École doctorale de sciences mathématiques de Paris centre
Laboratoire de Probabilités, Statistique et Modélisation - LPSM
Sorbonne Université

Thèse de doctorat

Discipline : Mathématiques
Specialité : Statistiques

présentée par

Gloria Buriticá

Assessing the time dependence of multivariate extremes for heavy rainfall modeling

---

dirigée par: Philippe NAVEAU (LSCE) and Olivier WINTENBERGER (LPSM)

Soutenue le 31 Mai 2022

Composition du Jury :

Prof. Dr. Axel BÜCHER	Heinrich-Heine-Universität Düsseldorf	Examinateur
Prof. Dr. Sebastian ENGELKE	University of Geneva	Examinateur
Prof. Dr. Stéphane GIRARD	Inria Grenoble Rhône-Alpes	Examinateur
Prof. Dr. Anja JANßEN	Otto-von-Guericke-Universität Magdeburg	Rapporteuse
Prof. Dr. Marie KRATZ	ESSEC Business School	Examinatrice
Prof. Dr. Olivier LOPEZ	Sorbonne Université	Examinateur
Prof. Dr. Mathieu RIBATET	École Centrale de Nantes	Rapporteur

École doctorale de sciences mathématiques de Paris centre
Laboratoire de Probabilités, Statistique et Modélisation - LPSM
Sorbonne Université

Thèse de doctorat

Discipline : Mathématiques
Specialité : Statistiques

présentée par

Gloria Buriticá

Assessing the time dependence of multivariate extremes for heavy rainfall modeling

dirigée par: Philippe NAVEAU (LSCE) and Olivier WINTENBERGER (LPSM)

Composition du Jury :

Prof. Dr. Axel BÜCHER	Heinrich-Heine-Universität Düsseldorf	Examinateur
Prof. Dr. Sebastian ENGELKE	University of Geneva	Examinateur
Prof. Dr. Stéphane GIRARD	Inria Grenoble Rhône-Alpes	Examinateur
Prof. Dr. Anja JANßEN	Otto-von-Guericke-Universität Magdeburg	Rapporteuse
Prof. Dr. Marie KRATZ	ESSEC Business School	Examinatrice
Prof. Dr. Olivier LOPEZ	Sorbonne Université	Examinateur
Prof. Dr. Mathieu RIBATET	Ecole Centrale de Nantes	Rapporteur

Laboratoire de Probabilités,
Statistiques et Modélisation.
UMR 8001
Boîte courrier 158
4 place Jussieu
75252 Paris Cedex 05

Sorbonne Université
École doctorale de sciences
mathématiques de Paris centre.
Boîte courrier 290
4 place Jussieu
75252 Paris Cedex 05

Para Mamá y Papá,

"The task of a poet is not to tell what has happened, but what might happen, and what is possible according to probability or necessity."

Aristotle, Poetics.

Agradecimientos

Tres años luego de descubrir mi escritorio actual, prender mi computadora y embarcarme en este proyecto de tesis. Cinco años más tarde de iniciar mi odisea y atravezar el oceano casi olivando las barreras físicas, las barreras imaginarias e inimaginables. Ahora, desde la ciudad de las luces, después del tiempo navegado entre sueños y proyectos, reconozco que el camino lo andamos juntos. Seguido de tanto tiempo les debo el reconocimiento a tantas personas que me han hecho este camino ameno. Este periplo hubiese sido irrelizable si no hubiese contado con ustedes para compartir los días alegres, los días azules, los días de trabajo y los días de fiesta. Gracias a quienes luego menciono y gracias a quienes olvido pero sabrán perdonarme.

Gracias a Olivier y a Philippe. Gracias a Olivier por su tiempo inconmesurable y generoso, por las horas que pasé aprendiendo en tu oficina. Gracias igualmente por exigirme incluso cuando yo lo hubiera creído irrealizable. Gracias por enseñarme a juntar el rigor y la amistad y por toda tu amabilidad. Philippe: gracias por compartir conmigo tu sabiduría y por guiarme hacía las formas prácticas de enfrentar los retos. Te agradezco la palabra justa en el momento justo. Me queda un gran aprecio por los desafíos de intentar implementar una idea a un problema real. Verdaderamente disfruté lo poco que aprendí de la lluvía y el clima, quizá mi influencia trópical o mi tiempo estacionario, me habían hecho pasar por alto lo fantástico de la meteorología. A ambos, les reconozco que este trabajo hubiera sido otro o no hubiese sido posible en absulto sin su afabilidad y su sabiduria.

Gracias a los profesores del jurado por asistir a este evento, por su voluntad y entusiasmo frente al deber de evaluar mi trabajo. Gracias a Thomas por todo su conocimiento, su apoyo, y por las anéctodas. Igualmente agradezco con mucho cariño a todos los profesores que encontré en el camino y que me han aportado de formas infinitas. Gracias también a la comunidad EVA por los, quizá pocos, pero ciertos eventos conviviales. Por Valpred, agradezco a los organizadores por la logística, a los profesores por su amabilidad: Valérie, Anne, Maud, Clément; y a los jóvenes por su energía: Juliette, Paula, Pauline, Philomène, Nicolas, Charles, Ricardo, Fabian, Juro, Olivier, y muchos más a quiénes ahora olvido. Gracias por los momentos que compartimos.

Gracias a papá y mamá por creer en mi incondicionalmente. Gracias por enseñarme a ser valiente pues todo probablemente tiene solución salvo la muerte. Necesité tanta sabiduria que no tengo pero que heredo de ustedes de vez en vez para seguir avanzando piano piano. Gracias por tanto amor siempre. Gracias a mi familia y a mis raíces por enseñarme la alegría, por su apoyo incondicional. Gracias Hugui por tus palabras dulces, tu cariño y tu consejo. A tu lado descubrí un territorio hermoso, complejo y rico en todas sus formas naturales, culturales y lingüisticas. Nadine, Didier, Fafa, es el momento de recordar los meses confinados. Rememoro los primeros días de la primavera que pasamos juntos como una bella ocasión para reir y compartir. Gracias por tanta paz en medio del caos y por toda su ayuda. Gracias a todas las personas que crucé en Colombia, mis amigos, mi infancia bogotana y el paseo mesuno.

Gracias a mi oficina por la diversión y las discuciones. Nicklas y Thibault: después de tantos almuerzos, pause café y el tiempo que compartimos en la oficina y fuera de ella, les deseo mucha suerte en los proyectos que emprenden al final del verano. La sala de nuestra oficina no hubiera sido posible sin ustedes. Gracias a mis colegas: Joseph, Grâce, Cyril y Camila por su simpatía y por los bonitos momentos que compartimos, que reímos, que atravezamos. Gracias al laboratorio. Gracias a los que se fueron antes: Adeline, Nicolas, Erick. Gracias a los que continuan en esta aventura: Franceso, Ludovic, Alice, Ariane, Pierre, Iqraa, Antonio, Miguel. Gracias al equipo GTT: David, Lucas, Joan, Emilien, por haber hecho la experience divertida. Gracias a Hugues por su trabajo, por su ayuda y por su ironía. Gracias al sécrétariat a Louise, Valérie y Nathalie. Por último, gracias a Corentin por hacer este día posible, por hacer navegable el proceso administrativo y mostrarme la luz en muchas ocasiones.

París, 19 de Mayo 2022.

Acknowledgments

Three years after discovering my current desk, turning on my computer and embarking on this thesis project. Five years after starting my odyssey and crossing the ocean, traveling all the physical barriers, the imaginary and unimaginable barriers. Today, thinking about the time navigated between dreams and projects, I recognize that we have walked this path together. After all this time, I owe recognition to all the people who have made this path enjoyable for me. This journey would have been impossible if I had not counted on you to share the happy days, the blue days, the days of work, and the party days. I warmly want to thank all those I will mention later and also those I now forget to mention. I hope you will find it in your heart to forgive me if so.

I start by thanking Olivier and Philippe. Thanks to Olivier for his immeasurable and generous time, and for the hours I spent learning in your office. Thank you also for demanding me even when I would have thought it was unfeasible. Thank you for teaching me to combine rigor and friendship and for all your kindness. Philippe: thank you for sharing your wisdom with me and guiding me towards practical ways to face challenges. I thank you for the right word at the right time. I will keep from you a great appreciation for the challenges of trying to implement an idea to a real problem. I truly enjoyed the little I learned about rain and weather; perhaps my tropical influence or stationary weather notion had made me overlook the fantastic ideas on meteorology. Thank you for helping me discover this field. To both of you, I acknowledge this work would have been different or not possible at all without your friendliness and wisdom.

Many thanks to all the professors of the jury for attending this event. I really appreciated your willingness and enthusiasm in your duty to evaluate my work. Thanks to Thomas for all his knowledge, support, and anecdotes. I also thank with kind affection all the teachers I met along the way who have contributed to my life in infinite ways. I also thank the EVA community for the perhaps few but always fun and convivial events. For Valpred, I thank the organizers for the logistics, the teachers for their amiability: Valérie, Anne, Maud, Clément; and the young people for their energy: Juliette, Paula, Pauline, Philomène, Nicolas, Charles, Ricardo, Fabian, Juro, Olivier, and many more whom I forget now. Thank you for the moments we shared.

