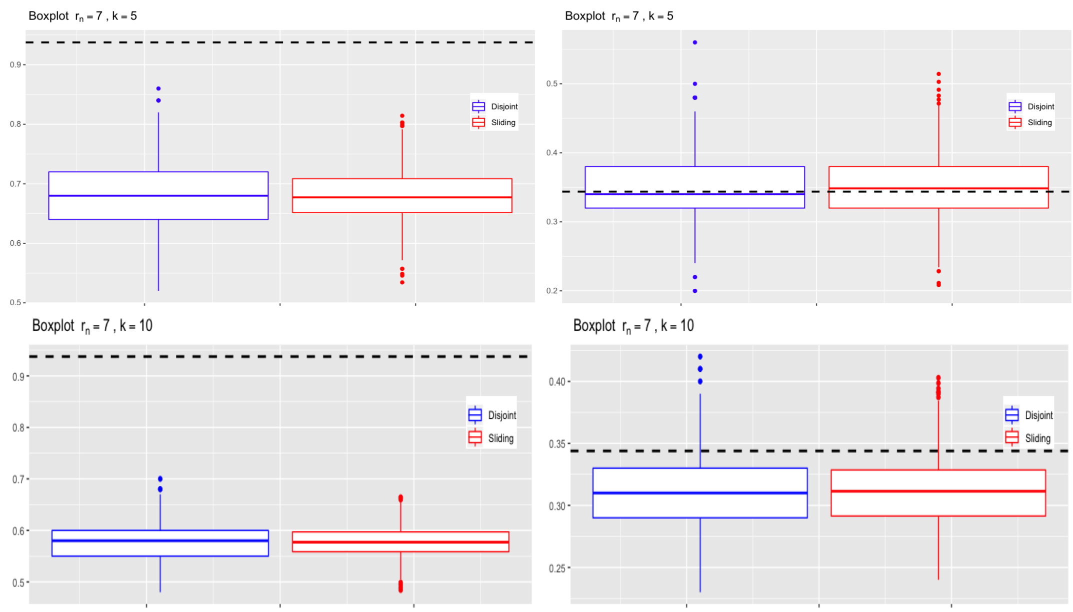

FIG 6.2. *Monte Carlo simulations of sliding and disjoint blocks estimators for the extremal index. Data are simulated from AR(1) with $\rho = 0.5$ and $\theta = 0.94$ (left panel), $\rho = 0.9$ and $\theta = 0.34$ (right panel), $\alpha = 4$, block size $r_n = 7$ and the number of order statistics $k = 5\%$ and 10%. Dotted lines indicated the true value of the extremal index.*

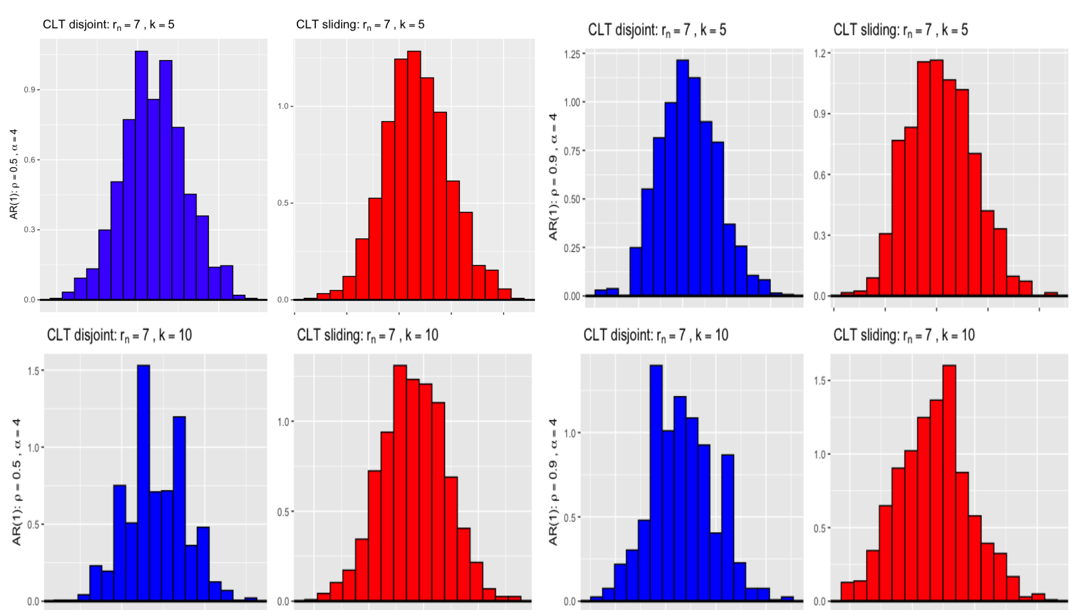

FIG 6.3. *Monte Carlo simulations of disjoint (blue) and sliding (red) blocks estimators for the extremal index. Data are simulated from AR(1) with $\rho = 0.5$ (left panel), $\rho = 0.9$ (right panel), $\alpha = 4$, block size $r_n = 7$ and the order statistics $k = 5\%$ and 10%.*

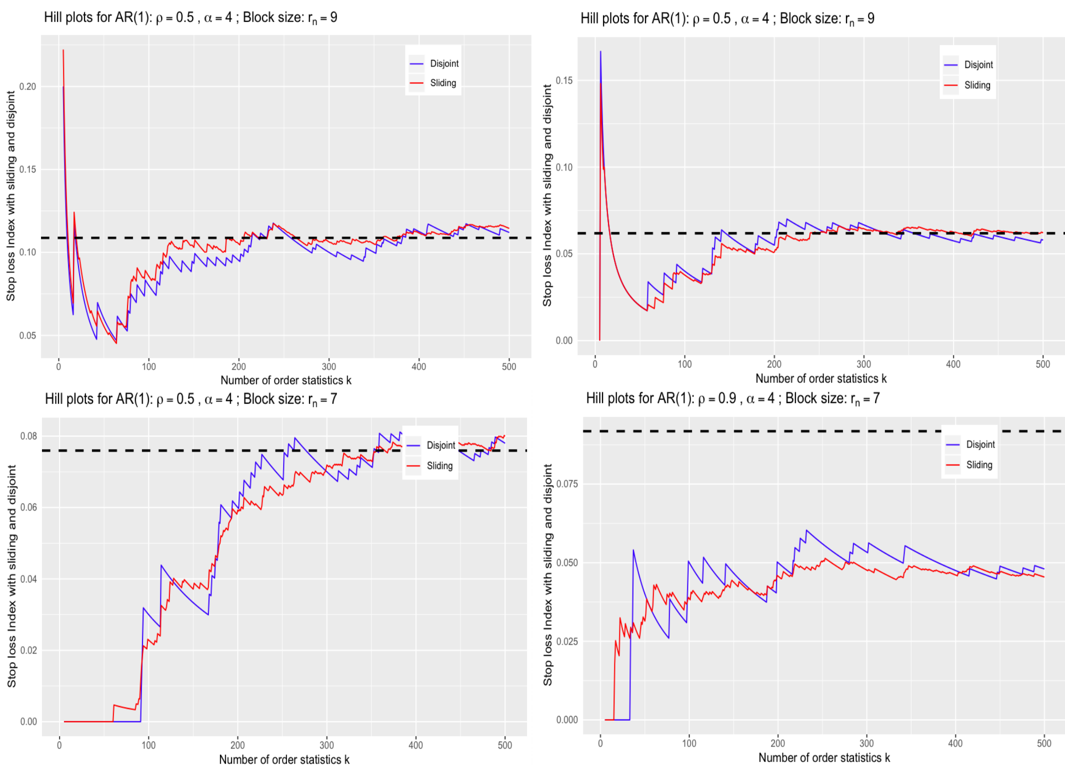

FIG 6.4. *Hill plots of disjoint and sliding blocks estimators for the stop-loss index plotted against the number of order statistics k. Data are simulated from $AR(1)$ with: $\rho = 0.5$, $S = 0.3$ (top left), $\rho = 0.5$, $S = 0.7$ (bottom left), $\rho = 0.5$, $S = 0.7$ (top right) and $\rho = 0.9$, $S = 0,7$, (bottom right), $\alpha = 4$, block size $r_n = 7,\ 9$.*

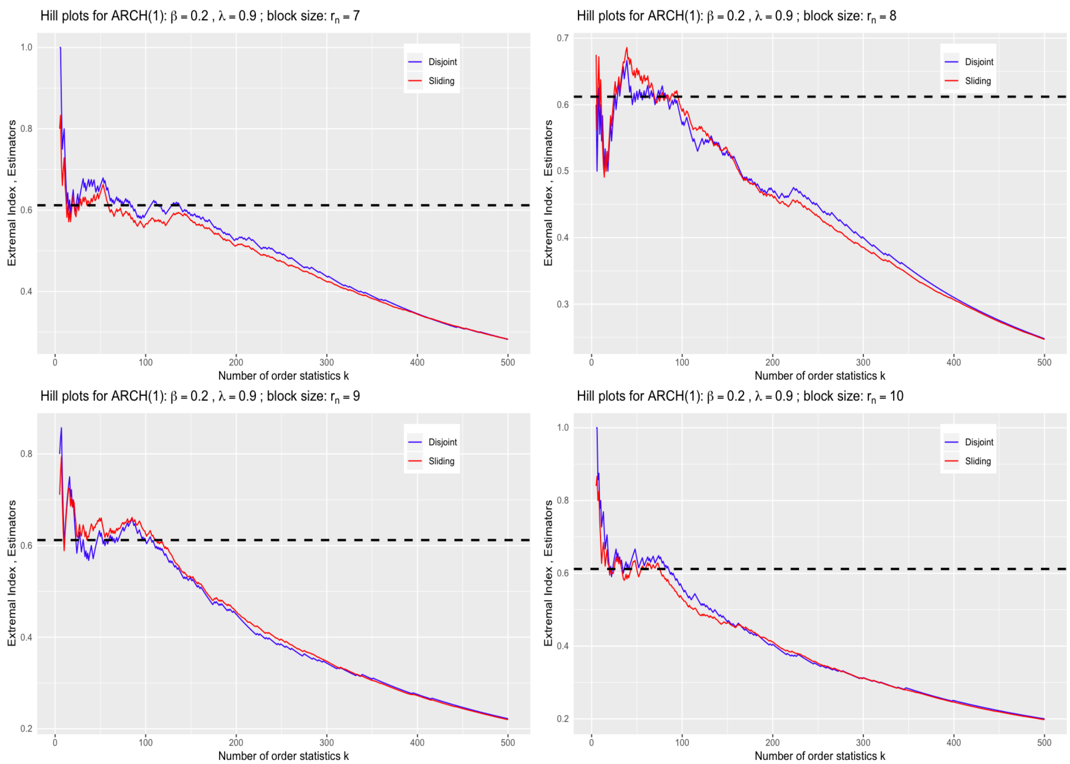

Fig 6.5. *Hill plots of disjoint and sliding blocks estimators for extremal index plotted against the number of order statistics k. Data are simulated from $ARCH(1)$ with: $\lambda = 0.9$.*

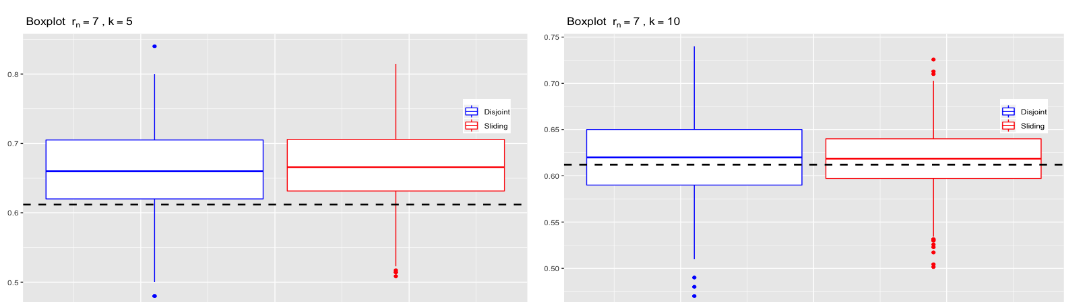

Fig 6.6. *Monte Carlo simulations of sliding and disjoint blocks estimators for the extremal index. Data are simulated from ARCH(1) with $\lambda = 0.9$, block size $r_n = 7$ and the number of order statistics $k = 5\%$ (left panel) and 10% (right panel). Dotted lines indicated the true value of the extremal index.*

Chapter 7

Central limit theorem for runs estimators

This chapter consists of the paper titled *Estimation of cluster functionals for regularly varying time series: runs estimators*. The paper is self-contained, with all the required definitions. The references can be found at the end of the paper, on page 116. This paper should appear soon in *Electronic Journal of Statistics*.

Electronic Journal of Statistics
Vol. 16(1) (2021) 3561–3607

Estimation of cluster functionals for regularly varying time series: runs estimators

Youssouph Cissokho* **and Rafal Kulik***

Abstract: Cluster indices describe extremal behaviour of stationary time series. We consider runs estimators of cluster indices. Using a modern theory of multivariate, regularly varying time series, we obtain central limit theorems under conditions that can be easily verified for a large class of models. In particular, we show that blocks and runs estimators have the same limiting variance.

Keywords and phrases: Regularly varying time series, Extremes, Cluster index, Extremal index.

7.1. Introduction

Consider a stationary, regularly varying $\mathbb{R}^d$-valued time series $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$. We are interested in its extremal behaviour. A classical approach to this problem is to calculate the *extremal index*. If $|\cdot|$ is an arbitrary norm on $\mathbb{R}^d$, then the extremal index θ (if exists) of $\{|\boldsymbol{X}_j|, j \in \mathbb{Z}\}$ is defined as a parameter in the limiting distribution of the maxima. With Q being the quantile function of $|\boldsymbol{X}_0|$ and $a_n = Q(1 - 1/n)$ we have

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1} \max_{j=1,\dots,n} \{|\boldsymbol{X}_1|, \dots, |\boldsymbol{X}_n|\} \le x) = \exp(-\theta x^{-\alpha}) \;, \quad x > 0 \;.$$

The parameter $\theta \in [0, 1]$ indicates the amount of clustering, with $\theta = 1$ (the case of extremal independence) meaning no-clustering of large values. If $\theta = 0$, then the limiting distribution is degenerated. This situation is commonly referred to as *long memory in extremes*. In this case, a different normalization of order $o(a_n)$ is required. See [Sam2016]. The case $\theta = 0$ is excluded from our studies. We would like to point out that there is a notion of the multivariate extremal index; see [BGTS09, Section 10.5.2].

The extremal index is just one parameter that describes clustering of extremes. Informally speaking, it arises as the limit

$$\lim_{n\to\infty} \frac{\mathbb{E}[H((\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r_n})/u_n)]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \;,$$

for the particular choice of function $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$, and a suitable choice of the scaling sequence $u_n \to \infty$ and the block size $r_n \to \infty$. (Formally speaking, $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r_n})$ is a random element of $(\mathbb{R}^d)^{r_n}$, while the domain of H is $(\mathbb{R}^d)^{\mathbb{Z}}$. This inconsistency will be explained later).

In particular, the extremal index is achieved by applying a suitable functional to a cluster:

$$H((\boldsymbol{X}_1, \dots, \boldsymbol{X}_{r_n})/u_n) = \mathbb{1}\{\max\{|\boldsymbol{X}_1|, \dots, |\boldsymbol{X}_{r_n}|\} > u_n\} \;.$$

That is,

$$\theta = \lim_{n\to\infty} \frac{\mathbb{P}(\max\{|\boldsymbol{X}_1|, \dots, |\boldsymbol{X}_{r_n}|\} > u_n)}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \;. \tag{7.1.1}$$

*Department of Mathematics and Statistics, University of Ottawa, 150 Louis Pasteur Private, Ottawa ON K1N 6N5

Informally speaking, a cluster is a triangular array $(\boldsymbol{X}_1/u_n, \ldots, \boldsymbol{X}_{r_n}/u_n)$ with $r_n, u_n \to \infty$ that converges in distribution in a certain sense. Cluster indices are obtained by applying the appropriate functional H to the cluster. The functionals are defined on $(\mathbb{R}^d)^{\mathbb{Z}}$, the space of $\mathbb{R}^d$-valued sequences, and are such that their values do not depend on coordinates that are equal to zero. More precisely, for $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z}$, we denote $\boldsymbol{X}_{i,j} = (\boldsymbol{X}_i, \ldots, \boldsymbol{X}_j) \in (\mathbb{R}^d)^{(j-i+1)}$. Then, we identify $H(\boldsymbol{X}_{i,j})$ with $H((\mathbf{0}, \boldsymbol{X}_{i,j}, \mathbf{0}))$, where $\mathbf{0} \in (\mathbb{R}^d)^{\mathbb{Z}}$ is the zero sequence. Such functionals H will be called *cluster functionals*.

Let $|\cdot|$ be an arbitrary norm on $\mathbb{R}^d$ and $\{u_n\}$, $\{r_n\}$ be such that

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} r_n = \lim_{n\to\infty} n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = \infty \;,$$
$$\lim_{n\to\infty} r_n/n = \lim_{n\to\infty} r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0 \;. \qquad (\mathcal{R}(r_n, u_n))$$

The sequence u_n will play a role of a threshold (that is, only the values of $|\boldsymbol{X}_t|$ that exceed u_n play a role), while r_n will be a block size in the formulation of the runs estimators. Given a cluster functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, we want to estimate the limiting quantity

$$\boldsymbol{\nu}^*(H) = \lim_{n\to\infty} \boldsymbol{\nu}^*_{n,r_n}(H) = \lim_{n\to\infty} \frac{\mathbb{E}[H(\boldsymbol{X}_{1,r_n}/u_n)]}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \;. \tag{7.1.2}$$

To guarantee existence of the limit we will require additional anticlustering assumptions on the time series $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$. For $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ define $\boldsymbol{x}^* = \sup_{j\in\mathbb{Z}} |\boldsymbol{x}_j|$. The cluster indices of interest are, among others:

- the extremal index obtained with $H_1(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x}^* > 1\}$, $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$;
- the cluster size distribution obtained with

$$H_{2,m}(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\} = m\right\} \;, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}} \;, m \in \mathbb{N} \;; \tag{7.1.3}$$

- the stop-loss index of a univariate time series obtained with

$$H_{3,\eta}(\boldsymbol{x}) = \mathbb{1}\left\{\sum_{j\in\mathbb{Z}} (x_j - 1)_+ > \eta\right\} \;, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \;, \quad \eta > 0 \;; \tag{7.1.4}$$

- the large deviation index of a univariate time series obtained with

$$H_4(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\} \;, \quad K(\boldsymbol{x}) = \left(\sum_{j\in\mathbb{Z}} x_j\right)_+ \;, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \;; \tag{7.1.5}$$

- the ruin index of a univariate time series obtained with

$$H_5(\boldsymbol{x}) = \mathbb{1}\{K(\boldsymbol{x}) > 1\} \;, \quad K(\boldsymbol{x}) = \sup_{i\in\mathbb{Z}} \left(\sum_{j\le i} x_j\right)_+ \;, \quad \boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in \mathbb{R}^{\mathbb{Z}} \;. \tag{7.1.6}$$

As indicated above, the extremal index is the classical quantity that arises in the extreme value theory for dependent sequences. Similarly, the cluster size distribution has been studied in [Hsi91] and [DR10]. The large deviation index was studied under the name *cluster index* in [MW13, MW14]. It quantifies the effect of dependence in large deviations results.

Several methods of estimation of the limit $\boldsymbol{\nu}^*(H)$ in (7.1.2) may be employed. The natural one is to consider a statistics based on disjoint blocks of size r_n, cf. [DR10] and [KS20],

$$\widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1, ir_n}/u_n) \;,$$

where $m_n = [n/r_n]$ is the number of disjoint blocks. The data-based estimator is constructed as follows. Let $k_n \to \infty$ be a sequence of integers and define u_n by $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Let $|\boldsymbol{X}|_{(n:1)} \le \cdots \le |\boldsymbol{X}|_{(n:n)}$ be order statistics from $|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_n|$. Define

$$\widehat{\boldsymbol{\nu}}^*_{n,r_n}(H) := \frac{1}{k_n} \sum_{i=1}^{m_n} H(\boldsymbol{X}_{(i-1)r_n+1, ir_n}/|\boldsymbol{X}|_{(n:n-k_n)}) \; . \tag{7.1.7}$$

The general asymptotic theory for disjoint blocks estimators was developed in [DR10]. See also [KS20, Chapter 10]. The limiting variance of the disjoint blocks estimator can be represented as

$$\boldsymbol{\nu}^*(\{H - \boldsymbol{\nu}^*(H)\mathcal{E}\}^2) \; , \tag{7.1.8}$$

where $\mathcal{E}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$. This result was established (implicitly) in [DR10], but the form of the limiting variance is again given in [KS20, Chapter 10].

Another approach to estimation of $\boldsymbol{\nu}^*(H)$ is to consider the sliding blocks statistics

$$\widetilde{\boldsymbol{\mu}}^*_{n,r_n}(H) := \frac{1}{q_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/u_n\right) \tag{7.1.9}$$

and the corresponding estimator defined in terms of order statistics:

$$\widehat{\boldsymbol{\mu}}^*_{n,r_n}(H) = \frac{1}{r_n k_n} \sum_{i=0}^{q_n-1} H\left(\boldsymbol{X}_{i+1,i+r_n}/|\boldsymbol{X}|_{(n:n-k_n)}\right) \; . \tag{7.1.10}$$

Here, $q_n = n - r_n - 1$ is the number of sliding blocks. In [DN21] the authors used the framework of [DR10] and showed that the limiting variance of the sliding blocks estimator never exceeds that of the disjoint blocks estimator. In case of the extremal index, both variances were proven to be equal. In [CK21] it was shown that the limiting variances for both disjoint and sliding blocks estimators agree and are given by the expression in (7.1.8) for an arbitrary choice of H. We note at this point that the methodology used in [DR10, DN21, KS20, CK21] fits into Peak Over Threshold (PoT) framework. On the other hand, in the Block Maxima (BM) framework, sliding blocks estimators yield typically smaller variance; see [BS18b, BS18a]. As of this moment, there is no thorough explanation of these phenomena and no formal comparison between PoT and BM framework. See [FdH15] for some partial results and [BZ21] for a recent review.

In the present paper we are interested in the so-called *runs estimators*. In the context of the extremal index, this approach goes back to [WN98] and stems from the following representation of the extremal index:

$$\theta = \lim_{n\to\infty} \mathbb{P}(\max\{|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_{r_n}|\} \le u_n \mid |\boldsymbol{X}_0| > u_n) \; . \tag{7.1.11}$$

We note that

$$\begin{aligned}
&\mathbb{P}(\max\{|\boldsymbol{X}_1|, \ldots, |\boldsymbol{X}_{r_n}|\} \le u_n \mid |\boldsymbol{X}_0| > u_n) \\
&= \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}[\mathbb{1}\{\mathcal{C}((\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{r_n})/u_n) = 0\}\mathbb{1}\{|\boldsymbol{X}_0| > u_n\}] \; ,
\end{aligned}$$

where

$$\mathcal{C}(\boldsymbol{x}) = \sup\{j : |\boldsymbol{x}_j| > 1\}$$

gives the position of the last exceedence above 1 in a particular block. Recall again the convention $\mathcal{C}((\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{r_n})/u_n) = \mathcal{C}((\boldsymbol{0}, \boldsymbol{X}_0, \ldots, \boldsymbol{X}_{r_n})/u_n)$. The latter expression provides an alternative representation for the extremal index. Then, $\mathcal{C}$ is an example of so-called *anchoring map*. Special cases of anchoring maps were considered in [Has18] and [BP18], while in [KS20] their connection to cluster indices $\boldsymbol{\nu}^*(H)$ was thoroughly investigated. It turns out that with an arbitrary choice of the anchoring map $\mathcal{C}$ we have

$$\boldsymbol{\nu}^*(H) = \mathbb{E}[H^{\mathcal{C}}(\boldsymbol{Y})] ,$$

where

$$H^{\mathcal{C}}(\boldsymbol{x}) = H(\boldsymbol{x})\mathbb{1}\{\mathcal{C}(\boldsymbol{x}) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > 1\}$$

and $\boldsymbol{Y}$ is the tail process, a distributional limit (as $x \to \infty$) of $\boldsymbol{X}/x$ given $|\boldsymbol{X}_0| > x$; see Section 7.2.3 for the precise definition.

This motivates the following runs statistics:

$$\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=r_n+1}^{n-r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) . \tag{7.1.12}$$

Indeed, under the appropriate conditions, Proposition 7.2.7 gives

$$\lim_{n\to\infty} \mathbb{E}[\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})] = \mathbb{E}[H^{\mathcal{C}}(\boldsymbol{Y})] = \boldsymbol{\nu}^*(H) .$$

The data-based runs estimator is then

$$\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{k_n} \sum_{i=r_n+1}^{n-r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/|\boldsymbol{X}|_{(n:n-k)}\right) .$$

The main result of this paper is Theorem 7.3.6, the asymptotic normality of the appropriately normalized estimator $\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$. We show, in particular, that the limiting variance agrees with the one for the disjoint blocks and sliding blocks estimators; cf. [DR10], [KS20, Chapter 10], [CK21]. This is also in accordance with the result of [DN21] for the extremal index.