Thanks to mom and dad for believing in me unconditionally. Thank you for teaching me to be brave because problems are always likely to have solutions. I needed so much wisdom that I don't have but that I inherit from you from time to time to keep moving forward piano piano. Thank you for so much love always. Thanks to my family and my roots for teaching me joy and for the unconditional support. Many thanks Hugui for your sweet words, love and keen advice every time I needed it. I discovered a beautiful territory at your side, complex and rich in natural, cultural, and linguistic forms. Nadine, Didier, Fafa, it's time to remember the lockdown months. I remember the first days of spring in lockdown as a beautiful occasion to laugh and share. Thank you for surrounding me with plenty of peace during the chaos of pandemia and for all your help. Thanks to all the people I crossed in Colombia, my friends, my childhood in Bogota, and the memorable trips to La Mesa.

Thanks to my office colleagues for all the fun and happy moments. Nicklas and Thibault: after so many lunches, coffee breaks, and time and laughs shared in and out of the office, I wish you the best of luck in the projects you are undertaking at the end of the summer. Without you, our office living room would not have been possible. Thanks to my colleagues: Joseph, Grâce, Cyril, and Camila, for their sympathy and for the lovely moments that we shared. Thanks to the laboratory in general. Thanks to those who left before me: Adeline, Nicolas, Erick. Thanks to those who continue this adventure: Franceso, Ludovic, Alice, Ariane, Pierre, Iqraa, Antonio, Miguel, and the ones I forget to mention. Thanks to the GTT team: David, Lucas, Joan, and Emilien, for having made the experience fun. Thanks to Hugues for his work, his help, and his irony. Thanks to the secretariat: Louise, Valérie, and Nathalie. Finally, thanks to Corentin for making this day possible by making the administrative process navigable and showing me the light on many occasions.

París, 19 de Mayo 2022.

Abstract

Assessing the time dependence of multivariate extremes for heavy rainfall modeling

Nowadays, it is common in environmental sciences to use extreme value theory to assess the risk of natural hazards. In hydrology, rainfall amounts reach high-intensity levels frequently, which suggests modeling heavy rainfall from a heavy-tailed distribution. In this setting, risk management is crucial for preventing the outrageous economic and societal consequences of flooding or land-sliding. Furthermore, climate dynamics can produce extreme weather lasting numerous days over the same region. However, even in the stationary setting, practitioners often disregard the temporal memories of multivariate extremes. This thesis is motivated by the case study of fall heavy rainfall amounts from a station's network in France. Its main goal is twofold. First, it proposes a theoretical framework for modeling time dependencies of multivariate stationary regularly varying time series. Second, it presents new statistical methodologies to thoughtfully aggregate extreme recordings in space and time.

To achieve this plan, we consider consecutive observations, or blocks, and analyze their extreme behavior as their ℓ^p-norm reaches high levels, for $p > 0$. This consideration leads to the theory of p-clusters, which model extremal ℓ^p-blocks. In the case $p = \infty$, we recover the classical *cluster (of exceedances)*. For $p < \infty$, we built on large deviations principles for heavy-tailed observations. Then, we study in depth two setups where p-cluster theory appears valuable. First, we design disjoint blocks estimators to infer statistics of p-clusters, e.g., the *extremal index*. Actually, p-clusters are linked through a change of norms functional. This relationship opens the road for improving cluster inference since we can now estimate the same quantity with different choices of p. We show cluster inference based on $p < \infty$ is advantageous compared to the classical $p = \infty$ strategy in terms of bias. Second, we propose the stable sums method for high return levels inference. This method enhances marginal inference by aggregating extremes in space and time using the ℓ^α-norm, where $\alpha > 0$ is the (tail) index of the series. In simulation, it appears to be robust for dealing with temporal memories and it is justified by the α-cluster theory.

Keywords[1]: *Extreme value theory, regularly varying time series, large deviations, environmental time series*

[1] *Primary 60G70 Secondary 60F10 62G32 60F05 60G57 60G55 60F99 60J10 62M10*

Résumé

Analyse multivariée de la dépendance temporelle des extrêmes pour la modélisation des précipitations sévères

Il est commun dans les sciences de l'environnement d'utiliser la théorie des valeurs extrêmes pour évaluer le risque des dangers naturels. En hydrologie, les précipitations atteignent fréquemment des hauts niveaux d'intensité, ce qui suggère de modéliser les pluies sévères à l'aide d'une distribution à queue lourde. Dans ce contexte, la gestion du risque est cruciale pour prévenir des conséquences économiques et sociétales majeures telles que des inondations ou des glissements de terrain. Par ailleurs, les dynamiques du climat peuvent produire des conditions météorologiques extrêmes pendant plusieurs jours sur la même région. Cependant, dans le cadre stationnaire, les praticiens négligent souvent les dépendances temporelles des extrêmes multivariées. Cette thèse propose un cadre théorique pour la modélisation des dépendances temporelles des séries chronologiques stationnaires à variation régulière et des nouvelles méthodologies statistiques pour agréger les observations spatiotemporelles des extrêmes. Plus précisement, nous développons l'étude de cas sur les précipitations extrêmes d'un réseau météo en France.

Nous considérons des observations consécutives, ou blocs, et analysons leur comportement lorsque leur ℓ^p-norme atteint des niveaux extrêmes, pour $p > 0$. Cette approche conduit à la théorie des p-clusters qui sert à modéliser les ℓ^p-blocs extrémaux. Dans le cas $p = \infty$, nous retrouvons la définition classique du *cluster extrémal*. Pour $p < \infty$, nous nous appuyons sur les principes de grandes déviations pour les observations à queue lourde. Nous approfondissons sur deux cas où la théorie des p-clusters semble pratique. Premièrement, nous proposons des estimateurs de blocs disjoints pour estimer des statistiques des p-clusters, e.g., *l'indice extrémal*. De plus, nous retraçons les p-clusters par une fonctionelle de changement de norme. Cette relation ouvre la voie à une possible amélioration de l'inférence des clusters puisque nous pouvons maintenant estimer la même quantité avec des choix de p différents. Nous montrons que l'inférence des clusters basée sur $p < \infty$ est avantageuse par rapport à la stratégie classique $p = \infty$ concernant le biais. Deuxièmement, nous proposons une méthode pour l'inférence des niveaux de retour extrémaux. Cette méthode améliore l'inférence marginale en agrégeant les extrêmes dans l'espace et dans le temps à l'aide de la norme ℓ^α, où $\alpha > 0$ est l'indice (de queue) de la série. En simulation, cette méthode paraît robuste pour traiter les dépendances temporelles et se justifie par la théorie des α-cluster.

Mots clés[2]: *Théorie des valeurs extrêmes, séries temporelles à variation régulière, grandes déviations, séries temporelles environnementales*

[2] *Primary 60G70 Secondary 60F10 62G32 60F05 60G57 60G55 60F99 60J10 62M10*

Table of Contents

Chapter 1: Introduction

Contents

1.1 Extreme value theory

Extreme value theory (EVT) is the area of statistics studying rare events. This domain calls the attention of many disciplines where risk assessment is necessary, and this feature encapsulates its realm of applications. Some examples are finance, insurance, network traffic, hydrology, geology, metallurgy, and electronics [59, 34, 12, 143]. Among all applications, their common point is their need to quantify how frequently unlikely events occur for designing prevention plans. Risk management of extremes has both a societal and an economic impact. Its ultimate goal is to gauge the magnitude of hazards and infer coverage levels for protection against disasters such as flooding, debris flows, land slides, heatwaves, financial crashes, high tides, network traffic crashes, etc. Overall, the techniques in extreme value statistics provide valuable information for defining risk policies and infrastructure plans attempting to reduce the ruinous consequences of environmental, industrial or economic disasters.

(EVT) groups the main statistical tools and methodologies for inferring the probability that an unusual phenomenon will happen. Naturally, this task entails significant statistical challenges as, by definition, rare events are scarce, and in practice, we rarely or never observe them beforehand. To overcome this statistical issue, both practitioners and theoreticians rely on probability theory and asymptotic analysis. The (EVT) primary goal is to extrapolate beyond observed data by modeling high records, and then to evaluate the uncertainties of this approximation. Typically, we have access to a sample of observations drawn under the same random mechanism. In this stationary setting, rare events lie on the tail of the underlying probability distribution, and we aim to infer/extrapolate the probability tail from the empirical observations. Stating the problem differently, we seek to model and infer large deviations from the distribution's median.

To list a few fields intersecting extreme value theory, we mention asymptotic development, theory of regular variation, measure convergence, and empirical processes theory. Aside from the asymptotic and probabilistic results, we need sharp statistical methodologies to infer the unusual. Commonly, we start by choosing among all observed values those relevant to model the tail probability. As expected, not all experimental recordings are helpful to extrapolate the tail. Once large values are selected, (EVT) proposes a rigorous theoretical background and various statistical methodologies for the purpose of tail inference. Textbooks discuss how to model univariate stationary data with temporal memories [34, 12, 59]. Also, assessing the risk of extremes from a spatial grid or network co-occurring is possible from spatial considerations modeling d-dimensional extremes [12, 80, 143]. However, practitioners often assume i.i.d. data for multivariate inference. I consider extremal temporal dependence in a stationary multivariate setting. I present below the case study of fall daily rainfall levels from a network of weather stations in France to illustrate my main motivation.