Furthermore, we prove that we cannot achieve variance reduction by considering a linear combination of runs estimators with a different choice of anchoring maps $\mathcal{C}$ and $\tilde{\mathcal{C}}$. Indeed, it turns out that $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$ and $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\tilde{\mathcal{C}}})$ are totally dependent in the limit. We note in passing that even though general ideas of proofs are similar to those of [CK21], however, technicalities are significantly different. Differences stem primarily from conditioning on $\{|\boldsymbol{X}_j| > u_n\}$ used in case of the runs estimators.

Thus, from the theoretical point of view the limiting behaviour of all (disjoint blocks, sliding blocks, runs) estimators is the same. However, for finite samples a bias has to be taken into account. We note first that the theoretical finite-sample bias for both disjoint and sliding blocks estimators is the same. This can be also seen in extensive simulation studies in [CK21]. On the other hand, we were not able to get an useful formula for the bias in the runs estimator case. As such we relied on simulations. It turns out that runs estimator are typically heavily biased when estimation of the extremal index is concerned. However, the runs estimators may have an advantage when other cluster indices are considered.

The paper is structured as follows. Section 7.2 contains definitions, notation and preliminary results on convergence of clusters. It is primarily based on [KS20, Chapters 5 and 6], with some results from [BS09], [BPS18], [PS18]. Section 7.3 defines runs pseudo-estimators and estimators. The main result of the paper is the central limit theorem for runs estimators in Theorem 7.3.6. We note again that the limiting variance agrees with the one for disjoint and sliding blocks estimators. A small simulation study is performed in Section 7.4, while the extended analysis can be found in the supplementary file. All the proofs are contained in Section 7.5.

7.2. Preliminaries

In this section we fix the notation and introduce the relevant classes of functions. In Section 7.2.3 we recall the notion of the tail and the spectral tail process (cf. [BS09]). Section 7.2.4 introduces anchoring maps (cf. [BP18], [Has18]). In Section 7.2.5 we define cluster indices. We refer to [KS20, Chapter 5] for more details. In Section 7.2.6 we discuss convergence of the cluster measure, following [KS20, Chapter 6].

The most important conclusion of these preliminaries is a representation of the cluster index $\boldsymbol{\nu}^*(H)$ (cf. (7.1.2)) as $\mathbb{E}[H^{\mathcal{C}}(\boldsymbol{Y})]$, with $H^{\mathcal{C}}$ defined in (7.2.7) and $\boldsymbol{Y}$ being the tail process. Also, Proposition 7.2.7 on conditional weak convergence and Proposition 6.3.9 on unconditional weak convergence play a central role in the rest of the paper.

7.2.1. Notation

Let $|\cdot|$ be a norm on $\mathbb{R}^d$. For a sequence $\boldsymbol{x} = \{\boldsymbol{x}_j, j \in \mathbb{Z}\} \in (\mathbb{R}^d)^{\mathbb{Z}}$ and $i \le j \in \mathbb{Z} \cup \{-\infty, \infty\}$ we denote $\boldsymbol{x}_{i,j} = (\boldsymbol{x}_i, \dots, \boldsymbol{x}_j) \in (\mathbb{R}^d)^{j-i+1}$, $\boldsymbol{x}^*_{i,j} = \max_{i \le l \le j} |\boldsymbol{x}_l|$ and $\boldsymbol{x}^* = \sup_{j \in \mathbb{Z}} |\boldsymbol{x}_j|$. By $\boldsymbol{0}$ we denote the zero sequence; its dimension can be different in each of its occurrences.

By $\ell_0(\mathbb{R}^d)$ we denote the set of $\mathbb{R}^d$-valued sequences which tend to zero at infinity. Likewise, $\ell_1(\mathbb{R}^d)$ consists of sequences such that $\sum_{j \in \mathbb{Z}} |\boldsymbol{x}_j| < \infty$.

7.2.2. Classes of functions

Functionals H are defined on $\ell_0(\mathbb{R}^d)$ with the convention $H(\boldsymbol{x}_{i,j}) = H((\boldsymbol{0}, \boldsymbol{x}_{i,j}, \boldsymbol{0}))$. For $s > 0$, the function $H_s : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ is defined by $H_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)$. We consider the following classes:

- $\mathcal{L}$ is the class of bounded real-valued functions defined on $(\mathbb{R}^d)^{\mathbb{Z}}$ that are either Lipschitz continuous with respect to the uniform norm or almost surely continuous with respect to the distribution of the tail process $\boldsymbol{Y}$. This class includes functions like $\mathbb{1}\{\boldsymbol{x}^* > 1\}$, $\mathbb{1}\left\{\sum_{j \in \mathbb{Z}} |\boldsymbol{x}_j| > 1\right\}$. See Remark 6.1.6 in [KS20].
- $\mathcal{A} \subset \mathcal{L}$ is the class of shift-invariant functionals with support separated from $\boldsymbol{0}$. In particular, for $H \in \mathcal{A}$, $H(\boldsymbol{0}) = 0$. The class $\mathcal{A}$ includes $\mathbb{1}\{\boldsymbol{x}^* > 1\}$.
- $\mathcal{K}$ is the class of shift-invariant functionals $K : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ defined on $\ell_1(\mathbb{R}^d)$ such that $K(\boldsymbol{0}) = 0$ and which are Lipschitz continuous with constant L_K, i.e.
$$|K(\boldsymbol{x}) - K(\boldsymbol{y})| \le L_K \sum_{j \in \mathbb{Z}} \left|\boldsymbol{x}_j - \boldsymbol{y}_j\right| \ , \quad \boldsymbol{x}, \boldsymbol{y} \in \ell_1(\mathbb{R}^d) \ .$$
- $\mathcal{B} \subset \mathcal{L}$ is the class of functionals H of the form $H = \mathbb{1}\{K > 1\}$, where $K \in \mathcal{K}$. Functionals in $\mathcal{B}$ may have support which is not separated from $\boldsymbol{0}$. The typical example is $H(\boldsymbol{x}) = \mathbb{1}\left\{\sum_j |\boldsymbol{x}_j| > 1\right\}$; note that $H \notin \mathcal{A}$.

We will also need the map $\mathcal{E}$ is defined on $\ell_0(\mathbb{R}^d)$ by $\mathcal{E}(\boldsymbol{x}) = \sum_{j \in \mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$. Note that $\mathcal{E}$ is shift-invariant, with the support separated from zero, but is not bounded.

7.2.3. Tail and spectral tail process

Let $\boldsymbol{X} = \{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying time series with values in $\mathbb{R}^d$ and tail index α. In particular,
$$\lim_{x \to \infty} \frac{\mathbb{P}(|\boldsymbol{X}_0| > tx)}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = t^{-\alpha}$$

for all $t > 0$. Then, there exists a sequence $\boldsymbol{Y} = \{\boldsymbol{Y}_j, j \in \mathbb{Z}\}$ such that

$$\mathbb{P}(x^{-1}(\boldsymbol{X}_i, \ldots, \boldsymbol{X}_j) \in \cdot \mid |\boldsymbol{X}_0| > x) \text{ converges weakly to } \mathbb{P}((\boldsymbol{Y}_i, \ldots, \boldsymbol{Y}_j) \in \cdot)$$

as $x \to \infty$ for all $i \leq j \in \mathbb{Z}$. We call $\boldsymbol{Y}$ the tail process. See [BS09]. We note that, in particular, $|\boldsymbol{Y}_0|$ has Pareto distribution with the density $\alpha x^{-\alpha-1}$, $x > 1$. As such, it follows automatically that $\boldsymbol{Y}^* = \sup_{j\in\mathbb{Z}} |\boldsymbol{Y}_j| > 1$. Equivalently, viewing $\boldsymbol{X}$ and $\boldsymbol{Y}$ as random elements with values in $(\mathbb{R}^d)^{\mathbb{Z}}$, we have for every bounded or non-negative functional H on $(\mathbb{R}^d)^{\mathbb{Z}}$, continuous with respect to the product topology,

$$\lim_{x\to\infty} \frac{\mathbb{E}[H(x^{-1}\boldsymbol{X})\mathbb{1}\{|\boldsymbol{X}_0| > x\}]}{\mathbb{P}(|\boldsymbol{X}_0| > x)} = \mathbb{E}[H(\boldsymbol{Y})] \ .$$

The spectral tail process $\{\boldsymbol{\Theta}_j, j \in \mathbb{Z}\}$ is defined by $\boldsymbol{\Theta} = |\boldsymbol{Y}_0|^{-1}\,\boldsymbol{Y}$ and is independent of the tail process $\boldsymbol{Y}$.

7.2.4. Anchoring maps

Definition 7.2.1 (Anchoring map)**.** *A measurable map $\mathcal{C} : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{Z} \cup \{-\infty, \infty\}$ is called an anchoring map if the following two properties hold:*

An(i)*: $\mathcal{C}(\boldsymbol{x}) = j$ implies $|\boldsymbol{x}_j| \geq |\boldsymbol{x}_0| \wedge 1$;*
An(ii)*: $\mathcal{C}(B\boldsymbol{x}) = \mathcal{C}(\boldsymbol{x}) + 1$, where B is a backsift operator.*

Three basic examples of anchoring maps are:

- The infargmax functional: $\mathcal{C}^{(0)}(\boldsymbol{y}) = \inf\{j : \boldsymbol{y}^*_{-\infty,j} = \boldsymbol{y}^*\}$;
- The first exceedence above one: $\mathcal{C}^{(1)}(\boldsymbol{x}) = \inf\{j : |\boldsymbol{x}_j| > 1\}$;
- The last exceedence above one: $\mathcal{C}^{(2)}(\boldsymbol{x}) = \sup\{j : |\boldsymbol{x}_j| > 1\}$.

In what follows we use the convention $\inf \emptyset = +\infty$. We note that $\mathcal{C}^{(0)}$ is 0-homogeneous, that is, $\mathcal{C}^{(0)}_s(\boldsymbol{x}) = \mathcal{C}^{(0)}(\boldsymbol{x})$ for all $s > 0$, while $\mathcal{C}^{(1)}_s(\boldsymbol{x}) = \mathcal{C}^{(1)}(\boldsymbol{x}/s)$ and $\mathcal{C}^{(2)}_s(\boldsymbol{x}) = \mathcal{C}^{(2)}(\boldsymbol{x}/s)$ are increasing and decreasing in s, respectively, but they are not 0-homogeneous. This will play a role in the proofs.

A special importance is given to the time index 0. In particular,

- If $\mathcal{C}^{(0)}(\boldsymbol{x}) = 0$, then $\boldsymbol{x}^*_{-\infty,-1} < |\boldsymbol{x}_0|$ and $\boldsymbol{x}^*_{1,\infty} \leq |\boldsymbol{x}_0|$;
- If $\mathcal{C}^{(1)}(\boldsymbol{x}) = 0$, then $\boldsymbol{x}^*_{-\infty,-1} \leq 1$ and $|\boldsymbol{x}_0| > 1$;
- If $\mathcal{C}^{(2)}(\boldsymbol{x}) = 0$, then $\boldsymbol{x}^*_{1,\infty} \leq 1$ and $|\boldsymbol{x}_0| > 1$.

Applying an anchoring map to a finite block, say $\boldsymbol{x}_{-r,r}$ with $r \in \mathbb{N}$, is equivalent to applying it $\boldsymbol{x} = (\boldsymbol{0}, \boldsymbol{x}_{-r,r}, \boldsymbol{0})$. For example, $\mathcal{C}^{(0)}(\boldsymbol{x}_{-r,r}) = 0$ means that $\boldsymbol{x}^*_{-r,-1} < |\boldsymbol{x}_0|$ and $\boldsymbol{x}^*_{1,r} \leq |\boldsymbol{x}_0|$. This in turn implies also that $\mathcal{C}^{(0)}(\boldsymbol{x}_{-s,s}) = 0$ for $0 < s < r$. Similarly,

$$\mathcal{C}(\boldsymbol{x}_{-r,r}) = 0 \Rightarrow \mathcal{C}(\boldsymbol{x}_{-s,s}) = 0 \ , \quad 0 < s < r \tag{7.2.1}$$

for $\mathcal{C} = \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. However, we do not know if this property (used explicitly in the proofs) holds for any anchoring map. Furthermore, there are crucial monotonicity and homogeneity properties, indicated above, used explicitly in the proofs (of tightness). As such, in the paper we focus on the three anchoring maps introduced above. We will clearly indicate which results hold for an arbitrary anchoring map, and which are specific to any of $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$.

Furthermore, note that An(ii) gives that

$$\mathcal{C}(\boldsymbol{x}_{h-r,h+r}) = h \Leftrightarrow \mathcal{C}(B^{-h}\boldsymbol{x}_{-r,r}) = 0 \ . \tag{7.2.2}$$

Indeed, consider for example $\mathcal{C}^{(0)}$. Then $\mathcal{C}^{(0)}(\boldsymbol{x}_{h-r,h+r}) = h$ means that $\boldsymbol{x}^*_{h-r,h-1} < |\boldsymbol{x}_h|$ and $\boldsymbol{x}_{h+1,h+r} \le |\boldsymbol{x}_h|$. Set $\tilde{\boldsymbol{x}} = B^{-h}\boldsymbol{x}$, so that $\tilde{\boldsymbol{x}}_{-r} = \boldsymbol{x}_{h-r}$, $\tilde{\boldsymbol{x}}_0 = \boldsymbol{x}_h$ and $\tilde{\boldsymbol{x}}_r = \boldsymbol{x}_{h+r}$. Thus, $\tilde{\boldsymbol{x}}_{-r,-1} < |\tilde{\boldsymbol{x}}_0|$ and $\tilde{\boldsymbol{x}}_{1,r} \le |\tilde{\boldsymbol{x}}_0|$. This in turn means that $\mathcal{C}^{(0)}(\tilde{\boldsymbol{x}}_{-r,r}) = \mathcal{C}^{(0)}(B^{-h}\boldsymbol{x}_{-r,r}) = 0$. The similar argument applies to the other two anchoring maps.

7.2.5. Cluster measure and cluster indices

Let $\mathcal{C}$ be an anchoring map. If $\mathbb{P}(\mathcal{C}(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$ then we can define

$$\vartheta = \mathbb{P}(\mathcal{C}(\boldsymbol{Y}) = 0) \ . \tag{7.2.3}$$

We want to emphasize that ϑ does not depend on the choice of the anchoring map (see [PS18] and [KS20, Corollary 5.5.4]). In particular,

$$\vartheta = \mathbb{P}(\mathcal{C}^{(1)}(\boldsymbol{Y}) = 0) = \mathbb{P}\left(\sup_{j \le -1} |\boldsymbol{Y}_j| \le 1\right) = \mathbb{P}(\mathcal{C}^{(2)}(\boldsymbol{Y}) = 0) = \mathbb{P}\left(\sup_{j \ge 1} |\boldsymbol{Y}_j| \le 1\right) .$$

The above identity follows from the time-change formula, see [CK21, Section 7.1]. Therefore, ϑ can be recognized as the (candidate) extremal index. It becomes the usual extremal index under additional mixing and anticlustering conditions (cf. Section 7.5 in [KS20]).

Recall that $\mathcal{E}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$. The property An(i) of the anchoring maps implies

$$\sum_{h\in\mathbb{Z}} \mathbb{P}(\mathcal{C}(\boldsymbol{Y}) = h) \le \sum_{h\in\mathbb{Z}} \mathbb{P}(|\boldsymbol{Y}_h| > 1) \ . \tag{7.2.4}$$

By [KS20, Lemma 9.2.3] the latter series is finite if an appropriate anticlustering condition holds (see $\mathcal{S}(r_n, u_n)$ to be introduced later on).

Definition 7.2.2 (Cluster measure)**.** *Let $\boldsymbol{Y}$ and $\boldsymbol{\Theta}$ be the tail process and the spectral tail process, respectively, such that $\mathbb{P}(\lim_{|j|\to\infty} \boldsymbol{Y}_j = \boldsymbol{0}) = 1$. The cluster measure is the measure $\boldsymbol{\nu}^*$ on $\ell_0(\mathbb{R}^d)$ defined by*

$$\boldsymbol{\nu}^* = \vartheta \int_0^\infty \mathbb{E}[\delta_{r\boldsymbol{\Theta}} \mathbb{1}\{\mathcal{C}(\boldsymbol{\Theta}) = 0\}] \alpha r^{-\alpha-1} \mathrm{d}r \ , \tag{7.2.5}$$

where δ is the Dirac measure.

The measure $\boldsymbol{\nu}^*$ is boundedly finite on $(\mathbb{R}^d)^{\mathbb{Z}} \setminus \{\boldsymbol{0}\}$, puts no mass at $\boldsymbol{0}$ and is α-homogeneous.

Furthermore, the cluster measure can be expressed in terms of another sequence.

Definition 7.2.3. *Assume that $\mathbb{P}(\mathcal{C}(\boldsymbol{Y}) \notin \mathbb{Z}) = 0$. The conditional spectral tail process $\boldsymbol{Q}$ is a random sequence with the distribution of $(\boldsymbol{Y}^*)^{-1}\boldsymbol{Y}$ conditionally on $\mathcal{C}(\boldsymbol{Y}) = 0$.*

The sequence $\boldsymbol{Q}$ appeared implicitly in the seminal paper [DH95]. See also [BS09], [PS18, Definition 3.5] and [KS20, Chapter 5]. An abstract setting is considered in [DHS18].

Note for example that $\mathcal{C}^{(0)}(\boldsymbol{Y}) = 0$ gives $\boldsymbol{Y}^* = |\boldsymbol{Y}_0|$. Thus, (7.2.5) and the definition of $\boldsymbol{Q}$ give for a bounded or non-negative measurable function H on $\ell_0(\mathbb{R}^d)$ (see Definition 5.4.11 in [KS20]),

$$\boldsymbol{\nu}^*(H) = \vartheta \int_0^\infty \mathbb{E}[H(r\boldsymbol{Q})] \alpha r^{-\alpha-1} \mathrm{d}r \ .$$

If moreover H is such that $H(\boldsymbol{y}) = 0$ if $\boldsymbol{y}^* \le \epsilon$ for one $\epsilon > 0$, then

$$\boldsymbol{\nu}^*(H) = \epsilon^{-\alpha} \mathbb{E}[H(\epsilon\boldsymbol{Y}) \mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}] \ . \tag{7.2.6}$$

For a shift-invariant $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ and an anchoring map $\mathcal{C}$ define

$$H^{\mathcal{C}}(\boldsymbol{x}) = H(\boldsymbol{x})\mathbb{1}\{\mathcal{C}(\boldsymbol{x}) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > 1\} . \tag{7.2.7}$$

Thus, since $|\boldsymbol{Y}_0| > 1$, if H is such that $H(\boldsymbol{y}) = 0$ whenever $\boldsymbol{y}^* \le 1$, then (7.2.6) gives

$$\boldsymbol{\nu}^*(H) = \mathbb{E}[H^{\mathcal{C}}(\boldsymbol{Y})] . \tag{7.2.8}$$

Note that the $\boldsymbol{\nu}^*(H)$ does not agree with $\mathbb{E}[H(\boldsymbol{Y})]$.

Definition 7.2.4 (Cluster index)**.** *We will call $\boldsymbol{\nu}^*(H)$ the cluster index associated to the functional H.*

7.2.6. Convergence of cluster measure

Define the measures $\boldsymbol{\nu}^*_{n,r_n}$, $n \ge 1$, on $\ell_0(\mathbb{R}^d)$ as follows:

$$\boldsymbol{\nu}^*_{n,r_n} = \frac{1}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[\delta_{u_n^{-1}\boldsymbol{X}_{1,r_n}}\right] .$$

We are interested in convergence of $\boldsymbol{\nu}^*_{n,r_n}$ to $\boldsymbol{\nu}^*$. The results of this section are extracted from [KS20, Chapter 6]. See also [PS18] and [BPS18].

Anticlustering conditions

For each fixed $r \in \mathbb{N}$, the distribution of $u_n^{-1}\boldsymbol{X}_{-r,r}$ conditionally on $|\boldsymbol{X}_0| > u_n$ converges weakly to the distribution of $\boldsymbol{Y}_{-r,r}$ (see Section 7.2.3). In order to let r tend to infinity, we must embed all these finite vectors into one space of sequences. By adding zeroes on each side of the vectors $u_n^{-1}\boldsymbol{X}_{-r,r}$ and $\boldsymbol{Y}_{-r,r}$ we identify them with elements of the space $\ell_0(\mathbb{R}^d)$. Then $\boldsymbol{Y}_{-r,r}$ converges (as $r \to \infty$) to $\boldsymbol{Y}$ in $\ell_0(\mathbb{R}^d)$ if (and only if) $\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)$ almost surely.

However, this is not enough for statistical purposes and we consider the following definition.

Definition 7.2.5 ([DH95], Condition 2.8)**.** *Condition $\mathcal{AC}(r_n, u_n)$ holds if for all $x, y > 0$,*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\max_{m \le |j| \le r_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) = 0 . \tag{$\mathcal{AC}(r_n, u_n)$}$$

Condition $\mathcal{AC}(r_n, u_n)$ is referred to as the anticlustering condition. It holds for i.i.d. regularly varying sequences if $\lim_{n\to\infty} r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$. Note that the latter condition is a part of the $\mathcal{R}(r_n, u_n)$ assumption. It is also fulfilled by many models, including geometrically ergodic Markov chains, short-memory linear or max-stable processes. $\mathcal{AC}(r_n, u_n)$ implies that $\boldsymbol{Y} \in \ell_0(\mathbb{R}^d)$. See [KSW19] and [KS20].

A stronger version of the anticlustering condition reads as follows.

Definition 7.2.6. *Condition $\mathcal{S}(r_n, u_n)$ holds if for all $s, t > 0$*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=m}^{r_n} \mathbb{P}(|\boldsymbol{X}_0| > u_n s, |\boldsymbol{X}_j| > u_n t) = 0 . \tag{$\mathcal{S}(r_n, u_n)$}$$

The main consequence of the anticlustering condition $\mathcal{AC}(r_n, u_n)$ is the following result.

Proposition 7.2.7 ([BS09], Proposition 4.2; [KS20], Theorem 6.1.4)**.** *Let $H \in \mathcal{L}$. If Condition $\mathcal{AC}(r_n, u_n)$ holds, then*

$$\lim_{n\to\infty} \mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{-r_n,r_n}) \mid |\boldsymbol{X}_0| > u_n] = \mathbb{E}[H(\boldsymbol{Y})] .$$

Vague convergence of cluster measure

We now state the unconditional convergence of $u_n^{-1}\boldsymbol{X}_{1,r_n}$. Contrary to Proposition 7.2.7, where an extreme value was imposed at time 0, a large value in the cluster can happen at any time. Moreover, the convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ to $\boldsymbol{\nu}^*(H)$ may hold only for shift-invariant functionals H. The following result is a re-formulation of Theorem 6.2.5 in [KS20].

Proposition 7.2.8. *Let condition $\mathcal{AC}(r_n,u_n)$ hold. For all $H\in\mathcal{A}$,*

$$\lim_{n\to\infty}\boldsymbol{\nu}^*_{n,r_n}(H)=\lim_{n\to\infty}\frac{\mathbb{E}[H(u_n^{-1}\boldsymbol{X}_{1,r_n})]}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=\boldsymbol{\nu}^*(H)\ .$$

The immediate consequence is the following limit (cf. (7.2.3)):

$$\lim_{n\to\infty}\frac{\mathbb{P}(\boldsymbol{X}^*_{1,r_n}>u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=\vartheta\ . \tag{7.2.9}$$

Indicator functionals not vanishing around zero

Proposition 7.2.8 entails convergence of $\boldsymbol{\nu}^*_{n,r_n}(H)$ for $H\in\mathcal{A}$. For functionals which are not defined on the whole space $\ell_0(\mathbb{R}^d)$ we need an additional assumption on Asymptotic Negligibility of Small Jumps.

Definition 7.2.9. *Condition $\mathrm{ANSJB}(r_n,u_n)$ holds if for all $\eta>0$,*

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{\mathbb{P}(\sum_{j=1}^{r_n}|\boldsymbol{X}_j|\mathbb{1}\{|\boldsymbol{X}_j|\le\epsilon u_n\}>\eta u_n)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=0\ . \tag{ANSJB(r_n,u_n)}$$

Proposition 7.2.10 (Theorem 6.2.16 in [KS20])**.** *Assume that $\mathcal{AC}(r_n,u_n)$ and $\mathrm{ANSJB}(r_n,u_n)$ hold. Then for $K\in\mathcal{K}$,*

$$\boldsymbol{\nu}^*(\mathbb{1}\{K>1\})=\lim_{n\to\infty}\frac{\mathbb{P}(K(\boldsymbol{X}_{1,r_n}/u_n)>1)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)}=\vartheta\int_0^\infty\mathbb{P}(K(z\boldsymbol{Q})>1)\alpha z^{-\alpha-1}\mathrm{d}z<\infty\ .$$

7.3. Central limit theorem for runs estimators

In this section we introduce and study runs estimators of cluster indices. A pseudo-estimator is defined in (7.3.4). Its limiting covariance (for different anchoring maps) is studied in Lemma 7.3.3. In particular, for two different anchoring maps, the runs statistics are totally dependent. As a consequence we cannot reduce the limiting variance for the estimation of $\boldsymbol{\nu}^*(H)$ by considering linear combinations of the runs statistics. In Lemma 7.3.4 we consider covariance between runs and disjoint blocks estimators. Again, we obtain total dependence in the limit. The main result of the paper is the central limit theorem for runs estimators; see Theorem 7.3.6. The limiting variance agrees with the one for the disjoint blocks and sliding blocks estimators.

7.3.1. Runs estimator

To introduce runs estimators recall that (cf. (7.2.2))

$$\begin{aligned}
&H^{\mathcal{C}}\left(\boldsymbol{X}_{(j-1)r_n+h,(j+1)r_n+h}/u_n\right)=H^{\mathcal{C}}(B^{-h-jr_n}\boldsymbol{X}_{-r_n,r_n}/u_n)\\
&=H(B^{-h-jr_n}\boldsymbol{X}_{-r_n,r_n}/u_n)\mathbb{1}\left\{\mathcal{C}(B^{-h-jr_n}\boldsymbol{X}_{-r_n,r_n}/u_n)=0\right\}\mathbb{1}\left\{\left|B^{-h-jr_n}\boldsymbol{X}_0\right|>u_n\right\}\\
&=H\left(\boldsymbol{X}_{(j-1)r_n+h,(j+1)r_n+h}/u_n\right)\times\\
&\qquad\mathbb{1}\left\{\mathcal{C}(\boldsymbol{X}_{(j-1)r_n+h,(j+1)r_n+h}/u_n)=h+jr_n\right\}\mathbb{1}\left\{|\boldsymbol{X}_{h+jr_n}|>u_n\right\}\ .
\end{aligned} \tag{7.3.1}$$

Set $q_n = n - r_n$ and $m_n = n/r_n$. Without loss of generality assume that m_n is an integer. Consider disjoint blocks

$$J_j := \{jr_n + 1, \ldots, (j+1)r_n\} \; , \quad j = 0, \ldots, m_n - 1 \; . \tag{7.3.2}$$

The union of these blocks gives $\{1, \ldots, n\}$. We assume we have data $\boldsymbol{X}_{1-r_n}, \ldots, \boldsymbol{X}_{n+r_n}$. For $j = 0, \ldots, m_n - 1$ define

$$\begin{aligned} H^{\mathcal{C}}_{n,j} &= \sum_{i=jr_n+1}^{(j+1)r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = \sum_{i=jr_n+1}^{(j+1)r_n} H^{\mathcal{C}}\left(B^{-i}\boldsymbol{X}_{-r_n,r_n}/u_n\right) \\ &= \sum_{i=jr_n+1}^{(j+1)r_n} H\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = i\}\mathbb{1}\{|\boldsymbol{X}_i| > u_n\} \; . \end{aligned} \tag{7.3.3}$$

Each $H^{\mathcal{C}}_{n,j}$ is a function of the block $\boldsymbol{X}_{(j-1)r_n+1,\ldots,(j+2)r_n}$ of size $3r_n$. The number j in the notation $H^{\mathcal{C}}_{n,j}$ indicates that the indicator $|\boldsymbol{X}_i| > u_n$ is applied with $i \in J_j$.

We consider a random process

$$\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \tag{7.3.4}$$

that can be decomposed as

$$\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=0}^{m_n-1} H^{\mathcal{C}}_{n,j} \; .$$

If the anticlustering condition $\mathcal{AC}(r_n, u_n)$ holds, then using stationarity, definition (7.2.7) of $H^{\mathcal{C}}$, Proposition 7.2.7 and (7.2.8) we have

$$\lim_{n\to\infty} \mathbb{E}[\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})] = \lim_{n\to\infty} \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \mathbb{E}\left[H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n)\right] = \mathbb{E}[H^{\mathcal{C}}(\boldsymbol{Y})] = \boldsymbol{\nu}^*(H) \; .$$

Now, let k_n be a sequence of integers (depending on n) such that $k_n \to \infty$ and $k_n/n \to 0$. Define u_n by $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$ and replace u_n in $H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)$ with (k_n+1)th order statistics $|\boldsymbol{X}|_{(n:n-k_n)}$ to get the runs estimator:

$$\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \frac{1}{k_n} \sum_{i=1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/|\boldsymbol{X}|_{(n:n-k_n)}\right) \; . \tag{7.3.5}$$

In what follows we will use interchangeably k_n and $n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$, whatever is more suitable.

7.3.2. Mixing assumptions

Dependence in $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ will be controlled by the β-mixing rates $\{\beta_n\}$. Recall $\mathcal{R}(r_n, u_n)$. Let $\{\ell_n\}$ be a sequence of integers such that $\lim_{n\to\infty} \ell_n = \infty$ and $\lim_{n\to\infty} \ell_n/r_n = 0$.

Definition 7.3.1. *Condition $\beta'(r_n)$ holds if:*

$$\lim_{n\to\infty} \left\{\frac{n}{r_n} + \frac{n}{k_n}\right\} \beta_{r_n} = 0 \; , \tag{7.3.6a}$$

$$\lim_{n\to\infty} \frac{1}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{i=r_n+1}^{\infty} \beta_i = \lim_{n\to\infty} \frac{n}{k_n} \sum_{i=r_n+1}^{\infty} \beta_i = 0 \; , \tag{7.3.6b}$$

$$\lim_{n\to\infty} \frac{1}{r_n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=1}^{\infty} \beta_{jr_n} = \lim_{n\to\infty} \frac{n}{r_n k_n} \sum_{j=1}^{\infty} \beta_{jr_n} = 0 \; . \tag{7.3.6c}$$

Remark 7.3.2. We note first that $\mathcal{R}(r_n, u_n)$ gives $r_n k_n/n \to 0$.

- Assume first that $\beta_j = O(j^{-\gamma})$, $\gamma > 1$. Then $\beta'(r_n)$ reduces to $n/r_n^{1+\gamma} + n/(r_n^\gamma k_n) \to 0$ and $(nr_n^{1-\gamma})/k_n \to 0$. Choose $r_n = n^{\delta_1}$, $k_n = n^{\delta_2}$, $\delta_1 < \delta_2$, $\delta_1 + \delta_2 < 1$. Then all the conditions hold if $\delta_1 > 1/(1+\gamma)$ and $\delta_2 > 1+\delta_1(1-\gamma)$. In other words, there are little restrictions on r_n and k_n if γ is big enough, that is the dependence in the time series is weak. We note further that the choice $r_n = (\log n)^\delta$, $\delta > 0$, is not allowed. Examples of processes with beta-mixing coefficients $\beta_j = O(j^{-\gamma})$ include: max-moving averages (Theorem 13.5.5 in [KS20]); infinite order moving averages with regularly varying innovations (Lemma 15.3.1 in [KS20]).
- Assume that $\beta_j = O(\exp(-\gamma j))$, $\gamma \geq 1$. Then we can choose $k_n = n^{\delta_2}$, $\delta_2 > 0$, and $r_n = (\log n)^{\delta_1}$, $\delta_1 \geq 1$. Examples of processes with beta-mixing coefficients $\beta_j = O(\exp(-\gamma j))$ include: subordinated Gaussian max-stable processes (Example 13.5.4 in [KS20]); geometrically ergodic Markov chains (Section 14.3 in [KS20]).

7.3.3. Limiting covariances

Runs statistics

The first result deals with covariance of the process $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}$ defined in (7.3.4).

Lemma 7.3.3. *Assume $\mathcal{R}(r_n, u_n)$, $\mathcal{AC}(r_n, u_n)$, $\mathcal{S}(r_n, u_n)$ and (7.3.6b) hold. Let $H, \widetilde{H} \in \mathcal{L}$ and $\mathcal{C}, \widetilde{\mathcal{C}}$ be any of the anchoring maps $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then*

$$\lim_{n\to\infty} n\mathbb{P}(|\boldsymbol{X}_0| > u_n)\mathrm{cov}\left(\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}), \widetilde{\boldsymbol{\xi}}^*_{n,r_n}(\widetilde{H}^{\widetilde{\mathcal{C}}})\right) = \boldsymbol{\nu}^*(H\widetilde{H}) . \tag{7.3.7}$$

We note that the limit does not depend on the choice of the anchoring maps. In other words, for two different anchoring maps, $\mathcal{C}$ and $\widetilde{\mathcal{C}}$, the runs statistics $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$ and $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\widetilde{\mathcal{C}}})$ are totally dependent. As a consequence we cannot reduce the limiting variance for the estimation of $\boldsymbol{\nu}^*(H)$ by considering a linear combination of $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$ and $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\widetilde{\mathcal{C}}})$.

Runs and disjoint blocks statistics

We analyse covariance between $\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$ defined in (7.3.4) and the disjoint blocks statistics

$$\frac{1}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{j=0}^{m_n-1} H\left(\boldsymbol{X}_{jr_n+1,(j+1)r_n}/u_n\right) = \widetilde{\boldsymbol{\nu}}^*_{n,r_n}(H) . \tag{7.3.8}$$

The disjoint blocks statistics are considered in [KS20, Chapter 10].

Lemma 7.3.4. *Assume $\mathcal{R}(r_n, u_n)$, $\mathcal{AC}(r_n, u_n)$, $\mathcal{S}(r_n, u_n)$ and (7.3.6c) hold. Let $H, \widetilde{H} \in \mathcal{L}$, $\widetilde{H}(\boldsymbol{0}) = 0$ and $\mathcal{C}$ be any of the anchoring maps $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then*

$$\lim_{n\to\infty} n\mathbb{P}(|\boldsymbol{X}_0| > u_n)\mathrm{cov}\left(\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}), \widetilde{\boldsymbol{\nu}}^*_{n,r_n}(\widetilde{H})\right) = \boldsymbol{\nu}^*(H\widetilde{H}) . \tag{7.3.9}$$

Again, irrespectively of the choice of the anchoring map $\mathcal{C}$, the runs and disjoint blocks statistics are totally dependent and we cannot reduce the limiting variance by considering their linear combinations.

7.3.4. Central limit theorem

Let $\mathbb{G}$ be the Gaussian process on $L^2(\boldsymbol{\nu}^*)$ with covariance

$$\mathrm{cov}(\mathbb{G}(H), \mathbb{G}(\widetilde{H})) = \boldsymbol{\nu}^*(H\widetilde{H}) .$$

Recall that for a functional $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}_+$ and $s > 0$ we define $H_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)$. Also, recall that $\mathcal{E}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z}} \mathbb{1}\{|\boldsymbol{x}_j| > 1\}$.

Consider the class

$$\mathcal{G} = \{H_s^{\mathcal{C}}, s \in [s_0, t_0]\} = \{H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > s\}, \;\; s \in [s_0, t_0]\}.$$

We need the following assumption on its random entropy.

Assumption 7.3.5. *There exists a random metric d_n on $\mathcal{G}$ and a measurable majorant $N^*(\mathcal{G}, d_n, \epsilon)$ of the covering number $N(\mathcal{G}, d_n, \epsilon)$ such that for every sequence $\{\delta_n\}$ which decreases to zero,*

$$\int_0^{\delta_n} \sqrt{\log N^*(\mathcal{G}, d_n, \epsilon)}\mathrm{d}\epsilon \xrightarrow{\mathbb{P}} 0 \; . \tag{7.3.10}$$

The main result of this paper is Theorem 7.3.6, the asymptotic normality of the appropriately normalized estimator $\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$. The limiting variance agrees with the one for the disjoint blocks and sliding blocks estimators; cf. [DR10], [KS20, Chapter 10], [CK21].

Theorem 7.3.6. *Let $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n, u_n)$, $\beta'(r_n)$, $\mathcal{S}(r_n, u_n)$ and*

$$\lim_{n\to\infty} \frac{r_n}{\sqrt{k_n}} = \lim_{n\to\infty} \frac{r_n}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}} = 0 \tag{7.3.11}$$

hold. Suppose that Assumption 7.3.5 is satisfied. Fix $0 < s_0 < 1 < t_0 < \infty$. Assume moreover that for $\mathcal{C} = \mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$,

$$\lim_{n\to\infty} \sqrt{k_n} \sup_{s\in[s_0,t_0]} \left| \frac{\mathbb{P}(|\boldsymbol{X}_0| > u_n s)}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} - s^{-\alpha} \right| = 0 \; , \tag{7.3.12a}$$

$$\lim_{n\to\infty} \sqrt{k_n} \sup_{s\in[s_0,t_0]} |\mathbb{E}[\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H_s^{\mathcal{C}})] - \boldsymbol{\nu}^*(H_s)| = 0 \; . \tag{7.3.12b}$$

If $H \in \mathcal{A}$, then

$$\sqrt{k_n}\left\{\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) - \boldsymbol{\nu}^*(H)\right\} \xrightarrow{\mathrm{d}} \mathbb{G}(H - \boldsymbol{\nu}^*(H)\mathcal{E}) \; . \tag{7.3.13}$$

If moreover $\mathrm{ANSJB}(r_n, u_n)$ is satisfied, then (7.3.13) holds for $H \in \mathcal{B}$.

Comments on the conditions

One chooses typically $k_n = n^\epsilon$ with some $\epsilon \in (0, 1)$. We note that $\beta'(r_n)$ holds if e.g. $\beta_n = O(n^{-\delta})$ with $\delta > 1$ big enough or if β_n decays logarithmically. In the latter case, we typically choose $r_n = (\log n)^{1+\delta}$ with some $\delta > 0$. Recalling the choice of k_n we can see that (7.3.11) is not a very stringent assumption.

Furthermore, (7.3.12a) controls the bias in the tail empirical process and can be related to the classical second order assumptions.

Assumption 7.3.5 controls the size of the class $\mathcal{G}$. We are not able to provide a general set of conditions under which this condition is satisfied, however, we will verify it for virtually all functionals H that appeared in the paper. See Section 7.5.9.

7.4. Simulation study

We conducted extensive simulations in order to study the finite sample performance of the runs estimators for selected cluster indices. We compare their performance with the disjoint and sliding blocks estimators (see [CK21]). Recall that the limiting variances are the same for all estimators. We do not have theoretical formulas for bias. We note that the bias for disjoint and sliding blocks estimators is the same, but different for runs estimators. We present an extensive discussion regarding the choice of the size of the block r_n and the number of order statistics k.

We include the most important findings below. An extensive discussion, supported by a number of graphs, is included in the supplementary material.

7.4.1. Stationary AR process

We start with a simple AR(1) process defined by $X_{j+1} = \rho X_j + Z_{j+1}$, where $\rho \in (0,1)$ and $\{Z_j, j \in \mathbb{Z}\}$ is a sequence of i.i.d. regularly varying random variables with tail index $\alpha > 0$. For this process we have the explicit formulas for all cluster indices. Samples of size $n = 1000$ are generated from AR(1) with $\alpha = 4$ and $\rho = 0.5, 0.9$. We perform simulations for the classical extremal index as well as for the stop-loss index.
The following parameters are used :

- Blocks sizes $r_n = 6,\ 10,\ 20,\ 30$;
- Number of order statistics $k = 6\%,\ 10\%,\ 20\%,\ 30\%$ of the sample size n.

Extremal index. For AR(1) with $\rho \in (0,1)$ the extremal index is $\theta = 1 - \rho^\alpha$; cf. [KS20, p. 396].

- The estimators for the extremal index perform the best in case of a) strong dependence; b) small number of order statistics; c) small block sizes. This is illustrated in Figure 7.1, where the following parameters are used: block size $r_n = 6$ and order statistics $k = 6\%,\ 10\%,\ 20\%,\ 30\%$ of the sample size $n = 1000$. We note that in case of strong dependence ($\rho = 0.9$), the sliding and disjoint blocks estimators perform well for small values of r_n and small values of k. For larger values of r_n and k, blocks estimators became heavily biased. Runs estimators are heavily biased irrespective of the choice of the block size and the order statistics. For weak dependence ($\rho = 0.5$), the results are bad for virtually all the estimators regardless of the considered parameters, primarily due to a strong bias. This is illustrated in Figures 1-8 in the Supplementary file.
- In this spirit, taking the "best" set of parameters suggested by the graphs, in Table 7.1 we included the results for Monte Carlo simulation for the extremal index based on disjoint blocks, sliding blocks and runs estimators with the block sizes $r_n = 5, 6, 7, 8$. We used the number of order statistics $k = 5\%$ and 10% of the sample size n. We note again that in case of strong dependence ($\rho = 0.9$), the sliding and disjoint blocks estimators outperform runs estimators for all considered parameters. For weak dependence ($\rho = 0.5$), the results are heavily biased for all considered parameters.
- In either case (weak and strong dependence) the variability of all estimators is approximately the same, supporting the theoretical findings of this paper.

Stop-loss index. For AR(1) with $\rho > 0$ the formula for the stop-loss index is given in [KS20, p. 619] :

$$\theta_{\text{stop-loss}}(S) = (1 - \rho^\alpha)\mathbb{P}\left(\sum_{j=0}^{\infty}(\rho^j Y_0 - 1)_+ > S\right) , \tag{7.4.1}$$

where Y_0 is a Pareto random variable with the parameter α.

- At the first step we used the formula (7.4.1) and performed the Monte-Carlo simulation to obtain the approximate value of the stop-loss index.

- With this in mind, we performed simulation studies for values of r_n and k defined above. As noted in [CK21], the stop-loss index estimation requires a higher number of order statistics.
- For the weak dependence ($\rho = 0.5$), we notice (see Figure 7.2) that, as opposed to the extremal index, disjoint and sliding blocks as well as the runs estimator $C^{(0)}$ perform well regardless of the choice of the size of the block (with a very good performance for a wide range of the block sizes r_n). In some cases, the estimator $C^{(0)}$ outperforms disjoint and sliding blocks. The estimators $C^{(1)}$, $C^{(2)}$ are heavily biased most of the time.
- For the strong dependence ($\rho = 0.9$), the simulation results are rather poor for all the estimators irrespective of the chosen parameters due to the bias. This may be quite intuitive, since the stop-loss functional is based on *sums* of large values. This is illustrated in Figures 9-16 in the Supplementary file.
- In this spirit, in Table 7.2 we included the results for Monte Carlo simulation for the stop-loss index based on disjoint blocks, sliding blocks and runs estimators $C^{(0)}$ with the block sizes $r_n = 6, 10, 20, 30$ (to show stability across the wide range of block sizes) and $k = 10\%, 40\%$ (to show good performance for a large number of order statistics).
- The variability of the runs estimators is not quite in line with the theoretical results of the paper.

7.4.2. Stationary ARCH process

We consider a stationary ARCH(1) process defined by $X_j = \sqrt{\beta + \lambda X_{j-1}^2} Z_j$, where $\{Z_j, j \in \mathbb{Z}\}$ are i.i.d standard normal random variables. For $\lambda = 0.9$ the extremal index is $\theta = 0.612$ (see [EKM97, p. 480]).

- Monte Carlo results are included in Table 7.3. In this case both disjoint and sliding blocks estimators yield better results as compared to runs. This is primarily due to bias.
- The variability of all estimators is approximately the same, in line with the theoretical results of the paper.
- The estimation works relatively well for a) small values of k; b) small block sizes ($r_n = 6, 10$).

7.4.3. Summary

Our findings are summarized here:

- **All estimators (blocks and runs) have virtually same variance, which is in line with the theoretical results.**
- **For the extremal index, runs estimators are inferior as compared to blocks estimators. This is primarily due to bias.**
- **For the extremal index, blocks estimators perform well in case of a) strong dependence; b) small number of order statistics; c) small block sizes.**
- **For the stop-loss index, both blocks and runs estimator $C^{(0)}$ are superior, in comparison to $C^{(1)}$, $C^{(2)}$; the latter being heavily biased in most cases.**
- **For the stop-loss index, blocks estimators and $C^{(0)}$ perform well in case of a) weak dependence; b) small number of order statistics; c) wide range of block sizes.**
- **Furthermore, it is rather not feasible that linear combinations of the blocks and runs will reduces the bias.**

7.5. Proofs

In Section 7.5.3 we prove several lemmas on conditional convergence when anchoring maps are involved. One needs to distinguish between finite blocks (when the conditional convergence follows basically from the conditional convergence to the tail process) and growing blocks (when the anticlustering condition is needed).

In Section 7.5.4 we prove the asymptotic behaviour of the covariances of runs estimators, that is we prove Lemma 7.3.3 and Lemma 7.3.4. Section 7.5.5 deals with the empirical cluster process of runs statistics. The functional central limit theorem (Theorem 7.5.6) established there yields immediately the central limit theorem for runs estimators. See Section 7.5.6. A long proof of Theorem 7.5.6 is given in Sections 7.5.7 and 7.5.8. Finally, in Section 7.5.9 we discuss the random entropy assumption.

7.5.1. Mixing

We recall the covariance inequality for bounded, beta-mixing random variables (in fact, the inequality holds for α-mixing). Let $\beta(\mathcal{F}_1, \mathcal{F}_2)$ be the β-mixing coefficient between two sigma fields. Then (cf. [Ibr62])

$$|\mathrm{cov}(H(Z_1), H(Z_2))| \leq \mathrm{cst}\ \|H\|_\infty \|\widetilde{H}\|_\infty \beta(\sigma(Z_1), \sigma(Z_2)) \,. \tag{7.5.1}$$

In (7.5.1) the constant cst does not depend on $H, \widetilde{H}$.