1.2 Case study of daily rainfall measurements in France

This thesis aims to develop a solid framework for assessing multivariate extremal temporal dependencies, and derive tactful inference strategies in the stationary setting. One main driver of this thesis is the data set of daily precipitation amounts from a France's network of weather stations. I consider daily rainfall records from 1975 to 2015 obtained from Météo France. Concerning stationarity, I focus on fall observations (September, October, and November) including both wet and dry days. By choosing different regions in France, I can contrast our forthcoming new methodologies on different climates. I have picked three sites featuring oceanic weather in the northwest (Brest, Lanveoc, and Quimper), three sites from the mediterranean in the south (Hyeres, Bormes-les-Mimosas, and Le Luc), and three continental sites in the northeast (Metz, Nancy, and Roville). Figure 1.1 indicates the three chosen regions. Concerning the meteorological dynamics, storms/fronts entail daily high rainfall levels recorded simultaneously over multiple sites and possibly at numerous time lags in a short period. Consequently, the underlying extremal event, let's say a storm, has both a spatial and temporal coverage.

To illustrate this point, we can see in Figure 1.2 that within the same region, fall rainfall records attain high levels with similar intensities. In addition, rainfall in the south is heavier than in the northwest and northeast regions. Also, extreme precipitations are often recorded simultaneously at nearby stations from the same region. As explained, the weather within a region is typically driven by the same meteorological dynamic, let's say, a convective storm, which produces high rainfall levels co-occurring at close locations with similar intensity. Moreover, the extremal dependencies in time are summarized by the temporal extremogram in[1] [44]. The empirical estimates are shown in Figure 1.3, and indicate that consecutive heavy daily rainfall amounts can not be considered

[1]The temporal extremogram for heavy-tailed series is defined over time lags by $t \mapsto \lim_{x\to+\infty} \mathbb{P}(X_t > x \,|\, X_0 > x)$.

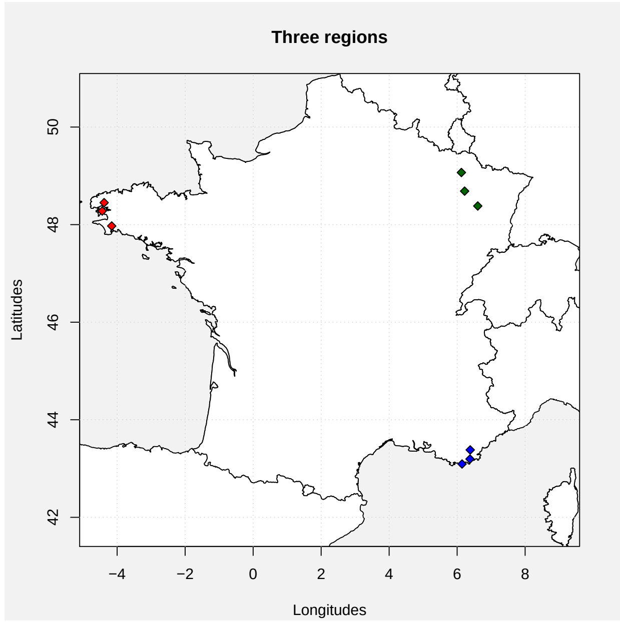


Figure 1.1: Map of France with localization of the weather stations considered in the analysis.

as i.i.d. Hence, the distance proximity of stations from the same region justifies addressing time dependence in a multivariate setting, and thus the main challenge in this framework is how to aggregate information about extremes in space and time.

Aside from the temporal and spatial dependencies, the third feature I want to highlight from this data set is its heavy-tailed nature. Daily rainfall take zero values on dry days, or low values on usual days, but exhibits large deviations from its median with extreme weather conditions. Then, as climate scientists and hydrology experts agree, the significant intensity levels reached by heavy rainfall are appropriately modeled with heavy-tailed distributions; cf. [102, 37, 126]. More precisely, in this stationary setting, I will model daily heavy rainfall at each station, denoted by X, using a survival function with polynomial decay, i.e., $\mathbb{P}(X > x) = \overline{F}(x) = x^{-\alpha}\ell(x)$, where $\alpha > 0$ is a (tail) index and $x \mapsto \ell(x)$ is a positive function verifying, for all $t > 0$, $\lim_{x\to+\infty} \ell(tx)/\ell(x) = 1$. In this heavy-tailed setting, my goal is to infer the extremal properties of the general mechanism driving heavy rainfall amounts while assessing uncertainties. For example, I would like to address high return level (quantile) estimation. My main claim is that integrating the temporal memories and spatial ties should enhance inference of confidence intervals. Indeed, uncertainties assessment is highly important for this case study. Erroneous heavy rainfall risk evaluation can have a considerably negative socio-economical impact as rare events are of significant magnitude in terms of flooding, debris flows or landslides.

Chapter 4 analyzes this data set with the new forthcoming methodologies. My goal in Chapter 4 is to compute high quantiles of the distribution of heavy rainfall in a stationary setting. The reason to introduce this case study here, is to contextualize the main objectives of this thesis that I now state below. Of course, this data set motivates the analysis, but the new methodologies cover further contexts based on stationary multivariate heavy-tailed time series.

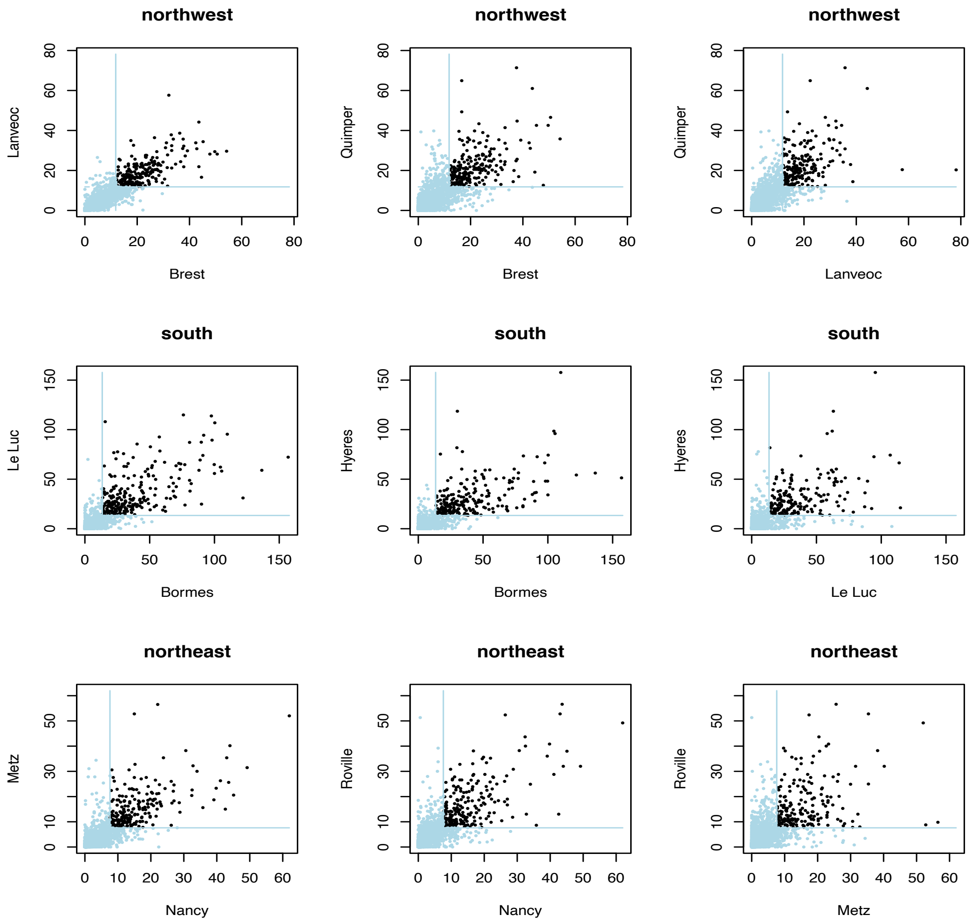


Figure 1.2: Scatter plots of fall daily rainfall in France from 1976 to 2015. The top, middle, and bottom panels refer to three climatological regions: continental (northwest), oceanic (west), and mediterranean (south), respectively. Simultaneous exceedances of the 95-th order statistic of the sample of daily maxima of a region are in black.

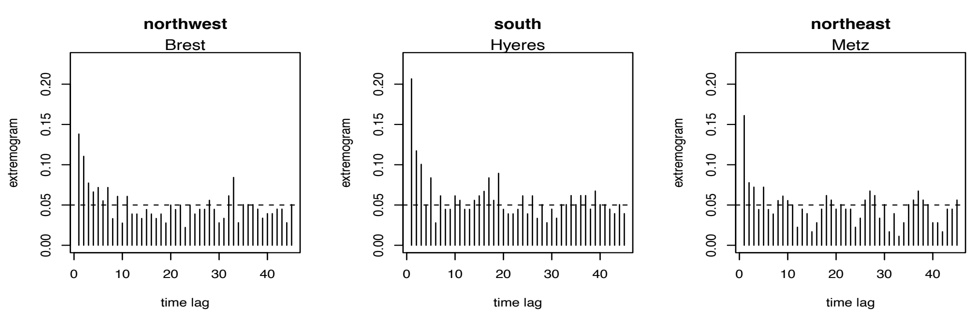


Figure 1.3: Empirical temporal extremogram of the 95-th order statistic of fall daily maximum rainfall record within regions in France from 1976 to 2015. The first, middle and last correspond to three climatological regions as in Figure 1.2. As a baseline, the extremogram takes the value pointed by the dotted line on independent time lags.