7.5.2. The anticlustering conditions

Lemma 7.5.1. *Assume that $\mathcal{R}(r_n, u_n)$ and $\mathcal{AC}(r_n, u_n)$ hold. If $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ is beta-mixing and $\lim_{n\to\infty} \beta_{r_n}/\mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$, then holds $\mathcal{AC}(r_n + h_n, u_n)$ with any $h_n \leq r_n/2$.*

Proof. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{m\leq|j|\leq r_n+h_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) \\
&\leq 2\mathbb{P}\left(\max_{m\leq|j|\leq r_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) \\
&\quad + 2\mathbb{P}\left(\max_{r_n+1\leq|j|\leq r_n+h_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) \,.
\end{aligned}$$

The second last term vanishes as $n \to \infty$ and then $m \to \infty$, on account of $\mathcal{AC}(r_n, u_n)$. Applying (7.5.1) to the last term we get

$$\begin{aligned}
&\mathbb{P}\left(\max_{r_n+1\leq|j|\leq r_n+h_n} |\boldsymbol{X}_j| > u_n x \mid |\boldsymbol{X}_0| > u_n y\right) \\
&\leq \mathbb{P}\left(\max_{r_n+1\leq|j|\leq r_n+h_n} |\boldsymbol{X}_j| > u_n x\right) + \frac{\beta_{r_n}}{\mathbb{P}(|\boldsymbol{X}_0| > u_n y)} \\
&= \mathbb{P}\left(\max_{1\leq j\leq 2h_n} |\boldsymbol{X}_j| > u_n x\right) + o(1) \,,
\end{aligned}$$

where the latter $o(1)$ holds when $n \to \infty$ by the assumptions. The first term in the last line vanishes as $n \to \infty$. Indeed, since $h_n \leq r_n/2$, the assumed $\mathcal{AC}(r_n, u_n)$ implies $\mathcal{AC}(2h_n, u_n)$. Then (7.2.9) gives

$$\lim_{n\to\infty} \frac{\mathbb{P}\left(\max_{1\leq j\leq 2h_n} |\boldsymbol{X}_j| > u_n x\right)}{2h_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)} = x^{-\alpha}\vartheta \in (0, \infty) \,.$$

Since $\mathcal{R}(r_n, u_n)$ holds, $\lim_{n\to\infty} r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$. This implies that

$$\lim_{n\to\infty} h_n \mathbb{P}(|\boldsymbol{X}_0| > u_n) = 0$$

and hence $\lim_{n\to\infty} \mathbb{P}\left(\max_{1\le j\le 2h_n} |\boldsymbol{X}_j| > u_n x\right) = 0$. This finished the proof. □

7.5.3. Conditional convergence

Lemma 7.5.2. *Let $\mathcal{C} = \mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then for $h \in \mathbb{Z}$,*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) = \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{h-r,h+r}) = h) .$$

Proof. We will provide the proof for $\mathcal{C}^{(0)}$ only. By the definition of the tail process

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\mathcal{C}^{(0)}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) \\
&= \lim_{n\to\infty} \mathbb{P}(\boldsymbol{X}^*_{h-r,h-1}/u_n < |\boldsymbol{X}_h|/u_n, \boldsymbol{X}^*_{h+1,h+r}/u_n \le |\boldsymbol{X}_h|/u_n \mid |\boldsymbol{X}_0| > u_n) \\
&= \mathbb{P}(\boldsymbol{Y}^*_{h-r,h-1} < |\boldsymbol{Y}_h|, \boldsymbol{Y}^*_{h+1,h+r} \le |\boldsymbol{Y}_h|) = \mathbb{P}(\mathcal{C}^{(0)}(\boldsymbol{Y}_{h-r,h+r}) = h) .
\end{aligned}$$

□

Lemma 7.5.3. *Let $\mathcal{C}, \widetilde{\mathcal{C}}$ be any of the anchoring maps $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then for $h \in \mathbb{Z}$,*

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{-r,r}/u_n) = 0, \widetilde{\mathcal{C}}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) \\
&= \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{-r,r}) = 0, \widetilde{\mathcal{C}}(\boldsymbol{Y}_{h-r,h+r}) = h) .
\end{aligned} \tag{7.5.2}$$

Proof. We verify the statement for one combination of the anchoring maps only. For $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\mathcal{C}^{(1)}(\boldsymbol{X}_{h-r,h+r}/u_n) = h, \mathcal{C}^{(2)}(\boldsymbol{X}_{-r,r}/u_n) = 0 \mid |\boldsymbol{X}_0| > u_n) \\
&= \mathbb{P}(\boldsymbol{X}^*_{h-r,h-1} \le u_n, |\boldsymbol{X}_h| > u_n, \boldsymbol{X}^*_{1,r} \le u_n, |\boldsymbol{X}_0| > u_n \mid |\boldsymbol{X}_0| > u_n) \\
&= \mathbb{P}(\boldsymbol{Y}^*_{h-r,h-1} \le 1, |\boldsymbol{Y}_h| > 1, \boldsymbol{Y}^*_{1,r} \le 1, |\boldsymbol{Y}_0| > 1) . \\
&= \mathbb{P}(\mathcal{C}^{(1)}(\boldsymbol{Y}_{h-r,h+r}) = h, \mathcal{C}^{(2)}(\boldsymbol{Y}_{-r,r}) = 0) .
\end{aligned}$$

□

Recall the definition of $H^{\mathcal{C}}$ in (7.3.1). Let $h_n \le r_n/2$ be a sequence of integers diverging to infinity. For bounded $H, \widetilde{H}$, a direct application of $\mathcal{AC}(r_n, u_n)$ gives

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n)\, \widetilde{H}^{\widetilde{\mathcal{C}}}(\boldsymbol{X}_{h_n-r_n,h_n+r_n}/u_n) \mid |\boldsymbol{X}_0| > u_n\right] \\
&\le \|H\|\|\widetilde{H}\| \lim_{n\to\infty} \mathbb{P}(|\boldsymbol{X}_{h_n}| > u_n \mid |\boldsymbol{X}_0| > u_n) = 0 .
\end{aligned}$$

Likewise, if $H, \widetilde{H}$ are bounded and $\widetilde{H} \in \mathcal{L}$ is such that $\widetilde{H}(\boldsymbol{0}) = 0$, then the Lipschitz continuity of $\widetilde{H}$, $\mathcal{AC}(r_n, u_n)$ and Lemma 7.5.1 imply

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n)\, \widetilde{H}(\boldsymbol{X}_{h_n,h_n+r_n}/u_n) \mid |\boldsymbol{X}_0| > u_n\right] \\
&\le \|H\| \lim_{n\to\infty} \mathbb{E}\left[\widetilde{H}(\boldsymbol{X}_{h_n,h_n+r_n}/u_n) - \widetilde{H}(\boldsymbol{0}) \mid |\boldsymbol{X}_0| > u_n\right] \\
&\le \|H\|\|\widetilde{H}\| \lim_{n\to\infty} \mathbb{P}\left(\boldsymbol{X}^*_{h_n,h_n+r_n} > u_n \mid |\boldsymbol{X}_0| > u_n\right) = 0 .
\end{aligned} \tag{7.5.3}$$

The statement in (7.5.3) is also valid if $\widetilde{H}\left(\boldsymbol{X}_{h_n,h_n+r_n}/u_n\right)$ is replaced with $\widetilde{H}\left(\boldsymbol{X}_{-h_n,-h_n-r_n}/u_n\right)$.

On the other hand, for fixed h we have the following lemma that extends Lemma 7.5.3 from fixed r to $r_n \to \infty$. For this, we need to assume additionally that $\mathcal{AC}(r_n, u_n)$ holds.

Lemma 7.5.4. *Assume the conditions of Lemma 7.5.1 are satisfied. Let $H, \widetilde{H} \in \mathcal{L}$ and $\mathcal{C}, \widetilde{\mathcal{C}}$ be any of the anchoring maps $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then*

$$\begin{aligned}&\lim_{n\to\infty} \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}^{\widetilde{\mathcal{C}}}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right) \mid |\boldsymbol{X}_0| > u_n\right]\\ &= \mathbb{E}\left[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\mathbb{1}\left\{\widetilde{\mathcal{C}}(\boldsymbol{Y}) = h\right\}\right] =: \mathcal{I}(H, \widetilde{H}, \mathcal{C}, \widetilde{\mathcal{C}}; h) . \qquad (7.5.4)\end{aligned}$$

Before we prove the above lemma, we make several comments. First, as a corollary we obtain

$$\lim_{n\to\infty} \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right) \mid |\boldsymbol{X}_0| > u_n\right] = \boldsymbol{\nu}^*(H) \qquad (7.5.5)$$

and

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{-r_n,r_n}/u_n) = 0 \mid |\boldsymbol{X}_0| > u_n) = \boldsymbol{\nu}^*(1) = 1 . \qquad (7.5.6)$$

Indeed, if we take $\widetilde{H} \equiv 1$, $\mathcal{C} = \widetilde{\mathcal{C}}$ and $h = 0$, then by (7.2.8),

$$\begin{aligned}\mathcal{I}(H, 1, \mathcal{C}, \mathcal{C}; 0) &= \mathbb{E}\left[H(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\right]\\ &= \mathbb{E}\left[H(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\mathbb{1}\{|\boldsymbol{Y}_0| > 1\}\right] = \boldsymbol{\nu}^*(H) .\end{aligned}$$

Since the definition of $\boldsymbol{\nu}^*$ does not depend on the anchoring map, we have $\mathcal{I}(H, \widetilde{H}, \mathcal{C}, \widetilde{\mathcal{C}}; 0) = \boldsymbol{\nu}^*(H\widetilde{H})$ for any $\mathcal{C}, \widetilde{\mathcal{C}}$. Since the value of any anchoring map is uniquely determined, we conclude immediately that $\mathcal{I}(H, \widetilde{H}, \mathcal{C}, \mathcal{C}; h) = 0$ for $h \neq 0$. Furthermore,

$$\begin{aligned}&\sum_{h\in\mathbb{Z}} \mathbb{E}\left[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\mathbb{1}\left\{\widetilde{\mathcal{C}}(\boldsymbol{Y}) = h\right\}\right]\\ &= \mathbb{E}\left[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\mathbb{1}\left\{\widetilde{\mathcal{C}}(\boldsymbol{Y}) \in \mathbb{Z}\right\}\right] = \boldsymbol{\nu}^*(H\widetilde{H}) .\end{aligned}$$

This implies that for arbitrary anchoring maps $\mathcal{C}$, $\widetilde{\mathcal{C}}$,

$$\mathcal{I}(H, \widetilde{H}, \mathcal{C}, \widetilde{\mathcal{C}}; h) = 0 , \quad h \neq 0 . \qquad (7.5.7)$$

Proof of Lemma 7.5.4. In [CK21] we proved a version of the lemma without anchoring maps included. Since $\boldsymbol{x} \to \mathbb{1}\{\mathcal{C}(\boldsymbol{x}) = 0\}$ is not Lipschitz continuous, Lemma 6.6 in [CK21] is not directly applicable. As such, we will focus on the anchoring maps only, assuming $H = \widetilde{H} \equiv 1$.

In the first step we prove that for all $h \in \mathbb{Z}$

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) = \mathbb{P}(\mathcal{C}(\boldsymbol{Y}) = h) . \qquad (7.5.8)$$

We already know that (cf. Lemma 7.5.2)

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) = \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{h-r,h+r}) = h) .$$

Since $r_n \to \infty$ and r is fixed we can assume $0 < r < r_n$. Now, for $\mathcal{C} = \mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$ the value of

$$\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}_{h-r,h+r}) = h\} - \mathbb{1}\{\mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h\}$$

is non zero if and only if $\mathbb{P}(h+r < \mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) \le h+r_n) > 0$ or $\mathbb{P}(h-r_n \le \mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) < h-r) > 0$. Indeed, take for simplicity $h=0$. If $\mathcal{C}^{(1)}(\boldsymbol{Y}_{-r,r}) = 0$ and $\mathcal{C}^{(1)}(\boldsymbol{Y}_{-r_n,r_n}) \ne 0$, then $\boldsymbol{Y}^*_{-r,-1} \le 1$, $|\boldsymbol{Y}_0| > 1$ and then $\boldsymbol{Y}^*_{-r_n,-r-1} > 1$, while $\mathcal{C}^{(1)}(\boldsymbol{Y}_{-r,r}) \ne 0$ and $\mathcal{C}^{(1)}(\boldsymbol{Y}_{-r_n,r_n}) = 0$ cannot happen. The same reasoning applied to the other anchoring maps.

Coming back to the general case of h, the first property of the anchoring map implies that $|\boldsymbol{Y}_j| > 1$ for some $j \in \{h+r+1,\dots,h+r_n\} \cup \{h-r_n,\dots,h-r-1\}$. Since we let $r, r_n \to \infty$, we can assume that $h < r$. Thus, using the property An(i) of the anchoring maps,

$$\begin{aligned}
&\lim_{r\to\infty}\lim_{n\to\infty} |\mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{h-r,h+r}) = h) - \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h)| \\
&\le \lim_{r\to\infty}\lim_{n\to\infty} \mathbb{P}\left(\max\left\{\max_{h+r\le j\le h+r_n} |\boldsymbol{Y}_j|, \max_{h-r_n\le j\le h-r} |\boldsymbol{Y}_j|\right\} > 1\right) = 0 \qquad (7.5.9)
\end{aligned}$$

since $\mathcal{AC}(r_n, u_n)$ implies $\boldsymbol{Y}_j \to \boldsymbol{0}$ almost surely as $|j| \to \infty$. Also, the vanishing property of $\boldsymbol{Y}_j$ and the property An(i) of the anchoring map imply that

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h) = \mathbb{P}(\mathcal{C}(\boldsymbol{Y}) = h) .$$

Similarly,

$$|\mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) - \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n)|$$

is non zero if and only if $\mathbb{P}(h+r < \mathcal{C}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) \le h+r_n) > 0$ or $\mathbb{P}(h-r_n \le \mathcal{C}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) < h-r) > 0$. The first property of the anchoring map implies that $|\boldsymbol{X}_j| > |\boldsymbol{X}_0| \wedge u_n$ for some $j \in \{h+r+1,\dots,h+r_n\} \cup \{h-r_n,\dots,h-r-1\}$. Again, we can assume that $h < r$. Keeping in mind the conditioning we have:

$$\begin{aligned}
&|\mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) - \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{h-r,h+r}/u_n) = h \mid |\boldsymbol{X}_0| > u_n)| \\
&\le \mathbb{P}\left(\max\left\{\max_{h+r\le j\le h+r_n} |\boldsymbol{X}_j|, \max_{h-r_n\le j\le h-r} |\boldsymbol{X}_j|\right\} > u_n \mid |\boldsymbol{X}_0| > u_n\right) . \qquad (7.5.10)
\end{aligned}$$

By Lemma 7.5.1, the latter expression vanishes by letting first $n \to \infty$ and then $r \to \infty$. This finishes the proof of (7.5.8).

Now, we will prove

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\mathcal{C}(\boldsymbol{X}_{-r_n,+r_n}/u_n) = 0, \widetilde{\mathcal{C}}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n) = h \mid |\boldsymbol{X}_0| > u_n) \\
&= \mathbb{P}(\mathcal{C}(\boldsymbol{Y}) = 0, \widetilde{\mathcal{C}}(\boldsymbol{Y}) = h) . \qquad (7.5.11)
\end{aligned}$$

In view of Lemma 7.5.3, (7.5.11) holds with r_n replaced with r. Now, the idea is to reduce the bivariate case to the univariate.

Note first that for the anchoring maps considered here, the event $A_1 := \{\mathcal{C}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h\}$ is included in $A_2 := \{\mathcal{C}(\boldsymbol{Y}_{h-r,h+r}) = h\}$. We also note that for any event B and any pair of ordered events A_1, A_2 we have

$$|\mathbb{P}(A_1 \cap B) - \mathbb{P}(A_2 \cap B)| \le |\mathbb{P}(A_1) - \mathbb{P}(A_2)| .$$

Thus, we can bound

$$\left|\mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{-r,r}) = 0, \widetilde{\mathcal{C}}(\boldsymbol{Y}_{h-r,h+r}) = h) - \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{-r_n,r_n}) = 0, \widetilde{\mathcal{C}}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h)\right| \qquad (7.5.12)$$

by

$$\begin{aligned}
&|\mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{-r,r}) = 0) - \mathbb{P}(\mathcal{C}(\boldsymbol{Y}_{-r_n,r_n}) = 0)| \\
&+ \left|\mathbb{P}(\widetilde{\mathcal{C}}(\boldsymbol{Y}_{h-r,h+r}) = h) - \mathbb{P}(\widetilde{\mathcal{C}}(\boldsymbol{Y}_{h-r_n,h+r_n}) = h)\right|
\end{aligned}$$

and we use the first step to conclude that the expression in (7.5.12) vanishes by letting first $n \to \infty$ and then $r \to \infty$.

Therefore, (7.5.9) can be extended to the bivariate case. The same argument allows to extend (7.5.10) to the bivariate case. In summary, the proof of (7.5.11) is finished. $\square$

In the next lemma, we analyse the conditional convergence for the product of $H^{\mathcal{C}}$ and $\widetilde{H}$. Its proof is almost the same as above and hence it is omitted.

Lemma 7.5.5. *Assume that $\mathcal{AC}(r_n, u_n)$ holds. Let $H, \widetilde{H} \in \mathcal{L}$, $\widetilde{H}(\mathbf{0}) = 0$ and $\mathcal{C}$ be any of the anchoring maps $\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$. Then, for $h, h' \geq 0$,*

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{h-r_n,h'+r_n}/u_n\right) \mid |\boldsymbol{X}_0| > u_n\right] \\
&= \mathbb{E}\left[H\left(\boldsymbol{Y}\right)\widetilde{H}\left(\boldsymbol{Y}\right)\mathbb{1}\{\mathcal{C}(\boldsymbol{Y}) = 0\}\right] = \boldsymbol{\nu}^*(H\widetilde{H}) \; . \qquad (7.5.13)
\end{aligned}$$

7.5.4. Limiting Covariances

The goal of this section is to prove Lemmas 7.3.3 and 7.3.4. Two situations will arise when dealing with the covariances:

- Situation 1: we will deal with $\sum_{h=-r_n}^{r_n} \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)]$, where
$$\lim_{n\to\infty} \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)] = \mathbb{E}[c_h(\boldsymbol{Y})] \; , \quad \sum_{h\in\mathbb{Z}} \mathbb{E}[|c_h(\boldsymbol{Y})|] < \infty \; .$$
We will fix an integer $r > 0$; the convergence of $\sum_{h=-r}^{r} \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)]$ to $\sum_{h=-r}^{r} \mathbb{E}[c_h(\boldsymbol{Y})]$ will follow. The reminder $\sum_{|h|>r} \mathbb{E}[c_h(\boldsymbol{Y})]$ is negligible (as $r \to \infty$) by the summability assumption, while $\sum_{h>|r_n|} \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)]$ will be treated by the anticlustering condition $\mathcal{S}(r_n, u_n)$.
- Situation 2: we will deal with $r_n^{-1} \sum_{h=1}^{r_n} \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)] = \int_0^1 g_n(\xi) \mathrm{d}\xi$, where $g_n(h) = \mathbb{E}[c_{h,n}(\boldsymbol{X}/u_n)]$ and $g_n(\xi) \to g(\xi)$ as $n \to \infty$. Bounded convergence argument will be applied.

In what follows, to shorten the notation, we set $v_n = \mathbb{P}(|\boldsymbol{X}_0| > u_n)$.

Proof of Lemma 7.3.3. Recall that

$$H_{n,j}^{\mathcal{C}} = \sum_{i=jr_n+1}^{(j+1)r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \; .$$

The covariance of the scaled statistics is

$$\begin{aligned}
n v_n \mathrm{cov}\left(\widetilde{\boldsymbol{\xi}}_{n,r_n}^*(H^{\mathcal{C}}), \widetilde{\boldsymbol{\xi}}_{n,r_n}^*(\widetilde{H}^{\widetilde{\mathcal{C}}})\right) &= \frac{1}{r_n v_n} \mathrm{cov}\left(H_{n,0}^{\mathcal{C}}, \widetilde{H}_{n,0}^{\widetilde{\mathcal{C}}}\right) \\
+ \frac{1}{r_n v_n} \sum_{j=1}^{m_n-1} \left(1 - \frac{j}{m_n}\right) &\left\{\mathrm{cov}(H_{n,0}^{\mathcal{C}}, \widetilde{H}_{n,j}^{\widetilde{\mathcal{C}}}) + \mathrm{cov}(\widetilde{H}_{n,0}^{\widetilde{\mathcal{C}}}, H_{n,j}^{\mathcal{C}})\right\} \; . \qquad (7.5.14)
\end{aligned}$$

Using $\mathcal{R}(r_n, u_n)$, we will show that $\mathrm{cov}\left(H_{n,0}^{\mathcal{C}}, \widetilde{H}_{n,0}^{\widetilde{\mathcal{C}}}\right)$ is determined by

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{r_n v_n} \mathrm{cov}\left(H_{n,0}^{\mathcal{C}}, \widetilde{H}_{n,0}^{\widetilde{\mathcal{C}}}\right) \\
&= \lim_{n\to\infty} \frac{1}{v_n} \sum_{h=-r_n}^{r_n} \left(1 - \frac{|h|}{r_n}\right) \mathbb{E}[H^{\mathcal{C}}\left(\boldsymbol{X}_{0,2r_n}/u_n\right)\widetilde{H}^{\widetilde{\mathcal{C}}}\left(\boldsymbol{X}_{h,h+2r_n}/u_n\right)] \\
&= \boldsymbol{\nu}^*(H\widetilde{H}) \; . \qquad (7.5.15)
\end{aligned}$$

We are in the Situation 1. For fixed r, using (7.5.4) and (7.5.7) we have

$$\lim_{n\to\infty}\frac{1}{v_n}\sum_{h=-r}^{r}\mathbb{E}[H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n)\,\widetilde{H}^{\widetilde{\mathcal{C}}}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n)]$$
$$=\sum_{h=-r}^{r}\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y})=0\}\mathbb{1}\left\{\widetilde{\mathcal{C}}(\boldsymbol{Y})=h\right\}]=\sum_{h=-r}^{r}\mathcal{I}(H,\widetilde{H},\mathcal{C},\widetilde{\mathcal{C}},h)=$$
$$=\mathcal{I}(H,\widetilde{H},\mathcal{C},\widetilde{\mathcal{C}},0)=\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y})=0\}]=\boldsymbol{\nu}^*(H\widetilde{H})\;. \tag{7.5.16}$$

The value above does not depend on r. Moreover,

$$\frac{1}{v_n}\sum_{r<|h|\le r_n}\mathbb{E}[H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n)\,\widetilde{H}^{\widetilde{\mathcal{C}}}(\boldsymbol{X}_{-r_n+h,r_n+h}/u_n)]$$
$$\le\|H\|\|\widetilde{H}\|\frac{1}{v_n}\sum_{r<|h|\le r_n}\mathbb{P}(|\boldsymbol{X}_0|>u_n,|\boldsymbol{X}_h|>u_n)\;. \tag{7.5.17}$$

Letting $n\to\infty$ and then $r\to\infty$, we finish the proof of (7.5.15) by applying $\mathcal{S}(r_n,u_n)$.

Now, we deal with the term in (7.5.14). For $j\ge 1$,

$$\frac{1}{r_nv_n}\left|\mathrm{cov}(H^{\mathcal{C}}_{n,0},\widetilde{H}^{\widetilde{\mathcal{C}}}_{n,j})\right|$$
$$=\frac{1}{r_nv_n}\left|\mathrm{cov}\left(\sum_{h=1}^{r_n}H^{\mathcal{C}}\left(B^{-h}\boldsymbol{X}_{-r_n,r_n}/u_n\right),\sum_{i=1}^{r_n}\widetilde{H}^{\widetilde{\mathcal{C}}}\left(B^{-i}\boldsymbol{X}_{(j-1)r_n,(j+1)r_n}/u_n\right)\right)\right|$$
$$\le\sum_{h=(j-1)r_n+1}^{jr_n}\left(\frac{h}{r_n}-(j-1)\right)|g_n(h)|+\sum_{h=jr_n+1}^{(j+1)r_n}\left((j+1)-\frac{h}{r_n}\right)|g_n(h)|$$
$$\le\sum_{h=(j-1)r_n+1}^{(j+1)r_n}|g_n(h)|=:I_j \tag{7.5.18}$$

with

$$g_n(h)=\frac{1}{v_n}\mathrm{cov}(H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}/u_n),\widetilde{H}^{\widetilde{\mathcal{C}}}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n))\;.$$

For $h>2r_n$ we have by (7.5.1),

$$|g_n(h)|\le\frac{\|H\|_\infty\|\widetilde{H}\|_\infty}{v_n}\beta_{h-2r_n}\;. \tag{7.5.19}$$

Thus,

$$\frac{1}{r_nv_n}\sum_{j=4}^{m_n-1}\left|\mathrm{cov}(H^{\mathcal{C}}_{n,0},\widetilde{H}^{\widetilde{\mathcal{C}}}_{n,j})\right|\le\frac{\|H\|_\infty\|\widetilde{H}\|_\infty}{v_n}\sum_{j=4}^{m_n-1}\sum_{h=(j-1)r_n+1}^{(j+1)r_n}\beta_{h-2r_n}$$
$$\le 2\frac{\|H\|_\infty\|\widetilde{H}\|_\infty}{v_n}\sum_{h=3r_n+1}^{\infty}\beta_{h-2r_n}=O(1)\frac{1}{v_n}\sum_{i=r_n+1}^{\infty}\beta_i=o(1)$$

by the assumption (7.3.6b).

The terms that correspond to $j=1,2,3$ in (7.5.14) have to be dealt with separately. We are again in the Situation 1. We have

$$I_1+I_2+I_3\le 2\sum_{h=1}^{4r_n}|g_n(h)|=2\left\{\sum_{h=1}^{r}+\sum_{i=r+1}^{4r_n}\right\}|g_n(h)|\;.$$

Both parts are negligible. Indeed, as in (7.5.16),

$$\begin{aligned}\lim_{n\to\infty}\sum_{h=1}^{r}|g_n(h)| &\le \lim_{n\to\infty}\frac{1}{v_n}\sum_{h=1}^{r}\mathbb{E}[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}^{\widetilde{\mathcal{C}}}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right)]\\ &= \sum_{h=1}^{r}\mathbb{E}[H(\boldsymbol{Y})\widetilde{H}(\boldsymbol{Y})\mathbb{1}\{\mathcal{C}(\boldsymbol{Y})=0\}\mathbb{1}\left\{\widetilde{\mathcal{C}}(\boldsymbol{Y})=h\right\}] = \sum_{h=1}^{r}\mathcal{I}(H,\widetilde{H},\mathcal{C},\widetilde{\mathcal{C}},h)\end{aligned}$$

and by (7.5.7) the last term vanishes.

For the term $\sum_{h=r+1}^{4r_n}$ we apply $\mathcal{S}(r_n,u_n)$; see the argument used in (7.5.17).