1.3 Goals of this thesis.

Consider $(\mathbf{X}_t)$ to be a multivariate stationary time series with heavy-tailed behavior; see Appendix B. In this setting, the objectives of this thesis are twofold. First, it aims to establish a theoretical framework for modeling extremal temporal dependence. Second, it aims to investigate statistical methods for multivariate extremes that account for the time dependencies. More precisely, in what follows I will address the following questions written in (**Quest. 1-4**).

Quest. 1. How can we model an extreme episode lasting for several time units in a multivariate setting?

Quest. 2. How can we detect extreme episodes from all observations in a sample-based approach?

Quest. 3. How can we infer the spatiotemporal features of extreme episodes?

Quest. 4. How can we compute high return levels and quantify its uncertainties while integrating the time and space dependencies?

1.4 Outline of the thesis

This thesis aims to assess the extremal time dependence of multivariate heavy-tailed extremes. To do so, I will address questions (**Quest. 1-4**) as follows. In the context of stationary heavy-tailed time series, extreme episodes can happen as short periods with numerous high recordings. Chapter 2 presents new large deviation principles of consecutive observations, or blocks, with a large ℓ^p-norm. It relies on extreme value theory and large deviations of sums of heavy-tailed time series. The asymptotics of extremal ℓ^p-blocks give a rigorous definition of p-clusters taking values in ℓ^p. Taking advantage of the p-clusters theory, Chapter 2 suggests to capture the characteristics of short periods with extremal behavior through p-clusters to answer (**Quest. 1**). The second part of Chapter 2 and Chapters 3, 4, 5, detail two applications of the new large deviation principles of extremal ℓ^p-blocks and p-cluster theory. For the first application, Chapter 2 develops an inference methodology based on extremal ℓ^p-blocks to estimate p-cluster features. It proposes new consistent disjoint blocks estimators for cluster inference which address (**Quest. 2-3**). Asymptotic normality of p-cluster blocks estimators is proven in Chapter 5. Chapter 3 will revisit the *extremal index* introduced in [109, 110], which can be interpreted as a summary of the extremal time dependencies. The extremal index has a reinterpretation using the α-cluster, where $\alpha > 0$ is the (tail) index. Chapter 3 studies cluster-based extremal index inference built on extremal ℓ^α-blocks, which further illustrates my contribution to (**Quest. 3**). Chapter 4 shows an application of the new large deviation principles to high quantile inference for heavy-tailed spatiotemporal data sets, and addresses in depth (**Quest. 4**). It presents the stable sums method for high return levels inference, which proposes to aggregate spatiotemporal extremes with the ℓ^α-norm and is justified by the α-cluster theory. Finally, Chapter 4 computes return levels for the case study of daily rainfall previously introduced in Section 1.2.

In the remaining of this chapter, I outline my contributions and publications, and sketch how they tackle the queries in (**Quest. 1-4**). Finally, appendices A and B provide a toolbox of background results needed for this thesis. Appendix A analyzes heavy daily rainfall in France, from the data set introduced Section 1.2, with classical tools in extremes. It also aims to underline the limitations of the most common practices. To simultaneously model the spatiotemporal aspects of our data, Appendix B presents a state-of-the-art framework of stationary heavy-tailed time series and complements the discussion on the upcoming new results and methodologies.

1.5 Personal contributions

1.5.1 Notation

The main theoretical contribution of this thesis is modeling extremal ℓ^p–blocks, i.e., short periods with large ℓ^p-norm. Let's start by recalling the definition of the sequential space ℓ^p. For $p \in (0, +\infty]$, consider the p-modulus $\| \cdot \|_p : (\mathbb{R}^d)^{\mathbb{Z}} \to [0, +\infty]$ defined by

$$\|\mathbf{x}_t\|_p \;=\; \big(\textstyle\sum_{t\in\mathbb{Z}}|\mathbf{x}_t|^p\big)^{1/p}, \tag{1.5.1}$$

and consider the space (ℓ^p, d_p) of sequences with finite p-modulus, where d_p is the metric induced by the p-modulus. If $p = \infty$, then $\| \cdot \|_\infty$ is the supremum norm. The $p-$modulus is a norm if $p \in [1, +\infty)$, and, abusing notation, we refer to $\| \cdot \|_p$ as the ℓ^p-norm for all $p \in (0, +\infty]$. The ℓ^p-norms are not equivalent, and they satisfy

$$\| \cdot \|_p \;\geq\; \| \cdot \|_{p'} \;\geq\; \| \cdot \|_\infty \,,$$

for all $p, p' > 0$ with $p < p'$. The supremum norm of a fixed sequence is also the monotone limit as $p \to +\infty$ of its ℓ^p-norm.

More generally, we write vectors $\mathbf{x} \in \mathbb{R}^d$ in bold, $\mathbb{R}^d$-valued time series as $(\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$. We sometimes use the notation $\mathbf{x}_{[a,b]}$, or $(\mathbf{x}_t)_{t=a,\dots,b}$, to write the vector $(\mathbf{x}_a, \dots, \mathbf{x}_b) \in (\mathbb{R}^d)^{b-a+1}$, for $a, b \in \mathbb{Z}$, $a \leq b$. For sequences $(x_n), (y_n) \in (\mathbb{R})^{\mathbb{Z}}$, the asymptotic relation $(x_n) \sim (y_n)$, as $n \to +\infty$, is interpreted as $x_n/y_n \to 1$, as $n \to +\infty$. We write random vectors in capital letter as $\mathbf{X}$ takes values in $\in \mathbb{R}^d$. We denote weak convergence of probability measures by $\xrightarrow{w}$, convergence in distribution of random processes by $\xrightarrow{d}$, convergence in probability of random variables by $\xrightarrow{\mathbb{P}}$. For random objects, almost sure relations are abbreviated as a.s.. Independent identically distributed random vectors is abbreviated by i.i.d.

1.5.2 Heavy-tailed stationary time series

Consider $(\mathbf{X}_t)$ to be a stationary time series taking values in $\mathbb{R}^d$, endowed with a norm $| \cdot |$. Assume it is heavy-tailed in the following sense: there exists $\alpha > 0$, and a time series (Θ_t), satisfying

$|\Theta_0| = 1$ a.s., such that for all $y > 0$, $h = 0, 1, \ldots,$

$$\mathbb{P}(|\mathbf{X}_0| > yx, \mathbf{X}_{t=\pm 0,1,\ldots,h}/|\mathbf{X}_0| \in \cdot \,|\, |\mathbf{X}_0| > x) \xrightarrow{w} y^{-\alpha}\mathbb{P}(\Theta_{t=\pm 0,1,\ldots,h} \in \cdot), \quad x \to +\infty. \tag{1.5.2}$$

We say in this case that $(\mathbf{X}_t)$ is regularly varying with (tail) index $\alpha > 0$ and spectral tail measure (Θ_t); see Appendix B.5. The time series (Θ_t) captures space and time extremal features over finite windows of time. To capture the full duration of spatiotemporal extremes, my goal is to let $h \to +\infty$ in (1.5.2) simultaneously as $x \to +\infty$. My aim is to model the extremal features of consecutive observations, or blocks, defined by

$$\mathbf{X}_{t=1,\ldots,n} \;=\; \mathbf{X}_{[1,n]},$$

as $n \to +\infty$. To pursue this approach, we must embed the blocks $\mathbf{X}_{[1,n]}$ into a suitable sequential space. I will study the extremal features of blocks in the space (ℓ^p, d_p).

1.5.3 Large deviations of extremal ℓ^p-blocks

To model extremal attributes of blocks $\mathbf{X}_{[1,n]}$ in (ℓ^p, d_p), as $n \to +\infty$, we must define what extreme means in (ℓ^p, d_p). Naturally, we say a block is extreme if its ℓ^p-norm is large. The main challenge then is to find suitable assumptions such that $\mathbb{P}(\mathbf{X}_{[1,n]}/x_n \in \cdot \,|\, \|\mathbf{X}_{[1,n]}\|_p > x_n)$ admits a limit, where (x_n) is a suitable sequence satisfying $\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_p > x_n) \to 0$, as $n \to +\infty$. In this case, we say (x_n) is a extremal sequence for $\mathbf{X}_{[1,n]} \in (\ell^p, d_p)$. In Chapter 2, Theorem 2.2.1 introduces anti-clustering and vanishing small values conditions: $\mathbf{AC}, \mathbf{CS}_p$, driving the sequence (x_n), such that, for $y > 0$,

$$\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_p > yx_n, \mathbf{X}_{[1,n]}/\|\mathbf{X}_{[1,n]}\|_p \in A \,|\, \|\mathbf{X}_{[1,n]}\|_p > x_n) \xrightarrow{w} y^{-\alpha}\mathbb{P}(\mathbf{Q}^{(p)} \in A), \quad n \to +\infty. \tag{1.5.3}$$

where $\mathbf{Q}^{(p)}$ takes values in ℓ^p, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., this limit holds for suitable shift-invariant continuity sets $A \subset (\ell^p, d_p)$, and $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$. We say in this case that the time series $(\mathbf{X}_t)$ admits a p-cluster $\mathbf{Q}^{(p)}$. We stress the close similarity between equations (1.5.3) and (1.5.2). Also, notice sum-type functionals and norms are examples of the shift-invariant sets we can consider in (1.5.3).