This finishes the proof of the lemma. □

Proof of Lemma 7.3.4. Recall that

$$\widetilde{H}_j = H\left(\boldsymbol{X}_{jr_n+1,(j+1)r_n}/u_n\right)\ .$$

Here, $\widetilde{H}_j$ is a function of the jth block $\boldsymbol{X}_{jr_n+1,(j+1)r_n}$, $j=0,\dots,m_n-1$. Since $H^{\mathcal{C}}_{n,j}$, $j=0,\dots,m_n-1$, is a function of the block $\boldsymbol{X}_{(j-1)r_n+1,\dots,(j+2)r_n}$ (recall that we assumed that we have data $\boldsymbol{X}_{1-r_n},\dots,\boldsymbol{X}_{n+r_n}$), for $|q|\ge 3$,

$$\operatorname{cov}(H^{\mathcal{C}}_{n,j},\widetilde{H}_{j+q}) \le \|H\|\|\widetilde{H}\|\beta_{(|q|-2)r_n}\ ; \tag{7.5.20}$$

cf. (7.5.1). We have

$$\begin{aligned} k_n\operatorname{cov}\left(\widetilde{\boldsymbol{\xi}}^{*}_{n,r_n}(H^{\mathcal{C}}),\widetilde{\boldsymbol{\nu}}^{*}_{n,r_n}(\widetilde{H})\right) &= \frac{1}{r_nv_n}\operatorname{cov}\left(H^{\mathcal{C}}_{n,0},\widetilde{H}_0\right)\\ &+\frac{1}{r_nv_n}\sum_{j=1}^{m_n-1}\left(1-\frac{j}{m_n}\right)\left\{\operatorname{cov}(H^{\mathcal{C}}_{n,0},\widetilde{H}_j)+\operatorname{cov}(\widetilde{H}_0,H^{\mathcal{C}}_{n,j})\right\}\ . \end{aligned}\tag{7.5.21}$$

We analyse $\operatorname{cov}\left(H^{\mathcal{C}}_{n,0},\widetilde{H}_0\right)$. We are in the Situation 2:

$$\begin{aligned} &\frac{1}{r_nv_n}\mathbb{E}\left[H^{\mathcal{C}}_{n,0}\widetilde{H}_0\right]\\ &=\frac{1}{r_nv_n}\sum_{i=1}^{r_n}\mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{1,r_n}/u_n\right)\right]\\ &=\frac{1}{r_nv_n}\sum_{i=1}^{r_n}\mathbb{E}\left[H\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}\right)=0\}\mathbb{1}\{|\boldsymbol{X}_i|>u_n\}\widetilde{H}\left(\boldsymbol{X}_{1,r_n}/u_n\right)\right]\\ &=\frac{1}{r_nv_n}\sum_{i=1}^{r_n}\mathbb{E}\left[H\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}\right)=0\}\mathbb{1}\{|\boldsymbol{X}_0|>u_n\}\widetilde{H}\left(\boldsymbol{X}_{1-i,r_n-i}/u_n\right)\right]\\ &=\frac{1}{r_n}\sum_{i=1}^{r_n}\mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{1-i,r_n-i}/u_n\right)\mid|\boldsymbol{X}_0|>u_n\right]=\int_0^1 h_{n,0}(\xi)\mathrm{d}\xi \end{aligned}$$

with

$$h_{n,0}(\xi)=\mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{1-[\xi r_n],r_n-[\xi r_n]}/u_n\right)\mid|\boldsymbol{X}_0|>u_n\right]\ ,\quad \xi\in(0,1)\ .$$

Note that the third equality follows by stationarity. By (7.5.13), for each $\xi\in(0,1)$, $h_{n,0}(\xi)\to\boldsymbol{\nu}^*(H\widetilde{H})$. Furthermore, the sequence $\{h_{n,0},n\ge 1\}$ is uniformly bounded in n and ξ. Thus, with help of $\mathcal{R}(r_n,u_n)$,

$$\lim_{n\to\infty}\frac{1}{r_nv_n}\operatorname{cov}(H^{\mathcal{C}}_{n,0},\widetilde{H}_0)=\lim_{n\to\infty}\frac{1}{r_nv_n}\mathbb{E}[H^{\mathcal{C}}_{n,0}\widetilde{H}_0]=\boldsymbol{\nu}^*(H\widetilde{H})\ .$$

The other covariances vanish. Indeed, we analyse $\mathrm{cov}(H^{\mathcal{C}}_{n,0}, \widetilde{H}_j)$, $j \geq 1$. We have, using again the stationarity as above,

$$
\begin{aligned}
&\frac{1}{r_n v_n}\mathbb{E}[H^{\mathcal{C}}_{n,0}, \widetilde{H}_j] \\
&= \frac{1}{r_n v_n}\sum_{i=1}^{r_n} \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{jr_n+1,(j+1)r_n}/u_n\right)\right] \\
&= \frac{1}{r_n v_n}\sum_{i=1}^{r_n} \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{jr_n+1-i,(j+1)r_n-i}/u_n\right)\right] \\
&= \int_0^1 h_{n,j}(\xi)\mathrm{d}\xi
\end{aligned}
$$

with a function $h_{n,j}$ defined on $(0,1)$ by

$$
h_{n,j}(\xi) = \mathbb{E}\left[H^{\mathcal{C}}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\widetilde{H}\left(\boldsymbol{X}_{jr_n-[\xi r_n]+1,(j+1)r_n-[\xi r_n]}/u_n\right) \mid |\boldsymbol{X}_0| > u_n\right] .
$$

Until now we proceeded as in the case $j = 0$ above. However, now we use (7.5.3). For each $\xi \in (0,1)$, $jr_n - [\xi r_n] \to +\infty$. Hence, $h_{n,j}(\xi) \to 0$. Bounded convergence and $\mathcal{R}(r_n, u_n)$ give

$$
\lim_{n\to\infty} \frac{1}{r_n v_n}\mathrm{cov}(H^{\mathcal{C}}_{n,0}, \widetilde{H}_j) = 0 . \tag{7.5.22}
$$

The same idea applies to $\mathrm{cov}(\widetilde{H}_0, H^{\mathcal{C}}_{n,j})$, $j \geq 1$:

$$
\lim_{n\to\infty} \frac{1}{r_n v_n}\mathrm{cov}(\widetilde{H}_0, H^{\mathcal{C}}_{n,j}) = 0 . \tag{7.5.23}
$$

Now, by (7.5.22)-(7.5.23), the terms that correspond to $j = 1, 2$ in (7.5.21) vanish, while (7.5.20) and (7.3.6c) give

$$
\begin{aligned}
&\frac{1}{r_n v_n}\sum_{j=3}^{m_n-1}\left\{\mathrm{cov}(H^{\mathcal{C}}_{n,0}, \widetilde{H}_j) + \mathrm{cov}(\widetilde{H}_0, H^{\mathcal{C}}_{n,j})\right\} \\
&= O(1)\frac{1}{r_n v_n}\sum_{j=1}^{\infty}\beta_{jr_n} = o(1) .
\end{aligned}
$$

□

7.5.5. Empirical cluster process of runs statistics

Recall that

$$
H^{\mathcal{C}}(\boldsymbol{x}) = H(\boldsymbol{x})\mathbb{1}\{\mathcal{C}(\boldsymbol{x}) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > 1\} .
$$

Define

$$
H^{\mathcal{C}}_s(\boldsymbol{x}) = H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > s\} . \tag{7.5.24}
$$

Recall that $0 < s_0 < 1 < t_0 < \infty$. Recall also that $k_n = n\mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Define also the classical tail empirical process by

$$
\mathbb{T}_n(s) = \sqrt{k_n}\left\{\frac{\sum_{j=1}^n \mathbb{1}\{|\boldsymbol{X}_j| > su_n\}}{k_n} - s^{-\alpha}\right\} , \quad s \in [s_0, t_0] .
$$

In order to deal with asymptotic normality of runs estimators, we study the empirical process

$$\begin{aligned}\mathbb{F}_n(H_s^{\mathcal{C}}) &:= \sqrt{k_n}\left\{\widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H_s^{\mathcal{C}}) - \boldsymbol{\nu}^*(H_s)\right\} \\ &= \sqrt{k_n}\left\{\frac{\sum_{i=1}^n H_s^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)}{k_n} - s^{-\alpha}\boldsymbol{\nu}^*(H)\right\} .\end{aligned}$$

The process $\mathbb{F}_n(H_s^{\mathcal{C}})$ is viewed as a random element with values in $\mathbb{D}([s_0,t_0])$. The next result is crucial to establish convergence of runs estimators.

Theorem 7.5.6. *Let $\{\boldsymbol{X}_j, j\in\mathbb{Z}\}$ be a stationary, regularly varying $\mathbb{R}^d$-valued time series. Assume that $\mathcal{R}(r_n,u_n)$, $\beta'(r_n)$, $\mathcal{S}(r_n,u_n)$, (7.3.11) and (7.3.12b) hold. Suppose that Assumption 7.3.5 is satisfied.*

Then $\mathbb{F}_n(H_\cdot^{\mathcal{C}})$ converges weakly in $(\mathbb{D}([s_0,t_0]),J_1)$ to a Gaussian process $\mathbb{G}(H_\cdot)$ with the covariance $\boldsymbol{\nu}^*(H_sH_t)$. If moreover ANSJB$(r_n,u_n)$ is satisfied, then the convergence holds for $H\in\mathcal{B}$. If additionally (7.3.12a) is satisfied, then the processes $\mathbb{F}_n(H_\cdot^{\mathcal{C}})$ and $\mathbb{T}_n(\cdot)$ converge jointly to $(\mathbb{G}(H_\cdot),\mathbb{G}(\mathcal{E}_\cdot))$.

Remark 7.5.7. The proof of Theorem 7.5.6 consists of two parts: fidi convergence (Section 7.5.7) and tightness (Section 7.5.8). We note that once the behaviour of covariances is established (see Lemmas 7.3.3 and 7.3.4) the fidi portion of the proof can be in principle omitted thanks to Theorem 2.1 and 2.3 in [DN21] (see the proof of Proposition B.4 there). This would require some modifications in the assumptions. As such, we keep the proof of fidi convergence for completeness.

7.5.6. Proof of Theorem 7.3.6

Write $\psi_n = |\boldsymbol{X}|_{(n:n-k_n)}/u_n$. Since $k_n = n\mathbb{P}(|\boldsymbol{X}_0|>u_n)$, we can rewrite $\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}})$ as $\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) = \widetilde{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}_{\psi_n})$ (cf. (7.3.4)-(7.3.5)). Therefore,

$$\sqrt{k_n}\left\{\widehat{\boldsymbol{\xi}}^*_{n,r_n}(H^{\mathcal{C}}) - \boldsymbol{\nu}^*(H)\right\} = \mathbb{F}_n(H^{\mathcal{C}}_{\psi_n}) + \sqrt{k_n}\left\{\boldsymbol{\nu}^*(H_{\psi_n}) - \boldsymbol{\nu}^*(H)\right\} .$$

We have local uniform convergence of $\{\mathbb{F}_n(H_s^{\mathcal{C}}),\ s\in[s_0,\,t_0]\}$ to a continuous Gaussian process $\mathbb{G}$ thanks to Theorem 7.5.6. Moreover, the convergence of $\{\mathbb{T}_n(\cdot),\ s\in[s_0,\,t_0]\}$ yields $\psi_n \xrightarrow{\mathrm{d}} 1$, jointly with $\mathbb{F}_n(H_s^{\mathcal{C}})$. Therefore, $\mathbb{F}_n(H^{\mathcal{C}}_{\psi_n}) \xrightarrow{\mathrm{d}} \mathbb{G}(H)$. Using Vervaat's theorem, we have, jointly with the previous convergence, $\sqrt{k_n}(\psi_n^{-\alpha}-1) \xrightarrow{\mathrm{d}} -\mathbb{G}(\mathcal{E})$. Therefore, by the homogeneity of $\boldsymbol{\nu}^*$,

$$\sqrt{k_n}\left\{\boldsymbol{\nu}^*(H_{\psi_n}) - \boldsymbol{\nu}^*(H)\right\} = \boldsymbol{\nu}^*(H)\sqrt{k_n}(\psi_n^{-\alpha}-1) \xrightarrow{\mathrm{d}} -\boldsymbol{\nu}^*(H)\mathbb{G}(\mathcal{E}).$$

Since the convergence hold jointly, we conclude the result.

7.5.7. Proof of Theorem 7.5.6 - fidi convergence

Recall the disjoint blocks of size r_n (cf. (7.3.2)):

$$J_j := \{jr_n+1,\ldots,(j+1)r_n\} ,\quad j=0,\ldots,m_n-1 .$$

These blocks were chosen to calculate the limiting covariance of the process $\mathbb{F}_n$. However, they are not appropriate for a proof of the central limit theorem. We need to introduce a large-small blocks decomposition.

For this purpose let z_n be a sequence of integers such that $z_n \to \infty$ and

$$\lim_{n\to\infty} z_n r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n) = \lim_{n\to\infty} \frac{z_n r_n}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}} = \lim_{n\to\infty} \frac{z_n r_n}{\sqrt{k_n}} = 0\;. \tag{7.5.25}$$

This is possible thanks to the assumptions $\mathcal{R}(r_n, u_n)$ and (7.3.11). We note that this assumption is needed for the Lindeberg condition only. Set

$$\widetilde{m}_n = \frac{q_n}{(z_n+3)r_n} = \frac{n-r_n}{(z_n+3)r_n} \sim \frac{n}{z_n r_n}$$

and assume for simplicity that $\widetilde{m}_n$ is an integer. Since $z_n \to \infty$, we have $\widetilde{m}_n = o(m_n)$. For $j = 1,\dots,\widetilde{m}_n$ define now large and small blocks as follows:

$$\begin{aligned}
&L_1 = \{1,\dots,z_n r_n\}\;, \quad S_1 = \{z_n r_n + 1 \dots, z_n r_n + 3r_n\}\;, \\
&L_2 = \{z_n r_n + 3r_n + 1,\dots,2z_n r_n + 3r_n\}\;, \quad S_2 = \{2z_n r_n + 3r_n + 1,\dots,2z_n r_n + 6r_n\}\;, \\
&L_j = \{(j-1)z_n r_n + 3(j-1)r_n + 1,\dots,jz_n r_n + 3(j-1)r_n\}\;, \\
&S_j = \{jz_n r_n + 3(j-1)r_n + 1,\dots,jz_n r_n + 3jr_n\}\;.
\end{aligned}$$

The block L_1 is obtained by merging z_n consecutive blocks $J_0,\dots,J_{z_n-1}$ of size r_n. Likewise, $S_1 = J_{z_n} \cup J_{z_n+1} \cup J_{z_n+2}$. Therefore, the large block of size $z_n r_n$ is followed by the small block of size $3r_n$, which in turn is followed by the large block of size $z_n r_n$ and so on. All together,

$$\bigcup_{j=1}^{\widetilde{m}_n} (L_j \cup S_j) = \{1,\dots,q_n\} = \{1,\dots,n-r_n\}\;.$$

Write

$$\begin{aligned}
\sum_{i=1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) &= \sum_{i=1}^{q_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) + \sum_{i=q_n+1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \\
&= \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(l)}(H^{\mathcal{C}}) + \sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(s)}(H^{\mathcal{C}}) + W_n\;,
\end{aligned} \tag{7.5.26}$$

where now

$$\Psi_j^{(l)}(H^{\mathcal{C}}) = \sum_{i\in L_j} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\;, \quad \Psi_j^{(s)}(H^{\mathcal{C}}) = \sum_{i\in S_j} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)$$

and

$$W_n = \sum_{i=q_n+1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = \sum_{i=n-r_n+1}^{n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\;.$$

With such the decomposition, $\boldsymbol{X}_{1-r_n},\dots,\boldsymbol{X}_{z_n r_n + r_n}$ used in the definition of $\Psi_1^{(l)}(H^{\mathcal{C}})$ are separated by at least r_n from the random variables that define $\Psi_2^{(l)}(H^{\mathcal{C}})$. The mixing condition (7.3.6a) allows us to replace $\boldsymbol{X}$ with the independent blocks process, that is, we can treat the random variables $\Psi_j^{(l)}(H^{\mathcal{C}})$, $j = 1,\dots,\widetilde{m}_n$, as independent. The same applies to $\Psi_j^{(s)}(H^{\mathcal{C}})$.

Set

$$\mathbb{Z}_n(H^{\mathcal{C}}) = \sum_{j=1}^{\widetilde{m}_n} \left\{Z_{n,j}(H^{\mathcal{C}}) - \mathbb{E}[Z_{n,j}(H^{\mathcal{C}})]\right\} =: \sum_{j=1}^{\widetilde{m}_n} \bar{Z}_{n,j}(H^{\mathcal{C}}) \tag{7.5.27}$$

with

$$Z_{n,j}(H^{\mathcal{C}}) = \frac{1}{\sqrt{k_n}} \Psi_j^{(l)}(H^{\mathcal{C}})\;. \tag{7.5.28}$$

The next steps are standard.

- First, we show that the limiting variance of the large blocks process $\mathbb{Z}_n$ is the same as that of the process $\mathbb{F}_n$;
- Next, we show that the small blocks process (the scaled second term in (7.5.26)) is negligible;
- We show that the boundary term W_n is also negligible;
- Finally, we will verify the Lindeberg condition for the large blocks process.

Again, to shorten the notation, we set $v_n = \mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Hence, $k_n = nv_n$.

Variance of the large blocks. We have (using the assumed independence of $\Psi_j^{(l)}(H^{\mathcal{C}})$)

$$\begin{aligned} \operatorname{var}\left(\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\widetilde{m}_n}\Psi_j^{(l)}(H^{\mathcal{C}})\right) &= \frac{\widetilde{m}_n}{k_n}\operatorname{var}(\Psi_1^{(l)}(H^{\mathcal{C}})) \\ \sim \frac{1}{z_nr_nv_n}\operatorname{var}\left(\sum_{i=1}^{z_nr_n}H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\right) &= \frac{1}{z_nr_nv_n}\operatorname{var}\left(\sum_{j=0}^{z_n-1}H^{\mathcal{C}}_{n,j}\right) \ , \end{aligned} \tag{7.5.29}$$

where $H^{\mathcal{C}}_{n,j}$ is defined in (7.3.3) and where in the last line we decomposed the block $L_1 = \{1,\dots,z_nr_n\}$ into z_n disjoint blocks $J_0,\dots,J_{z_n-1}$ and $\widetilde{m}_n \sim m_n/z_n$. The next steps follow easily from (7.5.14) with m_n replaced by z_n.

The term in (7.5.29) becomes

$$\frac{\operatorname{var}\left(H^{\mathcal{C}}_{n,0}\right)}{r_nv_n} + \frac{2}{r_nv_n}\sum_{j=1}^{z_n-1}\left(1-\frac{j}{z_n}\right)\operatorname{cov}(H^{\mathcal{C}}_{n,0},H^{\mathcal{C}}_{n,j}) \ . \tag{7.5.30}$$

It follows immediately from (7.5.15) that the limit of the first term above is

$$\lim_{n\to\infty}\frac{\operatorname{var}\left(H^{\mathcal{C}}_{n,0}\right)}{r_n\mathbb{P}(|\boldsymbol{X}_0|>u_n)} = \boldsymbol{\nu}^*(H^2) \ . \tag{7.5.31}$$

Now, for the second term in (7.5.30) we adapt the proof of Lemma 7.3.3 from m_n to z_n.

As in (7.5.18), for $j\geq 1$,

$$\frac{1}{r_nv_n}\left|\operatorname{cov}(H^{\mathcal{C}}_{n,0},H^{\mathcal{C}}_{n,j})\right| \leq \sum_{h=(j-1)r_n+1}^{(j+1)r_n}|g_n(h)| =: I_j$$

with (this time)

$$g_n(h) = \frac{1}{v_n}\operatorname{cov}(H^{\mathcal{C}}(\boldsymbol{X}_{-r_n,r_n}),H^{\mathcal{C}}(\boldsymbol{X}_{h-r_n,h+r_n})) \ .$$

For $h>2r_n$, similarly to (7.5.19), we have by (7.5.1),

$$|g_n(h)| \leq \frac{\|H\|_\infty^2}{v_n}\beta_{h-2r_n} \ .$$

Thus,

$$\begin{aligned} \frac{1}{r_nv_n}\sum_{j=4}^{z_n-1}|\operatorname{cov}(H^{\mathcal{C}}_{n,0},H^{\mathcal{C}}_{n,j})| &\leq \frac{\|H\|_\infty^2}{v_n}\sum_{j=4}^{z_n-1}\sum_{h=(j-1)r_n+1}^{(j+1)r_n}\beta_{h-2r_n} \\ \leq 2\frac{\|H\|_\infty^2}{v_n}\sum_{h=3r_n+1}^{\infty}\beta_{h-2r_n} &= O(1)\frac{1}{v_n}\sum_{i=r_n+1}^{\infty}\beta_i = o(1) \end{aligned}$$

by the assumption (7.3.6b). The terms that correspond to $j = 1,2,3$ in (7.5.14) are negligible.

In summary, we showed that

$$\lim_{n\to\infty} \mathrm{var}\left(\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(l)}(H^{\mathcal{C}})\right) = \boldsymbol{\nu}^*(H^2) \; .$$

Variance of the small blocks. We have (using again the assumed independence of $\Psi_j^{(s)}(H^{\mathcal{C}})$ thanks to the beta-mixing)

$$\mathrm{var}\left(\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(s)}(H^{\mathcal{C}})\right) = \frac{\widetilde{m}_n}{k_n}\mathrm{var}(\Psi_1^{(s)}(H^{\mathcal{C}})) \sim \frac{1}{z_n r_n v_n}\mathrm{var}(\Psi_1^{(s)}(H^{\mathcal{C}})) \; .$$

Since the size of $\Psi_1^{(s)}(H^{\mathcal{C}})$ is 3 times the size of $H_{n,1}^{\mathcal{C}}$ defined in (7.3.3), we have by (7.5.15)

$$\mathrm{var}\left(\frac{1}{\sqrt{k_n}}\sum_{j=1}^{\widetilde{m}_n} \Psi_j^{(s)}(H^{\mathcal{C}})\right) \sim \frac{1}{z_n r_n v_n} r_n v_n \boldsymbol{\nu}^*(H^2) = O(1/z_n) = o(1) \; .$$

Variance of the boundary term W_n. We have (cf. (7.3.3))

$$\mathrm{var}\left(\frac{1}{\sqrt{k_n}}W_n\right) = \mathrm{var}\left(\frac{1}{\sqrt{k_n}}\sum_{i=1}^{r_n} H^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\right) = \frac{\mathrm{var}(H_{n,0}^{\mathcal{C}})}{k_n} = \frac{\mathrm{var}(H_{n,0}^{\mathcal{C}})}{n v_n} \; .$$

The latter term vanishes when $n\to\infty$, using (7.5.31) and $r_n/n \to 0$.

Lindeberg condition for $\mathbb{Z}_n(H^{\mathcal{C}})$. We need to show that for all $\eta > 0$,

$$\lim_{n\to\infty} \widetilde{m}_n \mathbb{E}\left[Z_{n,1}^2(H^{\mathcal{C}})\mathbb{1}\{|Z_{n,1}(H^{\mathcal{C}})| > \eta\}\right] = 0 \; . \tag{7.5.32}$$

Since H is bounded, then by (7.5.25),

$$|Z_{n,1}(H^{\mathcal{C}})| \le \frac{\sqrt{k_n} z_n r_n}{n v_n}\|H\|_\infty \sim \frac{z_n r_n}{\sqrt{n v_n}}\|H\|_\infty = o(1) \; . \tag{7.5.33}$$

Thus, the indicator in (7.5.32) becomes zero for large n.

7.5.8. Proof of Theorem 7.5.6 - asymptotic equicontinuity

We need the following lemma which is an adapted version of Theorem 2.11.1 in [vdVW96]. Let $\mathbb{Z}_n$ be the empirical process indexed by a semi-metric space $(\mathcal{G},\rho)$, defined by

$$\mathbb{Z}_n(f) = \sum_{j=1}^{\widetilde{m}_n}\{Z_{n,j}(f) - \mathbb{E}[Z_{n,j}(f)]\} \; ,$$

where $\{Z_{n,j}, n\ge 1\}$, $j = 1,\dots,\widetilde{m}_n$, are i.i.d. separable, stochastic processes and $\widetilde{m}_n$ is a sequence of integers such that $\widetilde{m}_n \to \infty$. Define the random semi-metric d_n on $\mathcal{G}$ by

$$d_n^2(f,g) = \sum_{j=1}^{\widetilde{m}_n}\{Z_{n,j}(f) - Z_{n,j}(g)\}^2 \; , f,g\in\mathcal{G} \; .$$

Lemma 7.5.8. *Assume that $(\mathcal{G},\rho)$ is totally bounded. Assume moreover that:*

(i) For all $\eta > 0$,

$$\lim_{n\to\infty} \widetilde{m}_n \mathbb{E}[\|Z_{n,1}\|_{\mathcal{G}}^2 \mathbb{1}\{\|Z_{n,1}\|_{\mathcal{G}}^2 > \eta\}] = 0 \,. \tag{7.5.34}$$

(ii) For every sequence $\{\delta_n\}$ which decreases to zero,

$$\lim_{n\to\infty} \sup_{\substack{f,g\in\mathcal{G}\\ \rho(f,g)\le\delta_n}} \mathbb{E}[d_n^2(f,g)] = 0 \,. \tag{7.5.35}$$

(iii) There exists a measurable majorant $N^*(\mathcal{G}, d_n, \epsilon)$ of the covering number $N(\mathcal{G}, d_n, \epsilon)$ such that for every sequence $\{\delta_n\}$ which decreases to zero,

$$\int_0^{\delta_n} \sqrt{\log N^*(\mathcal{G}, d_n, \epsilon)} \mathrm{d}\epsilon \xrightarrow{\mathbb{P}} 0 \,.$$

Then $\{\mathbb{Z}_n, n \ge 1\}$ is asymptotically ρ-equicontinuous, i.e. for each $\eta > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\substack{f,g\in\mathcal{G}\\ \rho(f,g)<\delta}} |\mathbb{Z}_n(f) - \mathbb{Z}_n(g)| > \eta \right) = 0 \,.$$

Remark 7.5.9. The separability assumption is not in [vdVW96]. It implies measurability of $\|Z_{n,1}\|_{\mathcal{G}}$. Furthermore, the separability also implies that for all $\delta > 0$, $n \in \mathbb{N}$, $(e_j)_{1\le j\le \widetilde{m}_n} \in \{-1,0,1\}^{\widetilde{m}_n}$ and $i \in \{1,2\}$, the supremum

$$\sup_{\substack{f,g\in\mathcal{G}\\ \rho(f,g)<\delta}} \left| \sum_{j=1}^{\widetilde{m}_n} e_j \left(Z_{n,j}(f) - Z_{n,j}(g)\right)^i \right| = \sup_{\substack{f,g\in\mathcal{G}_0\\ \rho(f,g)<\delta}} \left| \sum_{j=1}^{\widetilde{m}_n} e_j \left(Z_{n,j}(f) - Z_{n,j}(g)\right)^i \right|$$

is measurable, which is an assumption of [vdVW96]. ⊕

Asymptotic equicontinuity of the empirical process of sliding blocks

Recall the big-blocks process $\mathbb{Z}_n(H^{\mathcal{C}})$ (cf. (7.5.27)-(7.5.28)). Recall also that thanks to the β-mixing we can consider random variables $\Psi_j^{(l)}(H^{\mathcal{C}})$, $j = 1, \ldots, \widetilde{m}_n$ to be independent. Recall that $H_s^{\mathcal{C}}$ is defined in (7.5.24). We need to prove the asymptotic equicontinuity of $\mathbb{Z}_n(H_s^{\mathcal{C}})$ indexed by the class $\mathcal{G} = \{H_s^{\mathcal{C}}, s \in [s_0, t_0]\}$ equipped with the metric $\boldsymbol{\rho}^*(H_s^{\mathcal{C}}, H_t^{\mathcal{C}}) = \boldsymbol{\nu}^*(\{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}^2)$. The same argument can be used to prove the asymptotic equicontinuity for the small blocks process. This yields asymptotic equicontinuity of $\mathbb{F}_n(H_s^{\mathcal{C}})$.