We call the limit in (1.5.3) a large deviation principle of extremal ℓ^p-blocks. Indeed, for $p < \infty$, notice that if (1.5.3) holds and (x_n) is a sequence such that $\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_p > x_n) \to 0$, as $n \to +\infty$, then necessarily

$$\textstyle\sum_{t=1}^{n} |\mathbf{X}_t/x_n|^p \;=\; \|\mathbf{X}_{[1,n]}/x_n\|_p^p \xrightarrow{\mathbb{P}} 0, \quad n \to +\infty. \tag{1.5.4}$$

The left-hand side of Equation (1.5.4) is a sum of regularly varying increments $|\mathbf{X}_t|^p$ with (tail) index α/p. Therefore, we can study the large deviations of this sum under the classical large deviation's literature appealing to the work in A.V. and S.V. Nagaev [125, 124] and Cline and Hsing [33], for i.i.d. time series, and Mikosch and Wintenberger [121, 123], for dependent heavy-tailed time series. Moreover, notice (1.5.4) holds if the sequence (x_n) satisfies $n\mathbb{E}[|\overline{\mathbf{X}_1/x_n}^1|^p] \to 0$, as $n \to +\infty$, with the truncation notation where $\overline{\mathbf{x}}^1 = \mathbf{x}\mathbb{1}(|\mathbf{x}| \leq 1)$. This last relation points straightforwardly to one of the main restrictions we impose on (x_n). We assume $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, for some $\kappa > 0$, thus $n\mathbb{E}[|\overline{\mathbf{X}_1/x_n}^1|^p] \to 0$, as $n \to +\infty$, which yields (1.5.4). This last relation follows by the Karamata's Theorem [100].

To sum up, Chapter 2 gives a theoretical framework to model extremal ℓ^p-blocks. The theory of p-clusters, which arises from Equation (1.5.3) is also developed therein. Formalizing the limit in (1.5.3) in Theorem 2.2.1 is my main contribution to **(Quest. 1)**. Actually, extremal ℓ^∞–blocks were previously considered in [40, 10, 9], and lead to the theory of *clusters (of exceedances)*, which coincide with our ∞-cluster defined by letting $p = \infty$ in Equation (1.5.3). From a theoretical point of view, it has been already challenging to formalize the definition of clusters (of exceedances); see [65, 153, 154, 10, 9], and references there in. However, the underlying idea of p-clusters has proven to be a very useful tool. For example, the considerations in [40] show how to model blocks $\mathbf{X}_{[0,n]}/a_n$ using the ∞-cluster, where (a_n) are moderate levels satisfying $\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_\infty > a_n) \to c$, for some positive constant $c > 0$, such that $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$, as $n \to +\infty$. Furthermore, [40] used then the limit representation of $\mathbf{X}_{[1,n]}/a_n$, as $n \to +\infty$, to state a central limit theory for heavy-tailed stationary time series.

Actually, we claim that to model blocks $\mathbf{X}_{[1,n]}/a_n$, with $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$, as $n \to +\infty$, it is advantageous to use the α-cluster rather than the ∞-cluster's approach for one main reason. The ∞-cluster, defined with respect to the ℓ^∞-norm, is in tight correspondance with the *extremal index* of the sequence $(\mathbf{X}_t)$, that we denote $\theta_{|\mathbf{X}|}$; see [109, 110]. Instead, using the α-cluster we can remove this parameter from the limit representation of $\mathbf{X}_{[1,n]}/a_n$, as $n \to +\infty$, in the following way. Arguing as in [40], we model $\mathbf{X}_{[1,n]}/a_n$ as a concatenation of independent extreme episodes, or p-clusters, such that (a_n) are moderate levels satisfying $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$, as $n \to +\infty$. Roughly speaking, we introduce an intermediate integer sequence (b_n), we denote $m_n = \lfloor n/b_n \rfloor$, and we define disjoint blocks as

$$\underbrace{\mathbf{X}_{[1,b_n]}}_{:=\mathcal{B}_1}, \underbrace{\mathbf{X}_{[b_n+1,2b_n]}}_{:=\mathcal{B}_2}, \dots, \underbrace{\mathbf{X}_{[n-b_n+1,n]}}_{:=\mathcal{B}_{m_n}}. \tag{1.5.5}$$

satisfying $x_{b_n} = a_n$, thus $n/b_n \to +\infty$, as $n \to +\infty$. Then, under sufficient mixing conditions like $\mathcal{A}(x_{b_n})$; see Appendix B.4, the subsample of disjoint blocks $(\mathcal{B}_t/x_{b_n})_{t=1,\dots,m_n}$ defined from Equation (1.5.5), is asymptotically independent, as $n \to +\infty$. This last observation yields the modeling strategy in [40] such that we model observations $\mathbf{X}_{[1,n]}/a_n$ as a concatenation of independent p-clusters, as $n \to +\infty$. However, in this setting we must keep track of two things to trace the limit representation of $\mathbf{X}_{[1,n]}/a_n$. First, we must model the limit of extremal ℓ^p-blocks: $\mathcal{B}_1/x_{b_n} = \mathbf{X}_{[1,b_n]}/x_{b_n}$, which we do using the p-clusters, and second, we care about the probability of hitting a p-cluster given by $\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n})$, as $n \to +\infty$. This last probability is described below by the asymptotics in (1.5.6), for $p = \infty$, and in (1.5.7), for $p = \alpha$.

For $p = \infty$, we mentioned that the ∞-cluster is linked to the extremal index $\theta_{|\mathbf{X}|}$. Actually, note the extremal index exists for time series with short-range memories, and that it can be interpreted as a summary statistic of the extremal time dependencies as follows. Under classical anti-clustering conditions like **AC** in Theorem 2.2.1, preventing the long-range dependence of extremes, the extremal index satisfies $\theta_{|\mathbf{X}|} \in (0, 1]$ and

$$\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_\infty > x_{b_n}) \;\sim\; \theta_{|\mathbf{X}|}\, b_n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty. \tag{1.5.6}$$

where (x_n) is as in (1.5.3). Instead, for our ℓ^α-norm approach, we show in Theorem 2.2.1 that, for $p = \alpha$, under anti-clustering and vanishing small values conditions **AC**, $\mathbf{CS}_\alpha$, stated therein, the following relation holds

$$\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_\alpha > x_{b_n}) \;\sim\; b_n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty. \tag{1.5.7}$$

Both, conditions **AC**, $\mathbf{CS}_\alpha$, in Theorem 2.2.1 typically hold for short-range memory time series. Condition $\mathbf{CS}_\alpha$ is tailored for neglecting the small values of sums in extreme blocks. In Chapter 2, Section 2.5 discusses these conditions thoroughly. Recall our goal to model $\mathbf{X}_{[1,n]}/a_n$ as a concatenation of independent p-cluster. In this case, the aforementioned modeling strategy based on $p = \alpha$ only relies on the α-cluster due to Equation (1.5.7). Indeed, both (1.5.6) and (1.5.7) compare the series behavior with the i.i.d. case. Roughly, we record extremal ℓ^α-blocks with the same probability for short-range memory series than we do for i.i.d. times series. This is not the case for extremal ℓ^∞-blocks with $\theta_{|\mathbf{X}|} < 1$. We interpret Equation (1.5.7) as a robust feature of the time series with respect to time dependencies.

Then, continuing the argument as in [40], we show in Theorem 3.3.8 the following empirical point process convergence towards a cluster Poisson process

$$N_n \;=\; \sum_{t=1}^{n} \epsilon_{\mathbf{X}_t/a_n} \;\xrightarrow{d}\; N \;=\; \sum_{i=1}^{\infty}\sum_{t\in\mathbb{Z}} \epsilon_{\Gamma_i Q^{(\alpha)}_{t,i}}\,, \quad n \to +\infty. \tag{1.5.8}$$

where (a_n) is a sequence of moderate levels satisfying $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$, $(\Gamma_i)_{i=1,\dots,\infty}$, $(\mathbf{Q}^{(\alpha)}_{t,i})_{i=1,\dots,\infty}$ are both i.i.d. sequences, independent between them, (Γ_i) are the points of a homogenous Poisson point process on $(0, \infty)$ with intensity measure $d(-y^{-\alpha})$, and $(\mathbf{Q}^{(\alpha)}_{t,1})$ is distributed as the α-cluster. Reviewing this approach, we see from (1.5.8) that we can fully identify the cluster component in the limit in terms of the α-cluster. We refer to Appendix B.6 for further discussion and references. Moreover, we mentioned that modeling the asymptotics of blocks $\mathbf{X}_{[1,n]}/a_n$ can be helpful to show further asymptotic results. Indeed, by modeling blocks $\mathbf{X}_{[1,n]}/a_n$ through α-clusters we can prove the central limit theorem of partial sums: $\sum_{t=1}^n \mathbf{X}_t/a_n$, towards a stable distribution, as $n \to +\infty$. We state central limit theory of partial sums in Proposition 2.4.4, for $\alpha \in (0, 1) \cup (1, 2)$, and in Theorem 4.7.1, for $\alpha = 1$.