In what follows, the proof of the Lindeberg-type condition (7.5.34) is easy. The proof of (7.5.35) is quite involved.

Thanks to Assumption 7.3.5, the condition (7.3.10) is satisfied. Its validity is discussed in Section 7.5.9.

Lindeberg condition: Proof of (7.5.34). We re-write (7.5.33) as follows:

$$\sup_{s\in[s_0,t_0)} |Z_{n,1}(H_s^{\mathcal{C}})| \le \frac{\sqrt{k_n} z_n r_n}{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \|H\|_\infty \le \frac{z_n r_n}{\sqrt{n\mathbb{P}(|\boldsymbol{X}_0| > u_n)}} \sup_{s\in[s_0,t_0]} \|H_s\|_\infty \,,$$

Since the class $\{H_s : s \in [s_0, t_0]\}$ is linearly ordered, $\sup_{s\in[s_0,t_0]} \|H_s\|_\infty$ is achieved either at $s = s_0$ or $s = t_0$. Hence, the Lindeberg condition (i) of Lemma 7.5.8 holds by (7.5.32).

Asymptotic continuity of random semi-metric: Proof of (7.5.35). The proof is rather long and technical. Again, to shorten the notation we set $v_n = \mathbb{P}(|\boldsymbol{X}_0| > u_n)$.

Define the random metric

$$d_n^2(H_s^{\mathcal{C}}, H_t^{\mathcal{C}}) = \sum_{j=1}^{\widetilde{m}_n} (Z_{n,j}(H_s^{\mathcal{C}}) - Z_{n,j}(H_t^{\mathcal{C}}))^2 \;.$$

Let (cf. (7.3.3))

$$H_{s,n,j}^{\mathcal{C}} = \sum_{i=jr_n+1}^{(j+1)r_n} H\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right) \mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right) = i\} \mathbb{1}\{|\boldsymbol{X}_i| > su_n\} \;.$$

We need to evaluate $\mathbb{E}[d_n^2(H_s^{\mathcal{C}}, H_t^{\mathcal{C}})]$:

$$\begin{aligned}
&\mathbb{E}[d_n^2(H_s^{\mathcal{C}}, H_t^{\mathcal{C}})] \\
&= \frac{k_n \widetilde{m}_n}{(nv_n)^2} \mathbb{E}\left[\left(\sum_{i=1}^{z_n r_n} \left\{H_s^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - H_t^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\right\}\right)^2\right] \\
&\sim \frac{1}{z_n r_n v_n} \mathbb{E}\left[\left(\sum_{j=0}^{z_n-1} \left\{H_{s,n,j}^{\mathcal{C}} - H_{t,n,j}^{\mathcal{C}}\right\}\right)^2\right] \;, \qquad (7.5.36)
\end{aligned}$$

where in the last line we decomposed the block L_1 into z_n disjoint blocks $J_0, \ldots, J_{z_n-1}$, $\widetilde{m}_n \sim m_n/z_n$; cf. (7.5.29). The term in (7.5.36) becomes

$$\begin{aligned}
&\frac{\mathbb{E}[\left\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\right\}^2]}{r_n v_n} \\
&+ 2\frac{1}{r_n v_n} \sum_{j=1}^{z_n-1} \left(1 - \frac{j}{z_n}\right) \mathbb{E}[\left\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\right\} \left\{H_{s,n,j}^{\mathcal{C}} - H_{t,n,j}^{\mathcal{C}}\right\}]
\end{aligned}$$

The above lines correspond to (7.5.14) with m_n replaced by z_n.

We are going to prove two statements:

$$\lim_{n\to\infty} \sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \frac{\mathbb{E}[\left\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\right\}^2]}{r_n v_n} = 0 \qquad (7.5.37)$$

and

$$\lim_{n\to\infty} \sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \frac{1}{r_n v_n} \sum_{j=1}^{z_n-1} \left(1 - \frac{j}{z_n}\right) \mathbb{E}[\left\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\right\} \left\{H_{s,n,j}^{\mathcal{C}} - H_{t,n,j}^{\mathcal{C}}\right\}] = 0 \;. \qquad (7.5.38)$$

Proof of (7.5.37). We will write $\{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}(\boldsymbol{x})$ for $H_s^{\mathcal{C}}(\boldsymbol{x}) - H_t^{\mathcal{C}}(\boldsymbol{x})$.

Similarly to (7.5.15),

$$\begin{aligned}
&\frac{\mathbb{E}[\left\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\right\}^2]}{r_n v_n} \\
&\le \|H\| \frac{1}{v_n} \sum_{h=-r_n}^{r_n} \left|\mathbb{E}[\{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right) \times \{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right)]\right| \\
&=: \|H\| \sum_{h=-r_n}^{r_n} |g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})| \qquad (7.5.39)
\end{aligned}$$

with

$$|g_n(h,G)| = \left|\frac{1}{v_n}\mathbb{E}[G(\boldsymbol{X}_{-r_n,r_n}/u_n)G(\boldsymbol{X}_{h-r_n,h+r_n}/u_n)]\right| . \tag{7.5.40}$$

Using the definition (7.5.24) of $H_s^{\mathcal{C}}$, the fact that $s,t \geq s_0$ and since H is bounded, we immediately get

$$|g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})| \leq \frac{4}{v_n}\|H\|^2\mathbb{P}(|\boldsymbol{X}_0| > s_0u_n, |\boldsymbol{X}_h| > s_0u_n) . \tag{7.5.41}$$

To get a more precise bound that involves the difference $s-t$ we need to consider two cases. The reason for this is that we need to keep the absolute value in (7.5.40) outside of the expectation. As such, computations below are quite technically involved.

To shorten our displays, we introduce the notation

$$\mathcal{J}(i,s) := \mathbb{1}\{|\boldsymbol{X}_i| > su_n\} . \tag{7.5.42}$$

Case 1. Assume here that $\mathcal{C}$ is 0-homogeneous. Then for any i,

$$\begin{aligned} &H_s^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \\ &= H\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right) = i\}\mathcal{J}(i,s) \\ &= H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = i\}\mathcal{J}(i,s) ; \end{aligned} \tag{7.5.43}$$

(we keep u_n in the argument of $\mathcal{C}$, although it can be omitted). What is important in this decomposition is that we can control monotonicity (withe respect to s) of each term.

Then

$$\begin{aligned} &H_s^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - H_t^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = \\ &= \mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = i\}H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\Big(\mathcal{J}(i,s) - \mathcal{J}(i,t)\Big) \\ &\quad + \mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = i\}\mathcal{J}(i,t)\Big(H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - H_t\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\Big) \\ &=: T_1(i) + T_2(i) \end{aligned}$$

and hence

$$\left|\mathbb{E}\left[\{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\{H_s^{\mathcal{C}} - H_t^{\mathcal{C}}\}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right)\right]\right|$$

is bounded by the sum of four nonnegative terms $W_{11} + W_{22} + W_{12} + W_{21}$ that we are going to define below. The general idea is that we will obtain rough bounds in terms of the difference $\Big(\mathcal{J}(i,s) - \mathcal{J}(i,t)\Big)$, except of one case which involves the product $T_2(0)T_2(h)$. Coming back to the definitions of $W_{i,j}$, the double sub-index $\{12\}$ of W indicates that W_{12} is related to multiplying $T_1(0)$ by $T_2(h)$; W_{21} means we multiply $T_2(0)$ and $T_1(h)$):

$$W_{11} := \|H\|^2\mathbb{E}\left[\left(\mathcal{J}(0,s) - \mathcal{J}(0,t)\right)\left(\mathcal{J}(h,s) - \mathcal{J}(h,t)\right)\right] ,$$

(above, the indicators of the anchoring map are omitted);

$$\begin{aligned} &W_{22} := \\ &\mathbb{E}\left[\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right) = 0\}\mathcal{J}(0,t)\left\{H_s - H_t\right\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\left\{H_s - H_t\right\}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right)\right] , \end{aligned}$$

(above, $\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right) = h\}$ and $\mathcal{J}(h,t)$ are omitted);

$$W_{12} := \pm\|H\|\mathbb{E}\left[\Big(\mathcal{J}(0,s) - \mathcal{J}(0,t)\Big)\left\{H_s - H_t\right\}\left(\boldsymbol{X}_{h-r_n,h+r_n}/u_n\right)\right] ,$$

(both indicators of the anchoring maps and $\mathcal{J}(h,t)$ are omitted);

$$W_{21} := \pm\|H\|\mathbb{E}\left[\Big(\mathcal{J}(h,s)-\mathcal{J}(h,t)\Big)\{H_s - H_t\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\right] ,$$

(both indicators of the anchoring maps and $\mathcal{J}(0,t)$ are omitted).

Note that the right-hand side of both W_{11}, W_{22} is nonnegative thanks to the monotonicity and $s<t$, while for the right-hand side of W_{12}, W_{21} we need to put $\pm$, since the sign of the expressions there depends on whether the map $s \to H_s$ is decreasing or increasing.

Recall that $v_n = \mathbb{P}(|\boldsymbol{X}_0| > u_n)$. Set also $v_n(s) = \mathbb{P}(|\boldsymbol{X}_0| > su_n)$, $s>0$. Thus,

$$W_{11} \le \|H\|^2 \left(v_n(s) - v_n(t)\right) , \tag{7.5.44}$$

$$W_{22} \le \pm 2\|H\|\mathbb{E}\left[\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)=0\}\mathcal{J}(0,t)\{H_s - H_t\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\right] , \tag{7.5.45}$$

$$W_{12} + W_{21} \le 4\|H\|^2 \left(v_n(s) - v_n(t)\right) . \tag{7.5.46}$$

The bound on W_{22} (with $+$) is obvious if $H_s - H_t \ge 0$ (thus, $s \to H_s$ is decreasing), while in an increasing case of $s \to H_s$ we use the following observation: if $a,b<0$, $|b|<c$, then $ab \le -ac$ (yielding $-$ on the right-hand side of (7.5.45)). The bound on $W_{12}+W_{21}$ follows from the same reasoning.

In summary, with $g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})$ defined in (7.5.40), we have

$$|g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})| \le 5\|H\|^2 \frac{v_n(s)-v_n(t)}{v_n} \tag{7.5.47}$$

$$\pm\, 2\|H\| \frac{\mathbb{E}\left[\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)=0\}\mathbb{1}\{|\boldsymbol{X}_0| > s_0u_n\}\{H_s - H_t\}\left(\boldsymbol{X}_{-r_n,r_n}/u_n\right)\right]}{v_n} , \tag{7.5.48}$$

where again the presence of $\pm$ depends on the sign of $H_s - H_t$.

We can ignore the scaling factor $2\|H\|$ in (7.5.48) and write it as (recall that the anchoring map is 0-homogeneous)

$$\begin{aligned}
&\mathbb{E}[H_{s/s_0}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)=0\} \mid |\boldsymbol{X}_0| > s_0u_n] \frac{v_n(s_0)}{v_n} \\
&\quad - \mathbb{E}[H_{t/s_0}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)=0\} \mid |\boldsymbol{X}_0| > s_0u_n] \frac{v_n(s_0)}{v_n} \\
&= \Big(\mu_{n,r_n}(s) - \mu_{n,r_n}(t)\Big)\frac{v_n(s_0)}{v_n} =: \Big(\tilde{\mu}_{n,r_n}(s) - \tilde{\mu}_{n,r_n}(t)\Big) ,
\end{aligned}$$

with

$$\mu_{n,r_n}(\cdot) = \mathbb{E}[H_{\cdot/s_0}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)\mathbb{1}\{\mathcal{C}\left(\boldsymbol{X}_{-r_n,r_n}/(s_0u_n)\right)=0\} \mid |\boldsymbol{X}_0| > s_0u_n]$$

and

$$\tilde{\mu}_{n,r_n}(s) = \mu_{n,r_n}(s)\frac{v_n(s_0)}{v_n} .$$

Thanks to (7.5.5), $\lim_{n\to\infty}\mu_{n,r_n}(s) = \boldsymbol{\nu}^*(H_{s/s_0})$. Thanks to the monotonicity of $s \to H_S$ and homogeneity of $\boldsymbol{\nu}^*$, the convergence of $\tilde{\mu}_{n,r_n}(s)$ to

$$s_0^{-\alpha}\boldsymbol{\nu}^*(H_{s/s_0}) = s_0^{-\alpha}(s/s_0)^{-\alpha}\boldsymbol{\nu}^*(H) = s^{-\alpha}\boldsymbol{\nu}^*(H)$$

is uniform on $[s_0, t_0]$. Thus, for $s,t \in [s_0,t_0]$,

$$|\tilde{\mu}_{n,r_n}(s) - \tilde{\mu}_{n,r_n}(t)| \le 2\sup_{s_0 \le u \le t_0}|\tilde{\mu}_{n,r_n}(u) - \boldsymbol{\nu}^*(H_u)| + \boldsymbol{\nu}^*(H)\{s^{-\alpha} - t^{-\alpha}\} .$$

Fix $\eta > 0$. For large enough n, the uniform convergence yields

$$\sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} |\tilde{\mu}_{n,r_n}(s) - \tilde{\mu}_{n,r_n}(t)| \le \eta + \boldsymbol{\nu}^*(H) \sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \{s^{-\alpha} - t^{-\alpha}\}$$
$$\le \eta + \alpha s_0^{-\alpha-1} \delta_n \boldsymbol{\nu}^*(H) \; . \quad (7.5.49)$$

The uniform convergence also yields that the term in (7.5.47) is bounded by $\eta + \alpha s_0^{-\alpha-1}\delta_n$. This, together with (7.5.49), gives

$$|g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})| \le \eta + \text{cst } \delta_n \quad (7.5.50)$$

with a generic constant cst.

Fix an integer r. Using (7.5.41) and (7.5.50) we have

$$\sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \frac{\mathbb{E}[\{H_{s,n,0}^{\mathcal{C}} - H_{t,n,0}^{\mathcal{C}}\}^2]}{r_n \mathbb{P}(|\boldsymbol{X}_0| > u_n)}$$
$$\le \sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \sum_{h=-r}^{r} |g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})| + \sup_{\substack{s_0 \le s,t \le t_0 \\ |s-t| \le \delta_n}} \sum_{|h|=r}^{r_n} |g_n(h, H_s^{\mathcal{C}} - H_t^{\mathcal{C}})|$$
$$\le \text{cst } r(\eta + \delta_n) + \frac{4}{\mathbb{P}(|\boldsymbol{X}_0| > u_n)} \sum_{|h|=r}^{r_n} \mathbb{P}(|\boldsymbol{X}_0| > s_0 u_n, |\boldsymbol{X}_h| > s_0 u_n) \; .$$

Applying the anticlustering conditions $\mathcal{S}(r_n, u_n)$ to the second term, letting $\delta_n \to 0$, since η is arbitrary, this proves (7.5.37).

Case 2. Now, we consider the anchoring maps $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$ which are not 0-homogeneous. Note that for $j = 1, 2$ we can write

$$H_s^{\mathcal{C}^{(j)}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = H\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right) \mathbb{1}\Big\{\mathcal{C}^{(j)}\left(\boldsymbol{X}_{i-r_n,i+r_n}/(su_n)\right) = i\Big\} \mathcal{J}(i,s)$$
$$= H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \mathcal{J}(i,s) F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \; ,$$

where

$$F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = \mathbb{1}\left\{\boldsymbol{X}^*_{i-r_n,i-1} \le su_n\right\}$$

or

$$F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) = \mathbb{1}\left\{\boldsymbol{X}^*_{i+1,i+r_n} \le su_n\right\}$$

in case $j = 1$ and $j = 2$, respectively. Note that regardless of the monotonicity of the map $s \to \mathcal{C}_s^{(j)}$, the map $s \to F_s$ is always non-decreasing. Then (7.5.43) gives, for any $i \in \mathbb{N}$, and $\mathcal{C} = \mathcal{C}^{(1)}, \mathcal{C}^{(2)}$,

$$H_s^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - H_t^{\mathcal{C}}\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) =$$
$$= H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \Big(\mathcal{J}(i,s) - \mathcal{J}(i,t)\Big)$$
$$+ \mathcal{J}(i,t) F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \left(H_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - H_t\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\right)$$
$$+ \mathcal{J}(i,t) H_t\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) \Big(F_s\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right) - F_t\left(\boldsymbol{X}_{i-r_n,i+r_n}/u_n\right)\Big)$$
$$=: T_1(i) + T_2(i) + T_3(i) \; .$$

Again, in this decomposition we can control monotonicity of each term. The argument now is very similar to that of Case 1. Hence, it is omitted.

Therefore, (7.5.37) is proved for $\mathcal{C}^{(j)}$, $j = 0, 1, 2$.

Proof of (7.5.38)**.** We proceed similarly. Recall that for $j\geq 1$ (cf. (7.5.18)),

$$\begin{aligned}
&\frac{1}{r_nv_n}\mathbb{E}[\{H^{\mathcal{C}}_{s,n,0}-H^{\mathcal{C}}_{t,n,0}\}\{H^{\mathcal{C}}_{s,n,j}-H^{\mathcal{C}}_{t,n,j}\}] \\
&=\frac{1}{r_nv_n}\mathbb{E}\left[\sum_{h=1}^{r_n}\left\{H^{\mathcal{C}}_s-H^{\mathcal{C}}_t\right\}(\boldsymbol{X}_{h-r_n,h+r_n}/u_n)\times\sum_{i=jr_n+1}^{(j+1)r_n}\left\{H^{\mathcal{C}}_s-H^{\mathcal{C}}_t\right\}(\boldsymbol{X}_{i-r_n,i+r_n}/u_n)\right] \\
&=\sum_{h=(j-1)r_n+1}^{jr_n}\left(\frac{h}{r_n}-(j-1)\right)g_n(h,H^{\mathcal{C}}_s-H^{\mathcal{C}}_t)+\sum_{h=jr_n+1}^{(j+1)r_n}\left((j+1)-\frac{h}{r_n}\right)g_n(h,H^{\mathcal{C}}_s-H^{\mathcal{C}}_t) \\
&\leq\sum_{h=(j-1)r_n+1}^{(j+1)r_n}g_n(h,H^{\mathcal{C}}_s-H^{\mathcal{C}}_t)=:I_j(s,t) \qquad (7.5.51)
\end{aligned}$$

with the same g_n as in (7.5.40).

Write $g_n(h,G)$ as

$$\begin{aligned}
&\frac{1}{v_n}\mathrm{cov}[G(\boldsymbol{X}_{-r_n,r_n}/u_n),G(\boldsymbol{X}_{h-r_n,h+r_n}/u_n)]+\frac{1}{v_n}\mathbb{E}^2[G(\boldsymbol{X}_{-r_n,r_n}/u_n)] \\
&=:\widetilde{g}_n(h,G)+\frac{1}{v_n}\mathbb{E}^2[G(\boldsymbol{X}_{-r_n,r_n}/u_n)]\;.
\end{aligned}$$

For $h>2r_n$ we have by (7.5.1),

$$|\widetilde{g}_n(h,H^{\mathcal{C}}_s-H^{\mathcal{C}}_t)|\leq\frac{\mathrm{cst}}{v_n}\beta_{h-2r_n}\;. \qquad (7.5.52)$$

Thus,

$$\begin{aligned}
&\frac{1}{r_nv_n}\sum_{j=4}^{z_n-1}\left(1-\frac{j}{z_n}\right)\mathbb{E}[\left(H^{\mathcal{C}}_{s,n,0}-H^{\mathcal{C}}_{t,n,0}\right)\left(H^{\mathcal{C}}_{s,n,j}-H^{\mathcal{C}}_{t,n,j}\right)] \\
&\leq\frac{\mathrm{cst}}{v_n}\sum_{j=4}^{z_n-1}\sum_{h=(j-1)r_n+1}^{(j+1)r_n}\beta_{h-2r_n}+\frac{\mathrm{cst}}{v_n}z_nr_n\mathbb{E}^2[\{H^{\mathcal{C}}_s-H^{\mathcal{C}}_t\}(\boldsymbol{X}_{-r_n,r_n}/u_n)] \\
&\leq\frac{\mathrm{cst}}{v_n}\sum_{h=3r_n+1}^{\infty}\beta_{h-2r_n}+\mathrm{cst}z_nr_n\mathbb{P}(|\boldsymbol{X}_0|>s_0u_n) \\
&=\frac{\mathrm{cst}}{v_n}\sum_{i=r_n+1}^{\infty}\beta_i+o(1)=o(1) \qquad (7.5.53)
\end{aligned}$$

uniformly in $s,t\in[s_0,t_0]$. In the last line we applied the assumption (7.3.6b), and the assumption (7.5.25).

The terms that correspond to $j=1,2,3$ in (7.5.51) have to be dealt with separately. We note that $I_1(s,t)=\sum_{h=1}^{r_n}g_n(h,H^{\mathcal{C}}_s-H^{\mathcal{C}}_t)$ is bounded by the term in (7.5.39). Hence,

$$\lim_{n\to\infty}\sup_{\substack{s_0\leq s,t\leq t_0\\|s-t|\leq\delta_n}}I_1(s,t)=0\;. \qquad (7.5.54)$$

Next, using (7.5.41) and $\mathcal{S}(r_n,u_n)$,

$$\begin{aligned}
&I_2(s,t)+I_3(s,t) \\
&\leq\mathrm{cst}\sum_{h=r_n+1}^{4r_n}g_n(h)\leq\mathrm{cst}\sum_{h=r_n+1}^{4r_n}\mathbb{P}(|\boldsymbol{X}_0|>s_0u_n,|\boldsymbol{X}_h|>s_0u_n)=0\;, \qquad (7.5.55)
\end{aligned}$$

uniformly in $s,t\in[s_0,t_0]$.

Combination of (7.5.53), (7.5.54), (7.5.55) finishes the proof of (7.5.38).

This, together with (7.5.37) concluded the proof of (7.5.35).

7.5.9. Random entropy

In this section we discuss validity of Assumption 7.3.5. We cannot check this condition for arbitrary functionals H and anchoring maps $\mathcal{C}$, however, we will see that the conditions is satisfied for most relevant cases considered in the paper.

Recall the class

$$\mathcal{G}=\{H_s^{\mathcal{C}},s\in[s_0,t_0]\}=\{H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s)=0\}\mathbb{1}\{|\boldsymbol{x}_0|>s\},\ \ s\in[s_0,t_0]\}.$$

We start first with H of the form

$$H_s(\boldsymbol{x})=\mathbb{1}\{K(\boldsymbol{x})>s\}\ , \tag{7.5.56}$$

where $K:\mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$. This is the case of the functionals that lead to the extremal index, the large deviation index and the ruin index.

Since $\mathcal{C}^{(0)}$ is 0-homogeneous, it does not play a role in calculating the class entropy. Then

$$H_s(\boldsymbol{x})\mathbb{1}\{|\boldsymbol{x}_0|>s\}=\mathbb{1}\{\min\{K(\boldsymbol{x}),|\boldsymbol{x}_0|\}>s\}\ .$$

Hence, the map $s\to H_s(\boldsymbol{x})\mathbb{1}\{|\boldsymbol{x}_0|>s\}$ is decreasing. Therefore, $\mathrm{VC}(\mathcal{G})=2$.