1.5.4 Statistical applications

We explained in Section 1.5.3 that for modeling purposes, the α-cluster is a good candidate for tracking the asymptotics of moderate observations. In this section, we argue that the strategy of considering extremal ℓ^α–blocks opens a new road for improving inference procedures for heavy-tailed stationary time series. For inference purposes, consider observations $\mathbf{X}_{[1,n]}$ and let (b_n) be again an intermediate integer sequence such that $b_n \to +\infty$ and $n/b_n \to +\infty$, as $n \to +\infty$. In a statistical setting, our main motivation for studying extremal ℓ^p–blocks, with $p < \infty$, is the following. Notice that as p decreases, the p-norms increase, and thus the probability of recording an extremal ℓ^p–block increase also. Actually, under the assumptions of Theorem 2.2.1 for p, there exists a non-increasing function $p \mapsto c(p)$, for $p > 0$, such that

$$\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n}) \sim c(p)\, b_n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty, \tag{1.5.9}$$

where $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$, and $p \mapsto c(p)$ satisfies

$$1 \;=\; c(\alpha) \;\geq\; c(p) \;\geq\; c(\infty) \;=\; \theta_{|\mathbf{X}|}; \tag{1.5.10}$$

see (1.5.6), (1.5.7). Hence, $c(p)$ summarizes the temporal memories of the time series for the ℓ^p-norm when compared to the i.i.d. setting. Actually, an i.i.d. sequence has extremal index one. Hence, from (1.5.9) and (1.5.10), we read that the number of extremal ℓ^∞-blocks available in a sample-based scenario is proportional to the extremal index, compared to the i.i.d. case. Instead, the extremal ℓ^α-blocks happen with the same probability than in the i.i.d. case. For this reason, the ℓ^α-blocks seem to be a robust solution for evaluating spatiotemporal extremes.

We develop statistical methodologies to enhance both cluster inference and marginal inference. Concerning cluster inference, note that Theorem 2.2.1 holds for $p > \alpha$ solely under **AC** and $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ such that this resctrictions determine the asymptotics of (x_n) . Equation (1.5.9) indicates that fewer extremal ℓ^∞-blocks are available in a sample compared to the number of extremal ℓ^p-blocks, with $p < \infty$. Then, estimation of cluster features using extremal ℓ^p-blocks should improve inference as we take $p \downarrow \alpha$. For $p = \alpha$, we also require condition $\mathbf{CS}_\alpha$ but in this case $c(\alpha) = 1$. This is advantageous since now the proportion of extremal ℓ^α blocks remains constant compared to the i.i.d. case, thus the estimation procedure is robust to time dependencies. We formalize this argument in the second part of Chapter 2 and further illustrate the advantages of ℓ^p-cluster-based inference in Chapter 3, for $p < \infty$. We detail briefly our main contributions to p-cluster inference in Section 1.5.5 below. Concerning marginal inference, our goal is to estimate features of the distribution of $\mathbf{X}_1$. In this case, the relationship between the extremal ℓ^∞-blocks and the extremal index also troubles the implementation of classical statistical methods in (EVT), tailored for i.i.d. observations. For example, the common block-maxima and exceedances approach for high quantile inference, built on Pareto models, must be corrected when applied to stationary time-dependent observations to obtain consistent estimates [60, 63]. Typically, estimating the extremal index is a necessary additional step to correct the asymptotics; cf. Appendix A. Instead, for high quantile inference, we can extrapolate the distribution of $\mathbf{X}_1$ using (1.5.7). The main idea is to aggregate short periods through the ℓ^α-norm and then model the α-power sums from a stable distribution. This seems to be a robust solution to account for the temporal memories. We will present our stable sums algorithm for marginal inference in Chapter 4. We explain in short the underlying idea in Section 1.5.6 below.

In what follows, it will be convenient to relate the p-clusters among them to compare the inference procedures tuned with different values of p. Recall the spectral tail process (Θ_t) in (1.5.2). In Proposition 2.3.1 we state that, provided $\|\Theta\|_\alpha + \|\Theta\|_p < +\infty$ a.s., and $\mathbb{E}[\|\Theta\|_p^\alpha / \|\Theta\|_\alpha^\alpha] < +\infty$, then p-clusters can be related as follows

$$\mathbb{P}(\mathbf{Q}^{(p)} \in \cdot) \;=\; c(p)^{-1}\mathbb{E}\big[\tfrac{\|\Theta\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\,\mathbb{1}(\Theta/\|\Theta\|_p \in \cdot)\big]\,, \tag{1.5.11}$$

such that $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., $c(p)$ coincides with (1.5.9) and

$$c(p) \;=\; \mathbb{E}\big[\|\Theta\|_p^\alpha / \|\Theta\|_\alpha^\alpha\big].$$

Moreover, $\|\Theta\|_\alpha < \infty$ a.s. holds if $\lim_{t\to+\infty} |\Theta_t| = 0$, i.e., for time series with short-range memory. It follows straightforwardly that $c(\alpha) = 1$ and the α-cluster satisfies

$$\mathbf{Q}^{(\alpha)} \;\stackrel{d}{=}\; \Theta/\|\Theta\|_\alpha.$$

1.5.5 Cluster inference

Concerning cluster inference, we can summarize the main features of extreme episodes with spatiotemporal coverage through cluster statistics. To address **(Quest. 3)**, we propose consistent disjoint blocks estimator based on extremal ℓ^p-blocks, for p-cluster inference. Our methodology uses order statistics of the ℓ^p-norms sample as empirical thresholds. Roughly speaking, for cluster inference we consider an intermediate integer sequence (b_n), the disjoint blocks defined in (1.5.5), and we write $m = m_n = \lfloor n/b_n \rfloor$. Then, for suitable continuous bounded shift-invariant functionals $f : \ell^p \to \mathbb{R}$, we define the statistic $f_p^{\mathbf{Q}}$ by

$$f_p^{\mathbf{Q}} \;=\; \mathbb{E}[f(\mathbf{Q}^{(p)})].$$

We propose to estimate the statistic $f_p^{\mathbf{Q}}$ by

$$\widehat{f_p^{\mathbf{Q}}} \;:=\; \frac{1}{k}\sum_{t=1}^{m} f\big(\mathcal{B}_t/\|\mathcal{B}_t\|_p\big)\,\mathbb{1}(\|\mathcal{B}_t\|_p > \|\mathcal{B}_t\|_{p,(k)}), \tag{1.5.12}$$

where $\|\mathcal{B}\|_{p,(1)} \geq \cdots \geq \|\mathcal{B}\|_{p,(m)}$ and

$$k \;:=\; k_n \;=\; \lfloor m_n \mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n})\rfloor. \tag{1.5.13}$$

Then, Theorem 2.4.1 states that, under suitable mixing conditions, $\|\mathcal{B}\|_{p,(k)}/x_{b_n} \xrightarrow{\mathbb{P}} 1$ and

$$\widehat{f_p^{\mathbf{Q}}} \xrightarrow{\mathbb{P}} f_p^{\mathbf{Q}}, \qquad n \to +\infty.$$

Furthermore, in Theorem 5.2.1 we state sufficient conditions yielding

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

One of the big challenges for implementing the estimators in (1.5.12) is to determine how to choose the sequence (k_n) in practice. The choice of $k = k_n$ points to the classical bias-variance trade-off in extreme value statistics (see Resnick [143]), where selecting k very large reduces the variance but it might increase the bias. This difficulty in choosing k appeals to **(Quest. 2)** since we have to choose a correct proportion of extremal blocks out of a sample to guarantee unbiased estimates. Furthermore, notice that in our setting the sequence (k_n) also depends on p, and, from Equation (1.5.13) and Equation (1.5.9), $k_n = k_n(p)$ must satisfy

$$k_n(\infty) \sim c(\infty)\, n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \leq c(p)\, n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \sim k_n(p), \quad n \to +\infty, \tag{1.5.14}$$

such that $p \mapsto c(p)$ is the non-increasing function in (1.5.9) and (1.5.10). Then, it follows straightforwardly from Equation (1.5.14) that inference based on extremal ℓ^p-blocks as in (1.5.12), for $p < \infty$, justifies taking k larger compared to inference built on extremal ℓ^∞-blocks. Moreover, note that the same quantity $f_p^{\mathbf{Q}}$ can be obtained using different pairs $p', f_{p'}$, thanks to the cluster representation in (1.5.11). In this case, we must keep in mind to multiply f by the change of norms derivative to obtain $f_{p'}$. We further illustrate in Chapter 2 that inference based on extremal ℓ^p-blocks with $p < \infty$ can be advantageous compared to the classical approach letting $p = \infty$, since for this last approach choosing k can be more challenging. Consistency of the disjoint blocks estimators in (1.5.12) is proven in Chapter 2, and asymptotic normality in Chapter 5.

As an example, we can apply cluster-based inference to estimate the extremal index. Interpretation of the extremal index through the ∞-cluster features is the basis for inference procedures from the early 1990s [160]. The main idea in this case is to infer the extremal index from the Equation (1.5.6). This strategy motivated the so-called extremal index *blocks estimator* in [88, 161], based on counts of exceedances per cluster, and also the *runs estimator* [161, 65], based on interexceedances lengths. In our setting, we can deduce from (1.5.11) the relation

$$\theta_{|\mathbf{X}|} = c(\infty) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha], \tag{1.5.15}$$

such that $\mathbf{Q}^{(\alpha)}$ is the α-cluster. Equation (1.5.15) suggests to use extremal ℓ^α-blocks to infer the extremal index. Letting $p = \alpha$ and $f = f : (\mathbf{x}_t) \mapsto \|\mathbf{x}_t\|_\infty^\alpha / \|\mathbf{x}_t\|_\alpha^\alpha$ in (1.5.12), we deduce an α-cluster-based consistent estimator $\widehat{\theta}_{|\mathbf{X}|}^{\mathbf{Q}}$ of the extremal index. We review the extremal index, and assess α-cluster-based inference of the estimator $\widehat{\theta}_{|\mathbf{X}|}^{\mathbf{Q}}$ in Chapter 3.