As for $\mathcal{C}^{(1)}$ we have

$$\mathbb{1}\left\{\mathcal{C}^{(1)}(\boldsymbol{x}/s)=0\right\}=\mathbb{1}\left\{\boldsymbol{x}^*_{-\infty,-1}\le s,|\boldsymbol{x}_0|>s\right\}\ .$$

Thus,

$$H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s)=0\}\mathbb{1}\{|\boldsymbol{x}_0|>s\}=\mathbb{1}\{\min\{K(\boldsymbol{x}),|\boldsymbol{x}_0|\}>s\}\mathbb{1}\left\{\boldsymbol{x}^*_{-\infty,-1}\le s\right\}\ .$$

Now, the class

$$\mathcal{F}=\{(-\infty,s)\times(s,+\infty):s\in\mathbb{R}\}$$

has the VC-index 3. By [KS20, Example C.4.14] the class $\mathcal{G}$ has VC-index at most 3.

Similarly, for $\mathcal{C}^{(2)}$ we have

$$\mathbb{1}\left\{\mathcal{C}^{(2)}(\boldsymbol{x}/s)=0\right\}=\mathbb{1}\left\{\boldsymbol{x}^*_{1,\infty}\le s,|\boldsymbol{x}_0|>s\right\}\ .$$

Thus,

$$H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s)=0\}\mathbb{1}\{|\boldsymbol{x}_0|>s\}=\mathbb{1}\{\min\{K(\boldsymbol{x}),|\boldsymbol{x}_0|\}>s\}\mathbb{1}\left\{\boldsymbol{x}^*_{1,\infty}\le s\right\}$$

and again the class $\mathcal{G}$ has VC-index at most 3.

In summary, for functionals H given in (7.5.56) and the anchoring maps $\mathcal{C}^{(0)}$, $\mathcal{C}^{(1)}$, $\mathcal{C}^{(2)}$ the class $\mathcal{G}$ has the VC-index at most 3 and hence the random entropy Assumption 7.3.5 is satisfied.

Now, assume that the map $s\to H_s$ is decreasing. This is the case of (again) the extremal index, the large deviation index and the ruin index. This is also the case of the stop-loss index and the cluster size distribution. If we choose $\mathcal{C}=\mathcal{C}^{(0)}$, since $\mathcal{C}^{(0)}$ is 0-homogeneous, the maps $s\to H_s^{\mathcal{C}}$ is also decreasing. Thus, the VC-index of $\mathcal{G}$ is at most 2. The random entropy condition is satisfied.

Bibliography

[BB18] Betina Berghaus and Axel Bücher. Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Annals of Statistics*, 46(5):2307–2335, 2018.

[BBKS20] Clemonell Bilayi-Biakana, Rafał Kulik, and Philippe Soulier. Statistical inference for heavy tailed series with extremal independence. *Extremes*, 23(1):1–33, 2020.

[BP18] Bojan Basrak and Hrvoje Planinić. A note on vague convergence of measures. *Statistics and Probability Letters*, 153:180–186, 2019.

[BPS18] Bojan Basrak, Hrvoje Planinić, and Philippe Soulier. An invariance principle for sums and record times of regularly varying stationary sequences. *Probability Theory and Related Fields*, 172(3-4):869–914, 2018.

[BS09] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[BGTS09] Jan Beirlant and Yuri Goegebeur and Jozef Teugels and Johan Segers. *Statistics of extremes* John Wiley & Sons, Ltd., Chichester, 2004.

[BS18a] Axel Bücher and Johan Segers. Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125, 2018.

[BS18b] Axel Bücher and Johan Segers. Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462, 2018.

[BZ21] Axel Bücher and Chen Zhou. A horse racing between the block maxima method and the peak–over–threshold approach. *Statistical Science*, 36(3), 360-378, 2021.

[CK21] Youssouph Cissokho and Rafal Kulik. Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electronic Journal of Statistics*, 15(1): 2777–2831, 2021.

[DM09] Richard A. Davis and Thomas Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 38A:977–1009, 2009. Probability, statistics and seismology.

[DH95] Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability*, 23(2):879–917, 1995.

[DHS18] Clément Dombry, Enkelejd Hashorva, and Philippe Soulier. Tail measure and spectral tail process of regularly varying time series. *Annals of Applied Probability*, 28(6):3884–3921, 2018.

[DK20] Holger Drees and Miran Knezevic. Peak-over-threshold estimators for spectral tail processes: Random vs deterministic thresholds. *Extremes*, 2020. DOI: https://doi.org/10.1007/s10687-019-00367-x.

[DN21] Holger Drees and Sebastian Neblung. Asymptotics for sliding blocks estimators of rare events. *Bernoulli* 27(2), 1239–1269, 2021.

[DR10] Holger Drees and Holger Rootzén. Limit theorems for empirical processes of cluster functionals. *Annals of Statistics*, 38(4):2145–2186, 2010.

[DSW15] Holger Drees, Johan Segers, and Michał Warchoł. Statistics for tail processes of Markov chains. *Extremes*, 18(3):369–402, 2015.

[EKM97] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, 1997.

[FdH15] Ana Ferreira, Laurens de Haan. On the block maxima method in extreme value theory: PWM estimators *Annals of Statistics*, 43(1): 276-298, 2015.

[Has18] Enkelejd Hashorva. Representations of max-stable processes via exponential tilting, *Stochastic Processes and their Applications*, 128(9): 2952–2978, 2018.

[Hsi91] Tailen Hsing. Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37(1):117–139, 1991.

[Ibr62] I. A. Ibragimov. Some limit theorems for stationary processes. *Theor. Probab. Appl.*, 7:349–382, 1962.

[KS20] Rafał Kulik and Philippe Soulier. *Heavy tailed time series.* Springer, 2020.

[KSW19] Rafał Kulik, Philippe Soulier, and Olivier Wintenberger. The tail empirical process of regularly varying functions of geometrically ergodic markov chains. *Stochastic Processes and their Applications*, 129(1):4209–4238, 2019.

[MW13] Thomas Mikosch and Olivier Wintenberger. Precise large deviations for dependent regularly varying sequences. *Probability Theory and Related Fields*, 156(3-4):851–887, 2013.

[MW14] Thomas Mikosch and Olivier Wintenberger. 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, 2014.

[MW16] Thomas Mikosch and Olivier Wintenberger. A large deviations approach to limit theorem for heavy-tailed time series. *Probability Theory and Related Fields*, 166(1-2):233–269, 2016.

[PS18] Hrvoje Planinić and Philippe Soulier. The tail process revisited. *Extremes*, 21(4):551–579, 2018.

[RLdH98] Holger Rootzén, Ross M. Leadbetter, and Laurens de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8(3):868–885, 1998.

[RSF09] Christian Y. Robert, Johan Segers, and Christopher A. T. Ferro. A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3:993–1020, 2009.

[Sam2016] Gennady Samorodnitsky. *Stochastic processes and long range dependence* Springer, 2016.

[SW94] Richard L. Smith and Ishay Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society. Series B. Methodological*, 56(3):515–528, 1994.

[vdVW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes.* Springer, New York, 1996.

[WN98] Ishay Weissman and S. Yu. Novak. On blocks and runs estimators of the extremal index. *Journal of Stat. Planning and Inference*, 66(2):281–288, 1998.

[ZVB21] Nan Zou, Stanislav Volgushev, and Axel Bücher. Multiple block sizes and overlapping blocks for multivariate time series extremes. *Annals of Statistics* 49(1), 295-320, 2021.

(k %)	$\rho = 0.9$, Extremal Index=0.34				$\rho = 0.5$, Extremal Index= 0.94			
	$k = 5$		$k = 10$		$k = 5$		$k = 10$	
$r_n = 5$								
Disjoint bl	**0.380**	(0.0499)	**0.350**	(0.0334)	**0.680**	(0.0513)	**0.620**	(0.0355)
Sliding bl	**0.376**	(0.0441)	**0.348**	(0.0293)	**0.692**	(0.0429)	**0.622**	(0.0285)
Runs $C^{(0)}$	**0.240**	(0.0488)	**0.180**	(0.0314)	**0.560**	(0.0442)	**0.450**	(0.0342)
Runs $C^{(1)}$	**0.220**	(0.0498)	**0.150**	(0.0318)	**0.540**	(0.0563)	**0.420**	(0.0357)
Runs $C^{(2)}$	**0.220**	(0.0499)	**0.150**	(0.0320)	**0.540**	(0.0564)	**0.420**	(0.0357)
$r_n = 6$								
Disjoint bl	**0.340**	(0.0498)	**0.310**	(0.0313)	**0.660**	(0.0540)	**0.580**	(0.0351)
Sliding bl	**0.340**	(0.0441)	**0.317**	(0.0275)	**0.670**	(0.0447)	**0.539**	(0.0283)
Runs $C^{(0)}$	**0.200**	(0.0468)	**0.170**	(0.0273)	**0.540**	(0.0555)	**0.420**	(0.0340)
Runs $C^{(1)}$	**0.180**	(0.0480)	**0.140**	(0.0288)	**0.520**	(0.0581)	**0.390**	(0.0361)
Runs $C^{(2)}$	**0.180**	(0.0479)	**0.140**	(0.0288)	**0.520**	(0.0577)	**0.390**	(0.0362)
$r_n = 7$								
Disjoint bl	**0.320**	(0.0515)	**0.290**	(0.0314)	**0.640**	(0.0524)	**0.550**	(0.0343)
Sliding bl	**0.317**	(0.0458)	**0.291**	(0.0281)	**0.649**	(0.0442)	**0.558**	(0.0281)
Runs $C^{(0)}$	**0.180**	(0.0465)	**0.160**	(0.0270)	**0.520**	(0.0537)	**0.400**	(0.0300)
Runs $C^{(1)}$	**0.180**	(0.0483)	**0.140**	(0.0275)	**0.500**	(0.0557)	**0.370**	(0.0316)
Runs $C^{(2)}$	**0.180**	(0.0482)	**0.140**	(0.0273)	**0.500**	(0.0560)	**0.370**	(0.0317)
$r_n = 8$								
Disjoint bl	**0.300**	(0.0479)	**0.270**	(0.0318)	**0.640**	(0.0537)	**0.540**	(0.0340)
Sliding bl	**0.303**	(0.0425)	**0.273**	(0.0284)	**0.633**	(0.0451)	**0.535**	(0.0281)
Runs $C^{(0)}$	**0.180**	(0.0437)	**0.150**	(0.0273)	**0.500**	(0.0517)	**0.370**	(0.0298)
Runs $C^{(1)}$	**0.180**	(0.0441)	**0.130**	(0.0281)	**0.480**	(0.0539)	**0.340**	(0.0323)
Runs $C^{(2)}$	**0.180**	(0.0440)	**0.130**	(0.0281)	**0.480**	(0.0539)	**0.340**	(0.0322)

TABLE 7.1

The median and the variance (in brackets) of disjoint, sliding blocks and runs (with the anchoring maps $C^{(0)}, C^{(1)}$ and $C^{(2)}$) estimators for the extremal index. Data are simulated from AR(1) with $\alpha = 4$, $\rho = 0.5$ (thus, $\theta = 0.94$), and $\rho = 0.9$ (thus $\theta = 0.34$). Block sizes are $r_n = 5,\ 6,\ 7,\ 8$. The number of order statistics are $k = 5\%,\ 10\%$ of the sample size n with $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

	$\rho = 0.9$, Stop-loss Index=0.085				$\rho = 0.5$, Stop-loss Index= 0.078			
(k %)	$k = 10$		$k = 40$		$k = 10$		$k = 40$	
$r_n = 6$								
Disjoint bl	**0.000**	(0.0118)	**0.008**	(0.0073)	**0.030**	(0.0192)	**0.055**	(0.0105)
Sliding bl	**0.000**	(0.0110)	**0.009**	(0.0069)	**0.032**	(0.0177)	**0.055**	(0.0093)
Runs $C^{(0)}$	**0.010**	(0.0106)	**0.020**	(0.0061)	**0.040**	(0.0192)	**0.070**	(0.0084)
Runs $C^{(1)}$	**0.000**	(0.0084)	**0.003**	(0.0032)	**0.030**	(0.0184)	**0.020**	(0.0070)
Runs $C^{(2)}$	**0.000**	(0.0014)	**0.000**	(0.0000)	**0.030**	(0.0177)	**0.015**	(0.0068)
$r_n = 10$								
Disjoint bl	**0.000**	(0.0122)	**0.018**	(0.0073)	**0.040**	(0.0196)	**0.068**	(0.0089)
Sliding bl	**0.006**	(0.0108)	**0.017**	(0.0067)	**0.049**	(0.0179)	**0.067**	(0.0074)
Runs $C^{(0)}$	**0.010**	(0.0110)	**0.022**	(0.0049)	**0.050**	(0.0191)	**0.070**	(0.0071)
Runs $C^{(1)}$	**0.010**	(0.0097)	**0.005**	(0.0038)	**0.030**	(0.0183)	**0.010**	(0.0054)
Runs $C^{(2)}$	**0.000**	(0.0082)	**0.000**	(0.0024)	**0.030**	(0.0180)	**0.010**	(0.0049)
$r_n = 20$								
Disjoint bl	**0.010**	(0.0114)	**0.023**	(0.0054)	**0.050**	(0.0188)	**0.073**	(0.0067)
Sliding bl	**0.010**	(0.0106)	**0.024**	(0.0046)	**0.049**	(0.0168)	**0.073**	(0.0051)
Runs $C^{(0)}$	**0.010**	(0.0102)	**0.023**	(0.0039)	**0.060**	(0.0169)	**0.053**	(0.0048)
Runs $C^{(1)}$	**0.010**	(0.0098)	**0.005**	(0.0037)	**0.030**	(0.0163)	**0.000**	(0.0020)
Runs $C^{(2)}$	**0.010**	(0.0098)	**0.003**	(0.0033)	**0.020**	(0.0165)	**0.000**	(0.0020)
$r_n = 30$								
Disjoint bl	**0.010**	(0.0106)	**0.023**	(0.0046)	**0.060**	(0.0188)	**0.065**	(0.0043)
Sliding bl	**0.011**	(0.0092)	**0.024**	(0.0038)	**0.056**	(0.0160)	**0.066**	(0.0032)
Runs $C^{(0)}$	**0.010**	(0.0096)	**0.020**	(0.0037)	**0.075**	(0.0161)	**0.035**	(0.0040)
Runs $C^{(1)}$	**0.010**	(0.0094)	**0.003**	(0.0035)	**0.020**	(0.0142)	**0.000**	(0.0006)
Runs $C^{(2)}$	**0.010**	(0.0094)	**0.003**	(0.0029)	**0.020**	(0.0142)	**0.000**	(0.0006)

TABLE 7.2

The median and the variance (in brackets) of disjoint, sliding blocks and runs ($C^{(0)}$) estimators for stop-loss index with $S = 0.9$. Data are simulated from AR(1) with $\alpha = 4,\ \rho = 0.5,\ 0.9$. The block sizes are $r_n = 6,\ 10,\ 20,\ 30$. The number of order statistics are $k = 10\%,\ 40\%$ of the sample size n with $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

(k %)	$k=6$		$k=10$		$k=20$		$k=30$	
	Extremal Index=0.612							
$r_n = 6$								
Disjoint bl	**0.650**	(0.0544)	**0.620**	(0.0340)	**0.550**	(0.0216)	**0.470**	(0.0127)
Sliding bl	**0.647**	(0.0479)	**0.625**	(0.0338)	**0.551**	(0.0175)	**0.472**	(0.0096)
Runs $C^{(0)}$	**0.483**	(0.0573)	**0.440**	(0.0392)	**0.320**	(0.0185)	**0.233**	(0.0118)
Runs $C^{(1)}$	**0.467**	(0.0594)	**0.410**	(0.0399)	**0.245**	(0.0183)	**0.1167**	(0.0122)
Runs $C^{(2)}$	**0.467**	(0.0594)	**0.410**	(0.0401)	**0.245**	(0.0183)	**0.1167**	(0.0121)
$r_n = 10$								
Disjoint bl	**0.567**	(0.0563)	**0.530**	(0.0370)	**0.410**	(0.0157)	**0.317**	(0.0060)
Sliding bl	**0.568**	(0.0497)	**0.526**	(0.0314)	**0.410**	(0.0120)	**0.314**	(0.0041)
Runs $C^{(0)}$	**0.417**	(0.0493)	**0.340**	(0.0305)	**0.210**	(0.0134)	**0.143**	(0.0085)
Runs $C^{(1)}$	**0.400**	(0.0493)	**0.300**	(0.0308)	**0.115**	(0.0145)	**0.0013**	(0.0082)
Runs $C^{(2)}$	**0.400**	(0.0499)	**0.300**	(0.0309)	**0.115**	(0.0146)	**0.0030**	(0.0081)
$r_n = 20$								
Disjoint bl	**0.450**	(0.0448)	**0.380**	(0.0244)	**0.240**	(0.0054)	**0.167**	(0.0009)
Sliding bl	**0.450**	(0.0340)	**0.374**	(0.0196)	**0.266**	(0.0036)	**0.163**	(0.0005)
Runs $C^{(0)}$	**0.300**	(0.0349)	**0.210**	(0.0201)	**0.110**	(0.0093)	**0.073**	(0.0062)
Runs $C^{(1)}$	**0.250**	(0.0351)	**0.130**	(0.0210)	**0.0015**	(0.0083)	**0.000**	(0.0021)
Runs $C^{(2)}$	**0.250**	(0.0356)	**0.130**	(0.0209)	**0.015**	(0.0084)	**0.000**	(0.0021)
$r_n = 30$								
Disjoint bl	**0.383**	(0.0352)	**0.290**	(0.0159)	**0.165**	(0.0018)	**0.110**	(0.0001)
Sliding bl	**0.373**	(0.0291)	**0.284**	(0.0118)	**0.160**	(0.0010)	**0.010**	(0.0000)
Runs $C^{(0)}$	**0.217**	(0.0280)	**0.140**	(0.0154)	**0.07**	(0.008)	**0.047**	(0.0005)
Runs $C^{(1)}$	**0.150**	(0.0300)	**0.060**	(0.0172)	**0.000**	(0.0036)	**0.000**	(0.0004)
Runs $C^{(2)}$	**0.150**	(0.0300)	**0.060**	(0.0170)	**0.000**	(0.0036)	**0.000**	(0.0004)

TABLE 7.3

The median and the variance (in brackets) of disjoint, sliding blocks and runs ($C^{(0)}$, $C^{(1)}$ and $C^{(2)}$) estimators for the extremal index in ARCH(1) model with $\lambda = 0.9$. The block size is $r_n = 6,\ 10,\ 20,\ 30$. The number of order statistics is $k = 6\%$, 10%, 20% and 30% of the sample size n with $n = 1000$ based on $N = 1000$ Monte Carlo simulations.

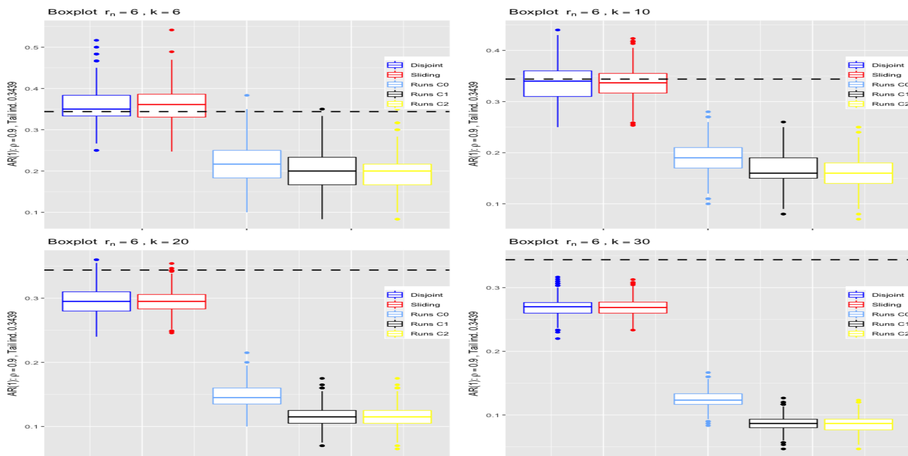

FIG 7.1. Monte Carlo simulations for a sample $n = 1000$ based on $N = 1000$ replicates of disjoint, sliding blocks and runs estimators ($C^{(0)}$, $C^{(1)}$ and $C^{(2)}$) for the extremal index. Data are simulated from AR(1) with $\alpha = 4$, $\rho = 0.9$ (thus, $\theta = 0.34$) The block size $r_n = 6$ and the number of order statistics $k = 6\%$ (top left), $k = 10\%$ (top right), $k = 20\%$ (bottom left), $k = 30\%$ (bottom right) of the sample size n. Dotted lines indicated the true value of the cluster index.

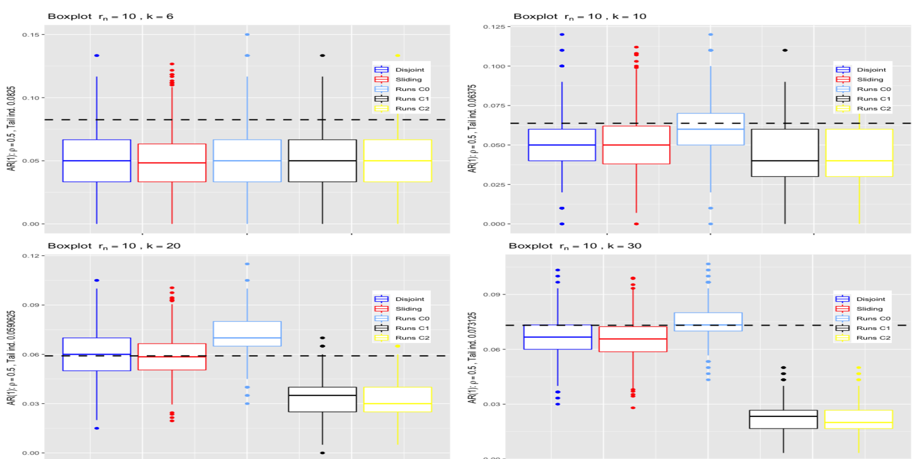

FIG 7.2. Monte Carlo simulations for a sample $n = 1000$ based on $N = 1000$ replicates of disjoint, sliding blocks and runs estimators ($C^{(0)}$, $C^{(1)}$ and $C^{(2)}$) for stop-loss index with $S = 0.9$. Data are simulated from AR(1) with $\rho = 0.5$, $r_n = 10$ and the number of order statistics $k = 6\%$ (top left), $k = 10\%$ (top right), $k = 20\%$ (bottom left), $k = 30\%$ (bottom right) of the sample size n. Dotted lines indicated the true value of the cluster index.

Chapter 8

Summary and future work

8.1. Summary

Stationary, multivariate time series often exhibits dependence between its coordinates and over time. As such, the analysis of dependence between large observations is a suitable approach to understand extremal behaviour of stationary time series. So-called cluster indices, including the well-known extremal index, are a crucial tool for this purpose. Of interest, is the estimation of cluster indices. In this regard, we have focused in this thesis, on asymptotic behaviour of three types of estimators: *disjoint blocks*, *sliding blocks* and *runs estimators*. We used a modern theory of multivariate, regularly varying time series (see Chapter 2). Results for disjoint blocks estimators are taken from the existing literature. We obtained consistency results (chapter 4), central limit theorem Chapter 6 for sliding block estimator, as well as central limit theorem for runs estimators (chapter 7) under conditions that can be easily verified for a large class of short-range dependent models. In particular, we show that, from the theoretical point of view, in the Peak-over-Threshold framework, all the estimators (disjoint blocks, sliding blocks, runs) have the same limiting variances. This solves a longstanding open problem and is in contrast to the Block Maxima method.

However, for finite samples a bias has to be taken into account. We note first that the theoretical finite-sample bias for disjoint and sliding blocks estimators is the same. We do not have results on bias for runs estimators.

The theoretical results have been illustrated in simulation studies. The findings are summarized as follows:

- All estimators (blocks and runs) have virtually the same variance, which is in line with the theoretical results.
- For the extremal index, runs estimators are inferior as compared to blocks estimators. This is primarily due to bias.
- For the extremal index, blocks estimators perform well in case of a) strong dependence; b) small number of order statistics; c) small block sizes.
- For the stop-loss index, both blocks and runs estimator $C^{(0)}$ are superior, in comparison to runs estimators $C^{(1)} and\, C^{(2)}$; the latter being heavily biased in most cases.
- For the stop-loss index, blocks estimators and $C^{(0)}$ perform well in case of a) weak dependence; b) small number of order statistics; c) wide range of block sizes.
- Furthermore, it is rather not feasible that linear combinations of the blocks and runs will reduce the bias.

8.2. Further Research Directions

In what follows, we envision the following open questions and research directions:

- Extend CLT for sliding blocks estimators (Theorem 6.4.3) to piecewise stationary processes. This line of research was proposed recently by Axel Bücher and his student. Piecewise stationary processes may be used in climate modeling.
- Obtain the results of Theorem 6.4.3 under minimal conditions (that is, without relying on β-mixing and linear ordering of function classes). Do these results are valid under long range dependence?
- Can we extend the asymptotic results presented here to Gumbel domain of attraction? note that the probabilistic methods have to be completely different.
- Since the disjoint and sliding blocks statistics have the same asymptotic behaviour, is it possible to obtain an asymptotic expansion for the difference between these two statistics?
- Can we compare results between Peak-over-Threshold and Block Maxima methods?