Furthermore, a theoretical framework for proving asymptotic normality of cluster-based disjoint blocks estimators was first established in [52]. Actually, [52] treated in detail estimation of the aforementioned extremal index so-called blocks estimator in [88, 161]. Notice that in general for disjoint blocks-type estimation, we can start the series at a lagged time unit and compute a possibly different estimate. We can then take means of the disjoint blocks estimator computed over lagged observations aiming to improve inference. This strategy yields to the sliding blocks estimators. Asymptotic normality of the cluster-based sliding blocks estimators was studied in [51, 28]; see also [108]. Variance computations indicate cluster-based sliding blocks estimators typically have the same asymptotic variance than disjoint blocks estimators. Furthermore, already in [39, 50], the authors discussed how to improve cluster inference using the (tail) index of the time series $\alpha > 0$. Our α-cluster theory gives a solid theoretical framework to further study this strategy exploiting the information on the (tail) index α.

1.5.6 Marginal inference

Concerning marginal inference, Chapter 4 introduces the stable sums method for inferring multivariate high quantiles. Algorithm 1 summarizes the new inference methodology which aims to address (**Quest. 4**). We claim that looking at ℓ^α-norms rather than looking at ℓ^∞-norms, can be seen as a robust approach for aggregating the spatiotemporal information within an extremal block. Consider observations $\mathbf{X}_{[1,n]}$, an integer sequence (b_n) satisfying $n/b_n \to +\infty$, as $n \to +\infty$, and the disjoint blocks defined in (1.5.5). We deduce from (1.5.2) and (1.5.9) the following asymptotic approximation

$$\mathbb{P}(X_{0,j} > x_{b_n}) \ \sim \ m(j)\,(b_n\, c(p))^{-1}\mathbb{P}(\|\mathbf{X}_{[1,b_n]}\|_p > x_{b_n}), \quad n \to +\infty, \tag{1.5.16}$$

such that $X_{0,j}$ is the j-th coordinate of $\mathbf{X}_0$, for $j = 1, \dots, d$, and $m(j)$ is a positive constant satisfying

$$\mathbb{P}(X_{0,j} > x_{b_n}) \ \sim \ m(j)\,\mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty,$$

where $m(j) = \mathbb{E}[(\Theta_{0,j})_+^\alpha]$ tracks the spatial features of the vector $\mathbf{X}_0$ with respect to the j-th coordinate, and does not depend on the temporal aspects. Then, (1.5.16) suggests to aggregate the d-variate observations within blocks $\mathcal{B}_1 = \mathbf{X}_{[1,b_n]}$, in space and in time, through the ℓ^α–norm to infer the marginal extremal features of $\mathbf{X}_1$. In particular, in Equation (1.5.16) we focus on inference of marginal coordinates. Of course, this approach can be implemented, albeit the knowledge of the two constants $m(j)$, $c(p)$. Though, choosing $p = \alpha$, the (tail) index, in Equation (1.5.16), yields $c(\alpha) = 1$ which simplifies to

$$\mathbb{P}(X_{0,j} > x_{b_n}) \ \sim \ m(j)\,(b_n)^{-1}\,\mathbb{P}(\textstyle\sum_{t=1}^{b_n}|\mathbf{X}_t|^\alpha > x_{b_n}^\alpha), \quad n \to +\infty. \tag{1.5.17}$$

Hence, we can infer the coordinate aspects of $X_{0,j}$ through (1.5.17), for $j = 1, \dots, d$. We pursue this strategy in Chapter 4 for high return levels inference.

Chapter 4 explains our new inference methodology in Algorithm 1, which responds to **(Quest. 4)**. Roughly speaking, suppose the constants $m(j)$, for $j = 1, \dots, d$, are known or have been already estimated. From the right-hand side of Equation (1.5.17), we can see that the extrapolation of high levels is possible from the distribution function

$$x \mapsto \mathbb{P}\big(\textstyle\sum_{t=1}^{b_n} |\mathbf{X}_t|^\alpha \le x \big). \tag{1.5.18}$$

Actually, since $|\mathbf{X}_0|^\alpha$ is regularly varying with (tail) index equal to one, the central limit theory justifies modeling the distribution in (1.5.18) with a stable distribution. In practice, we propose to fit the sub-sample of disjoint sums: $(\|\mathcal{B}_t\|_\alpha^\alpha)_{t=1,\dots,m_n}$ (see (1.5.5)), with a stable distribution of stable parameter one. From this fit we can infer the distribution in (1.5.18). Finally, for high return levels inference, we rely on the asymptotic approximation from Equation (1.5.17) and extrapolate the right-hand side of Equation (1.5.17) from the fitted stable distribution.

Two major advantages can be highlighted from this procedure. First, we fit a univariate stable distribution to the α-powers of sums only once, and information on the whole spatiotemporal extreme event is available in this first step. In simulation, this seems to enhance inference. Then, in the second step, we allocate the weights $m(j)$, for $j = 1, \cdots, d$, to recover coordinate features. Second, (1.5.17) keeps track of the temporal memories of the time series but avoids typical declustering methods [65, 34], which are needed for the Pareto-based methods. Indeed, these methods often rely on the asymptotics of ℓ^∞-blocks. To conclude, this new approach is justified since we will show $c(\alpha) = 1$ for a number of classical models exhibiting short-range time dependence behavior. Moreover, this method also requires estimating the (tail)-index $\alpha > 0$, but this is a mandatory step also in Pareto-based high quantile estimation. Again, it is important to highlight the important role of α for assessing temporal dependencies. Chapter 4 concludes with the analysis of fall daily heavy rainfall in France from the data set introduced in Section 1.2.

1.6 List of publications and code

1. G. Buriticá, N. Meyer, T. Mikosch, O. Wintenberger. (2021). Some variations on the extremal index. Zap. Nauchn. Semin. POMI. Volume 501, Probability and Statistics. 30, 52—77. To be translated in J.Math.Sci. (Springer). [arXiv]. (code)

2. G. Buriticá, T. Mikosch, O. Wintenberger. (2022) Large deviations of ℓ^p-blocks of regularly varying time series and applications to cluster inference. (Submitted). [arXiv].

3. G. Buriticá, P. Naveau. (2022). Stable sums to infer high return levels of multivariate rainfall time series. (Submitted). [arXiv]. (code)

Ongoing projects:

1. G. Buriticá, O. Wintenberger. Asymptotic normality for ℓ^p-cluster inference.

Chapter 2: Large deviations of ℓ^p–blocks of regularly varying time series and applications to cluster inference

Abstract

In the regularly varying time series setting, a cluster of exceedances is a short period for which the supremum norm exceeds a high threshold. We propose to study a generalization of this notion considering short periods, or blocks, with ℓ^p–norm above a high threshold. Our main result derives new large deviation principles of extremal ℓ^p-blocks. We show an application to cluster inference to motivate our result, where we design consistent disjoint blocks estimators. We focus on inferring important indices in extreme value theory, e.g., the *extremal index*. Our approach motivates cluster inference based on extremal ℓ^p–blocks with $p < \infty$ rather than the classical one with $p = \infty$. Our estimators also promote the use of large empirical quantiles from the ℓ^p–norm of blocks as threshold levels which eases implementation and facilitates comparison.

keywords[a]: *Extremal index, cluster Poisson process, extremal cluster, regularly varying time series, affine stochastic recurrence equation, autoregressive process*

[a] *Primary 60G70 Secondary 60F10 62G32 60F05 60G57*

Main contributions

I address (**Quest. 1-3**) in this Chapter. The following are my main contributions:

- (**Quest. 1**): Recall my goal to model extremal characteristics of blocks $\mathbf{X}_{[0,n]}$, as $n \to +\infty$. Theorem 2.2.1 presents large deviations of blocks embedded in the shift-invariant sequential space $(\tilde{\ell}^p, \tilde{d}_p)$, and defines p-clusters $\mathbf{Q}^{(p)} \in \ell^p$, satisfying $\|\mathbf{Q}^{(p)}\|_p = 1$, a.s.

- (**Quest. 3**): I present in Theorem 2.4.1 consistent disjoint blocks estimators tailored to infer the features of extremal ℓ^p-blocks. The estimators in (2.4.2) can be used to infer statistics

$$f_p^{\mathbf{Q}} \;=\; \mathbb{E}[f(Y\mathbf{Q}^{(p)})], \tag{2.0.1}$$

  for suitable functionals $f : \ell^p \to \mathbb{R}$, Y Pareto distributed, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, and independent of $\mathbf{Q}^{(p)}$.

  Moreover, Proposition 2.3.1 show that all p-cluster are related through a change of norms functional. Hence, the same statistic, let's say $f_p^{\mathbf{Q}}$ in (2.0.1), can be obtained by combining pairs $q, f_q : \ell^q \to \mathbb{R}$, for $p, q > 0$.