Chapter 9

Appendix - VC classes

9.1. VC and VC-subgraph classes

Let $\mathcal{F}$ be a class of subsets of a measurable space E. For a finite set $A \subseteq \mathsf{E}$, denote by $\mathrm{Trace}^{\mathcal{F}}(A)$ the trace of $\mathcal{F}$ on A, that is the collection of all subsets of A obtained by intersection of A with sets F of $\mathcal{F}$. We also denote by $\Delta^{\mathcal{F}}(A)$ the cardinality of $\mathrm{Trace}^{\mathcal{F}}(A)$. Note that $\Delta^{\mathcal{F}}(A) \leq 2^{\mathrm{Card(A)}}$. We say that $\mathcal{F}$ **shatters** A if $\Delta^{\mathcal{F}}(A) = 2^{\mathrm{Card(A)}}$, that is, if every subset of A is the intersection of A with some set $F \in \mathcal{F}$. Let

$$m^{\mathcal{F}}(k) = \sup_{A \subseteq \mathsf{E}:\ \mathrm{card(A)}=k} \Delta^{\mathcal{F}}(A)$$

and

$$VC(\mathcal{F}) = \inf\{k \ :\ m^{\mathcal{F}}(k) < 2^k\}.$$

Definition 9.1.1 (VC-classes)**.** *A collection of sets $\mathcal{F}$ is called a VC-class if VC($\mathcal{F}$) $< \infty$, that is there exists a $k < \infty$ such that $\mathcal{F}$ does* ***not shatter*** *any subsets of E of cardinality k. The quantity VC($\mathcal{F}$) is called the VC-index of the class $\mathcal{F}$.*

Examples

Example 9.1.2 (Example C.4.10, [KS20]). Let $\mathsf{E} = \mathbb{R}$, $\mathcal{F} = \{(-\infty, c) \ :\ c \in \mathbb{R}\}$ and $A = \{x_1, x_2\}$, $x_1 < x_2$. Then $\mathrm{Trace}^{\mathcal{F}}(A) = \{\emptyset, \{x_1\}, \{x_1, x_2\}\}$ and $\Delta^{\mathcal{F}}(A) = 3 < 4$. Hence, the VC-index of $\mathcal{F}$ is 2 since no two points can be shattered. In general, let $\mathcal{F} = \{(-\infty, c) \ :\ c \in \mathbb{R}^d\}$. Then the VC-index of $\mathcal{F}$ is d+1.

Example 9.1.3. We continue with the situation of Example 9.1.2. We are going to provide a bound on the VC index using the **hyperplanes method**. This method is useful when it is hard to draw sets. The method does not give the precise value of the VC-index, but provides useful upper bounds.

Let $d = 2$, $\mathcal{F} = \{M_{c,d},\ c,\, d \in \mathbb{R}\} = \{(-\infty, c) \times (d, +\infty) : c,\, d \in \mathbb{R}\}$ and let $A = \{x_1, \ldots, x_k\}$ be a set of k points $x_j = (x_{j,1}, x_{j,2})$ in $\mathbb{R}^2$.

1. If the symmetric difference, $M_{c_1,d_1} \Delta M_{c_2,d_2}$ (**cross hatching area in Figure 9.1**) does not include any of the points $x_1, \ldots, x_k$, then $M_{c_1,d_1} \cap A = M_{c_2,d_2} \cap A$.

2. Consider the hyperplanes

$$\{y = (y_1, y_2) \in \mathbb{R}^2 \,:\, y_1 = x_{j,1}, j = 1, \cdots, k\}$$
$$\cup \{y = (y_1, y_2) \in \mathbb{R}^2 \,:\, y_2 = x_{j,2}, j = 1, \cdots, k\}\,.$$

Then there are at most $(k+1)^2$ hypercubes created. If (c_1, d_1) and (c_2, d_2) are in the same hypercube then the symmetric difference $M_{c_1,d_1} \Delta M_{c_2,d_2}$ does not include any of the points from the set A. Thus, M_{c_1,d_1} and M_{c_2,d_2} can pick out different subsets of A only if the points (c_1, d_1) and (c_2, d_2) are in different hypercubes. Therefore, the collection $\mathcal{F} = \{M_{c,d},\ c,\, d \in \mathbb{R}\}$ can pick out at most $(k+1)^2$ subsets of A.
3. Since there are 2^k subsets of A, the collection cannot shutter A if $(k+1)^2 < 2^k$ which holds for $k \geq 6$. This means that no $\{x_1, \cdots, x_6\}$ can be shattered. Hence $VC(\mathcal{F}) \leq 6$.

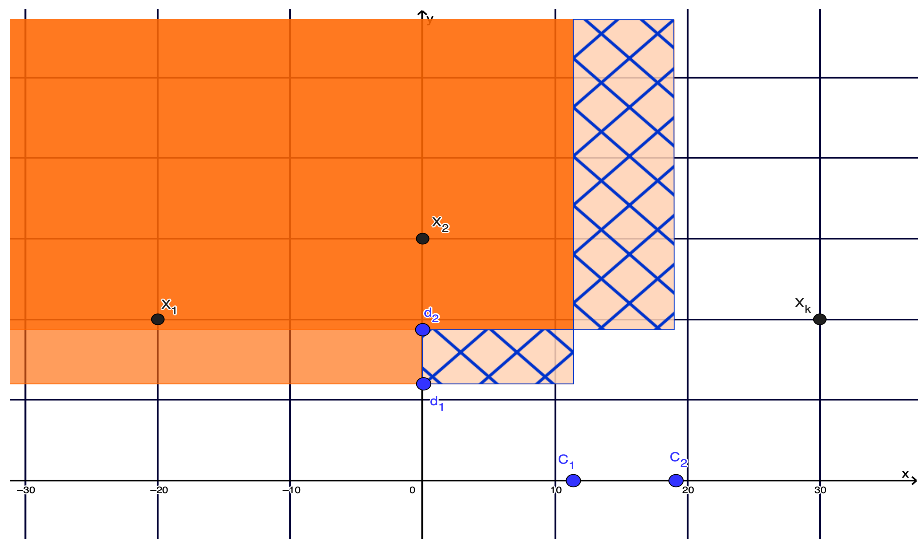

Fig 9.1. *The symmetric difference* $M_{c_1,d_1} \Delta M_{c_2,d_2}$

Let $d = 3$, $\mathcal{F} = \{M_{b,c,d},\ c,\, d,\, b \in \mathbb{R}\} = \{(-\infty, b) \times (-\infty, c) \times (-\infty,\, d) \ :\ c,\, b,\, d \in \mathbb{R}\}$ and let $A = \{x_1, \ldots, x_k\}$ be a set of k points $x_j = (x_{j,1}, x_{j,2}, x_{j,3})$ in $\mathbb{R}^3$.

1. If the symmetric difference, $M_{b_1,c_1,d_1} \Delta M_{b_2,c_2,d_2}$ does not include any of the points $x_1, \ldots, x_k$, then $M_{b_1,c_1,d_1} \cap A = M_{b_2,c_2,d_2} \cap A$.
2. Consider the boxes

$$\{y = (y_1, y_2, y_3) \in \mathbb{R}^3 \,:\, y_1 = x_{j,1}, j = 1, \cdots, k\}$$
$$\cup \{y = (y_1, y_2, y_3) \in \mathbb{R}^3 \,:\, y_2 = x_{j,2}, j = 1, \cdots, k\}$$
$$\cup \{y = (y_1, y_2, y_3) \in \mathbb{R}^3 \,:\, y_3 = x_{j,3}, j = 1, \cdots, k\}\,.$$

Then they are at most $(k+1)^3$ boxes created. If (b_1, c_1, d_1) and (b_2, c_2, d_2) are in the same box then the symmetric difference $M_{b_1,c_1,d_1} \Delta M_{b_2,c_2,d_2}$ does not include any of the points from the set A. Thus, M_{b_1,c_1,d_1} and M_{b_2,c_2,d_2} can pick out different subsets of A only if the points (b_1, c_1, d_1) and (b_2, c_2, d_2) are in different boxes. Therefore, the collection $\mathcal{F} = \{M_{b,c,d},\ c,\, d,\, b \in \mathbb{R}\}$ can pick out at most $(k+1)^3$ subsets of A.
3. Since there are 2^k subsets of A, the collection cannot shutter A if $(k+1)^3 < 2^k$ which holds for $k \geq 11$. This means that no $\{y_1, \cdots, y_{11}\}$ can be shattered. Hence $VC(\mathcal{F}) \leq 11$.

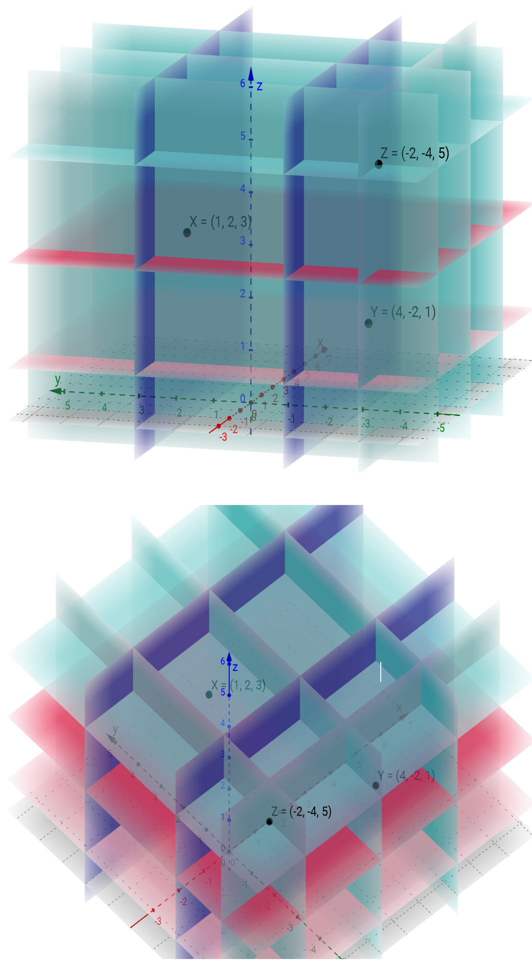

FIG 9.2. *The number of boxes created in $\mathbb{R}^3$ by three points. We have 64 boxes. Figures are produced with Geogebra*

Definition 9.1.4 (VC-subgraph class)**.** *The subgraph of a real valued function $f : \mathsf{E} \to \mathbb{R}$ is the set*

$$S_f = \{(\boldsymbol{x}, s) \in \mathsf{E} \times \mathbb{R} \, : \, s < f(\boldsymbol{x})\}.$$

A class of functions $\mathcal{G}$ is a VC-subgraph class if the set $A = \{S_f, \ f \in \mathcal{G}\}$ is a VC-class.

Examples

Example 9.1.5 (Example C.4.13, [KS20]). The family of indicator functions of sets in a VC-class $\mathcal{F}$ is a VC-subgraph class of functions with the VC-index equal to $\mathrm{VC}(\mathcal{F})$.

1. The family of indicators of half-open intervals: Let $f(x) = \mathbb{1}\{x \leq c\}, \ \ c \in \mathbb{R}$. Then

$$S_f = \{(x, s) \in \mathbb{R} \times \mathbb{R} \, : \, s < f(x)\} = \{(-\infty, c) \times (-\infty, 1) \cup (c, +\infty) \times (-\infty, 0), \ c \in \mathbb{R}\},$$

is a VC-subgraph of index 2. Indeed,

$$S_f = \{(-\infty, c) \times (-\infty, 1) \cup (c, +\infty) \times (-\infty, 0), \ c \in \mathbb{R}\}$$

is a VC-class of index 2. In order to see this, consider $A = \{x_1, x_2\},\ x_1, x_2 \in \mathbb{R}^2$. If $x_i = (x_{i,1}, x_{i,2})$ and $x_{i,2} > 1,\ i = 1, 2$ then x_i cannot be picked out. However, if $x_{i,2} < 0,\ i = 1, 2$ then x_i is always picked out. Thus we consider $x_i = (x_{i,1}, x_{i,2})$ with $x_{i,2} \in [0, 1]$. We can never pick out x_2 (the point with larger 1^{st} coordinate) without picking out x_1. Hence, no two points set can be shattered. Therefore $\mathrm{VC}(\mathcal{F}) = 2$.
In general indicator functions of $(-\infty,\ c)$ or $(c,\ +\infty)$, $c \in \mathbb{R}^d$ are VC-subgraph classes with VC-index $d + 1$.

2. If $f(x) = d\mathbb{1}\{x \le c\},\ \ c, d \in \mathbb{R}$. Then

$$S_f = \{(x, s) \in \mathbb{R} \times \mathbb{R} \,:\, s < f(x)\} = \{(-\infty, c) \times (-\infty, d) \cup (c, +\infty) \times (-\infty, 0),\ c,\ d \in \mathbb{R}\},$$

is a VC-subgraph of index 3. Indeed,

$$\{(-\infty, c) \times (-\infty, d) \cup (c, +\infty) \times (-\infty, 0),\ c,\ d \in \mathbb{R}\}$$

is a VC-class of index 3.
We will not calculate the VC-index precisely, but we will obtain a bound using the **hyperplanes method**. Let

$$\{M_{c,d},\ c,\ d \in \mathbb{R}\} = \{(-\infty, c) \times (-\infty, d) \cup (c, +\infty) \times (-\infty, 0),\ c \in \mathbb{R}\}$$

and $A = \{x_1, \ldots, x_k\}$ a set of k points in $\mathbb{R}^2$.

a) If the symmetric difference, $M_{c_1,d_1} \Delta M_{c_2,d_2}$ (**highlighted area in blue in Figure 9.3**) does not include any of the points $x_1, \cdots, x_k$, then $M_{c_1,d_1} \cap A = M_{c_2,d_2} \cap A$.

b) They are at most $(k+1)^2$ hypercubes created and the collection $\{M_{c,d},\ c,\ d \in \mathbb{R}\}$ can pick out at most $(k+1)^2$ subsets of A. Hence, it cannot shutter A if $(k+1)^2 < 2^k$ which holds for $k \ge 6$. This means that no $\{y_1, \cdots, y_6\}$ can be shattered. Hence $VC(\mathcal{F}) \le 6$.

Example 9.1.6. We say that the class $\mathcal{F} = \{f_s, s \in \mathbb{R}\}$ of functions $f_s : \mathsf{E} \to \mathbb{R}$ is linearly ordered if for all $s < t$ and all $\boldsymbol{x} \in \mathsf{E}$ we have $f_s(\boldsymbol{x}) \le f_t(\boldsymbol{x})$. Any linearly ordered class has VC-index equal to 2.

Let $\mathsf{E} = \mathbb{R}^{\mathbb{Z}}$, $H : \mathsf{E} \to \mathbb{R}$ and $s > 0$. Recall that $\mathcal{C}$ is an anchoring map introduced in Section 2.4. For $\boldsymbol{x} = (\boldsymbol{x}_j, j \in \mathbb{Z}) \in \mathbb{R}^{\mathbb{Z}}$ define

$$H_s^{\mathcal{C}} = H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > s\}\ .$$

Let $0 < s_0 < t_0 < \infty$.

VC index of the class $\{H_s^{\mathcal{C}}, s \in [s_0, t_0]\}$

Consider the class

$$\mathcal{G} = \{H_s^{\mathcal{C}}, s \in [s_0, t_0]\} = \{H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s) = 0\}\mathbb{1}\{|\boldsymbol{x}_0| > s\},\ \ s \in [s_0, t_0]\}.$$

In order to find the VC index of this class, it is enough to identify the situations when the class $\mathcal{G}$ is monotone. Since the map $s \to \mathbb{1}\{|\boldsymbol{x}_0| > s\}$ is decreasing, it remains to discuss the monotonicity of $H(\boldsymbol{x}/s)\mathbb{1}\{\mathcal{C}(\boldsymbol{x}/s) = 0\}$.
Case 1: The anchoring maps $\mathcal{C} = \inf\arg\max$ is 0-homogeneous. If $s \to H_s$ is decreasing then $s \to H_s^{\mathcal{C}}$ is decreasing, hence $\mathcal{G}$ is linearly ordered. Therefore $\mathrm{VC}(\mathcal{G}) = 2$.
In general, when $s \to H_s$ is not decreasing, we cannot say too much, so we focus on some particular cases. Let $H_{s,d}(\boldsymbol{x}) = d\mathbb{1}\{K(\boldsymbol{x}) > s\} = d\mathbb{1}\{(s, \infty)\}(K(\boldsymbol{x}))$, $s \in [s_0, t_0]$, $d \in \mathbb{R}$, and functional $K : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Hence

$$H_{s,d}^{\mathcal{C}}(\boldsymbol{x}) = d\mathbb{1}\{(s, \infty)\}(K(\boldsymbol{x}))\mathbb{1}\{(s, \infty)\}(|\boldsymbol{x}_0|) = d\mathbb{1}\{(s, \infty)\}(K(\boldsymbol{x}) \wedge |\boldsymbol{x}_0|),$$

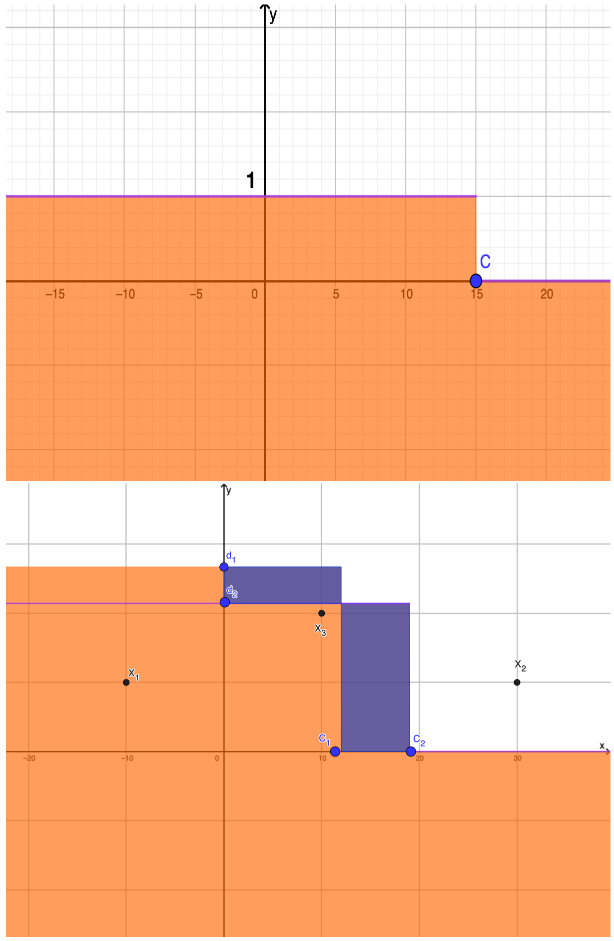

FIG 9.3. *The subgraph, S_f (colored in brown) of $f(x) = \mathbb{1}\{x \le c\}$ (left) and of $f(x) = d\mathbb{1}\{x \le c\}$, $c, d \in \mathbb{R}$ (right) with the symmetric difference, $M_{c_1,d_1} \Delta M_{c_2,d_2}$ (highlighted area in blue)*

then $\mathcal{G}$ has $\mathrm{VC}(\mathcal{G}) = 3$ thanks to Example 9.1.5.
Case 2: We assume that $s \to \mathcal{C}_s$ is decreasing. This is the case of the anchoring map $\mathcal{C}^{(2)}$.
If $s \to H_s$ is decreasing then $s \to H_s^{\mathcal{C}}$ is decreasing, hence $\mathcal{G}$ is linearly ordered. Therefore $\mathrm{VC}(\mathcal{G}) = 2$.
Case 3: We assume that $s \to \mathcal{C}_s$ is increasing and specifically we consider the anchoring map $\mathcal{C}^{(1)}$.
We have $\mathcal{C}^{(1)}(\boldsymbol{x}) = \inf\{j : |\boldsymbol{x}_j| > 1\}$. Hence

$$\mathbb{1}\left\{\mathcal{C}_s^{(1)}(\boldsymbol{x}) = 0\right\}\mathbb{1}\{|\boldsymbol{x}_0| > s\} = \mathbb{1}\left\{\boldsymbol{x}^*_{-\infty,-1} \le s\right\}\mathbb{1}\{|\boldsymbol{x}_0| > s\} = \mathbb{1}\{(-\infty, s) \times (s, \infty)\}\big(\boldsymbol{x}^*_{-\infty,-1}, |\boldsymbol{x}_0|\big)\ .$$

Now, recall from Example 9.1.3, that the class

$$\mathcal{F} = \{M_{c,d},\ c,\, d \in \mathbb{R}\} = \{(-\infty, c) \times (d, +\infty) : c,\, d \in \mathbb{R}\}$$

has the VC-index at most 6. Thus, the class

$$\{(-\infty, s) \times (s, +\infty) : s \in \mathbb{R}\}$$

has also the VC-index at most 6. Thanks to Example 9.1.5, the class

$$\{\mathbb{1}\{(-\infty, s) \times (s, \infty)\}\big(\boldsymbol{x}^*_{-\infty,-1}, |\boldsymbol{x}_0|\big), s \in [s_0, t_0]\}$$

has also the VC-index at most 6.

Chapter 10

References

Bibliography

[CK21] Youssouph Cissokho and Rafal Kulik. Estimation of cluster functionals for regularly varying time series: sliding blocks estimators. *Electronic Journal of Statistics*, 15(1): 2777–2831, 2021.

[CK22] Youssouph Cissokho. Rafal Kulik. Estimation of cluster functionals for regularly varying time series: Runs estimators. *Electronic Journal of Statistics*, 16(1): 3561–3607, 2022.

[BP18] Bojan Basrak and Hrvoje Planinić. Compound poisson approximation for random fields with application to sequence alignment. *Bernoulli Society for Mathematical Statistics and Probability*, 27(2): 1371 – 1408, 2021

[BPS18] Bojan Basrak, Hrvoje Planinić, and Philippe Soulier. An invariance principle for sums and record times of regularly varying stationary sequences. *Probability Theory and Related Fields*, 172(3-4):869–914, 2018.

[BS09] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Processes and their Applications*, 119(4):1055–1080, 2009.

[BS18a] Axel Bücher and Johan Segers. Inference for heavy tailed stationary time series based on sliding blocks. *Electronic Journal of Statistics*, 12(1):1098–1125, 2018.

[BS18b] Axel Bücher and Johan Segers. Maximum likelihood estimation for the fréchet distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462, 2018.

[DH95] Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability*, 23(2):879–917, 1995.

[DHS18] Clément Dombry, Enkelejd Hashorva, and Philippe Soulier. Tail measure and spectral tail process of regularly varying time series. *Annals of Applied Probability*, 28(6):3884–3921, 2018.

[DN20] Holger Drees and Sebastian Neblung. Asymptotics for sliding blocks estimators of rare events. arXiv:2003.01016, 2020.

[DR10] Holger Drees and Holger Rootzén. Limit theorems for empirical processes of cluster functionals. *Annals of Statistics*, 38(4):2145–2186, 2010.

[Has18] Enkelejd Hashorva. Representations of max-stable processes via exponential tilting. *Stochastic Processes and their Applications*, 128(9):2952—2978, 2018.

[Hsi91] Tailen Hsing. Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37(1):117–139, 1991.

[KS20] Rafał Kulik and Philippe Soulier. *Heavy tailed time series.* Springer, 2020.

[MW13] Thomas Mikosch and Olivier Wintenberger. Precise large deviations for dependent regularly varying sequences. *Probability Theory and Related Fields*, 156(3-4):851–887, 2013.

[MW14] Thomas Mikosch and Olivier Wintenberger. 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, 2014.

[MW16] Thomas Mikosch and Olivier Wintenberger. A large deviations approach to limit theorem for heavy-tailed time series. *Probability Theory and Related Fields*, 166(1-2):233–269, 2016.

[PS18] Hrvoje Planinić and Philippe Soulier. The tail process revisited. *Extremes*, 21(4):551–579, 2018.

[RLdH98] Holger Rootzén, Ross M. Leadbetter, and Laurens de Haan. On the distribution of tail array sums for strongly mixing stationary sequences. *Annals of Applied Probability*, 8(3):868–885, 1998.

[SW94] Richard L. Smith and Ishay Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society. Series B. Methodological*, 56(3):515–528, 1994.

[BZ18] Bücher, Axel and Zhou, Chen. A horse racing between the block maxima method and the peak–over–threshold approach. arXiv:1807.00282, 2018.

[FdH15] Ana Ferreira, Laurens de Haan. On the block maxima method in extreme value theory: PWM estimators *Annals of Statistics*, 43(1): 276-298, 2015.

[ZVB20] Nan Zou, Stanislav Volgushev, and Axel Bücher. Multiple block sizes and overlapping blocks for multivariate time series extremes. *Annals of Statistics*, 49(1): 295-320, 2021.