- (**Quest. 2**): The new estimators in (2.4.2) use order statistics of the ℓ^p-norms subsample as threshold levels. This strategy aims to stress the intrinsic relation between the large deviation results from Theorem 2.2.1, and the choice of p indicating we infer $f_p^{\mathbf{Q}}$ in (2.0.1) using extremal ℓ^p-blocks. I demonstrate inference with extremal ℓ^p-blocks, for $p < \infty$, is advantageous compared to the classical approach $p = \infty$. This is further illustrated with numerical experiments.

Contents

2.1 Introduction

For various applications of extreme value statistics with stationary time series, it is natural to wonder how a recorded high level can affect the future behavior of the sequence or how often extreme events occur. However, the extremal dependencies of the time series can disturb the inference procedures tailored for independent observations; cf. Leadbetter [109], Embrechts *et al.* [59]. We consider the setting of $\mathbb{R}^d$-valued stationary regularly varying time series $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ with generic element $\mathbf{X}_0$; see Section 2.2.1 for a definition, cf. Basrak and Segers [10]. In this framework, an exceedance of a high threshold by the norm $|\mathbf{X}_t|$ at time t might trigger consecutive exceedances in some small time interval around t. To model these short periods with at least one exceedance Davis and Hsing introduced the *clusters (of exceedances)* implicitly in the seminal paper [40]. These were further reviewed in Basrak and Segers [10] and Basrak *et al.* [9].

The main motivation for studying clusters (of exceedances) can be traced back to Theorem 2.5. in Davis and Hsing [40]. Roughly speaking, for weakly dependent regularly varying time series, the limit distribution of $a_n^{-1}\mathbf{X}_{[0,n]} = a_n^{-1}(\mathbf{X}_0, \ldots, \mathbf{X}_n)$ is fully characterized in terms of the index of regular variation, the cluster (of exceedances), and the *extremal index* of $(|\mathbf{X}_t|)$, denoted by $\theta_{|\mathbf{X}|}$, where $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$. In this representation, the notion of a cluster aims at modeling how rare events of $\mathbf{X}_{[0,n]}$ happen, i.e., its behavior if its supremum norm exceeds the high level x_n, and as the probability of recording a cluster tends to zero, i.e., $\mathbb{P}(\|\mathbf{X}_{[1,n]}\|_\infty > x_n) \sim \theta_{|\mathbf{X}|} n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$. From this last equivalence, we can also see that $\theta_{|\mathbf{X}|}$ summarizes the ratio between the expected number of exceedances in the stationary time-dependent scenario and the independent setting.

By the previous discussion, the cluster (of exceedances) is tied together with the extremal index by the supremum norm. Our main theoretical result extends the aforementioned ideas from the ℓ^∞–

norm to ℓ^p–norms. In Theorem 2.2.1 we investigate the behavior of $\mathbf{X}_{[0,n]}$ when its ℓ^p–norm exceeds high levels (x_n) satisfying $\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) = \mathbb{P}(\sum_{t=1}^n |\mathbf{X}_t|^p > x_n^p) \to 0$ as $n \to +\infty$. We call this a large deviation result since it describes the block behavior when the probability that the partial sums of the norm pth powers of $\mathbb{R}^d$-valued regularly varying vectors exceed the extreme threshold x_n^p. This leads us to a new definition of a *cluster process* in the space $\ell^p = \ell^p(\mathbb{R}^d)$ and, in the limiting case $p = \infty$, one recovers the classical clusters (of exceedances). Similarly, large deviation principles for sums were considered by Nagaev [124], Cline and Hsing [33] in the independent heavy-tailed case, and by Mikosch and Wintenberger [121, 122, 123], Mikosch and Rodionov [118] in the dependent heavy-tailed case. Instead, we derive large deviation principles for blocks in ℓ^p and focus on ℓ^p–norms: they increase when p decreases, hence the proportion of extremal clusters increases. This fact points at an advantageous road for improving inference procedures which we will follow in the second part of this article.

For inference purposes we divide a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ into disjoint blocks $(\mathbf{B}_t)_{1 \le t \le \lfloor n/b_n \rfloor}$, $\mathbf{B}_t := \mathbf{X}_{(t-1)\, b_n + [1, b_n]}$, for a sequence of block lengths (b_n) such that $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$. To apply our findings we study cluster inference. Typically, inference on clusters (of exceedances) starts with selecting blocks whose supremum norm exceeds a high threshold x_{b_n}; cf. Kulik and Soulier [108]. Instead, we aim at considering blocks whose ℓ^p–norms exceed a high threshold. As a matter of fact, for cluster inference the thresholds (x_{b_n}) should adapt to the block lengths (b_n) and take into account the value of p. In the existing literature for $p = \infty$ no detailed advice is given of how the couple (b_n) and (x_{b_n}) must be chosen; see for example Drees, Rootzén [52], Drees, Neblung [51], Cissokho, Kulik [28], Drees *et al.* [50], who assume bias conditions on the sequence (x_n) such as $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ as $n \to \infty$. It is common practice to replace x_{b_n} by an upper order statistic of $(|\mathbf{X}_t|)_{1 \le t \le n}$, turning the choice of b_n into a delicate problem which has not been studied carefully. This problem occurs for example in the context of the blocks estimator of the *extremal index* proposed by Hsing [88].

To address the aforementioned threshold/block length difficulties of cluster inference, we design consistent disjoint blocks methods in Theorem 2.4.1 with thresholds chosen as order statistics of ℓ^p–norms that adapt to the length of the block. In this way one deals with the classical bias-variance trade-off in extreme value statistics (see Resnick [143]) where choosing a small number of order statistics increases the variance while a large number might lead to biased estimates. We retrieve the need for a rigorous definition of extremal ℓ^p–blocks for cluster inference to reveal how p can play a key role. Indeed, we can infer the same quantity using extremal ℓ^p–blocks, for different values of p, by applying a change of norms functional. When comparing these estimators, our large deviation principles support our approach for $p < \infty$ is more robust compared to the classical perspective based on $p = \infty$.A similar change of norms approach was investigated in Drees *et al.* [53] and Davis *et al* [39] for inference of the tail process. In this paper, we focus on inference of cluster processes and show consistency of the blocks estimators. Asymptotic normality of our estimators can be derived by combining arguments from Theorem 4.3 in Cissokho and Kulik [28] and the large deviation arguments developed below; this topic is the subject of ongoing work.

We apply our inference procedure to estimate the extremal index using extremal ℓ^α–blocks, with α equal to the index of regular variation of $(\mathbf{X}_t)$. We also consider inference of *cluster indices* as defined by Mikosch and Wintenberger [122] on partial sum functionals by considering extremal ℓ^1–blocks. These functionals are shift-invariant with respect to the backward shift in sequence spaces; see Kulik and Soulier [108] for details. To define extremal ℓ^p–blocks for $p \leq \alpha$, we require a so-called vanishing-small-values condition too; see Section 2.5. Another advantage of this consideration is that, using a simple continuity argument, we can also study αth-power sum functionals acting on ℓ^p. Then, coupled with the random shift analysis of Janssen [96] based on the αth moment of the cluster process, we also extend cluster inference to functionals acting on ℓ^p rather than on shift-invariant spaces. We heavily rely on Theorem 2.2.1 to provide the main argument for comparing inference methods as we build therein on the definition of cluster processes from the new large deviation principles.

2.1.1 Outline of the paper.

Section 2.2 states the main large deviation principle in Theorem 2.2.1 after introducing the preliminaries on regular variation and fixing the notation. In Section 2.3 we study the cluster processes in the space ℓ^p introduced in Theorem 2.2.1. In Section 2.4 we apply Theorem 2.2.1 to inference for shift-invariant functionals acting on these cluster processes through Theorem 2.4.1, choosing thresholds as empirical quantiles of the ℓ^p–norms of blocks. In an example, we illustrate our approach for $p < \infty$ and compare it with the setting $p = \infty$. This is done at the end of Section 2.4. In Section 2.5 we supplement the discussion on Theorem 2.2.1 and analyze in depth its assumptions. In Section 2.6 we provide inference for non-shift-invariant functionals. We defer all proofs to Section 2.7.

2.2 Preliminaries and main result

2.2.1 About regular variation of time series

We consider an $\mathbb{R}^d$-valued stationary process $(\mathbf{X}_t)$. Following Davis and Hsing [40], we call it regularly varying if the finite-dimensional distributions of the process are regularly varying. This notion involves the vague convergence of certain tail measures; see Resnick [143]. Avoiding the concept of vague convergence and infinite limit measures, Basrak and Segers [10] showed that regular variation of $(\mathbf{X}_t)$ is equivalent to the weak convergence relations: for any $h \geq 0$,

$$\mathbb{P}\big(x^{-1}(\mathbf{X}_t)_{|t|\leq h} \in \cdot \mid |\mathbf{X}_0| > x\big) \xrightarrow{w} \mathbb{P}\big(Y\,(\Theta_t)_{|t|\leq h} \in \cdot\big)\,, \qquad x \to \infty\,, \tag{2.2.1}$$

where Y is Pareto(α)-distributed, i.e., it has tail $\mathbb{P}(Y > y) = y^{-\alpha}$, $y > 1$, independent of the vector $(\Theta_t)_{|t|\leq h}$ and $|\Theta_0| = 1$. According to Kolmogorov's consistency theorem, one can extend the latter finite-dimensional vectors to a sequence $\Theta = (\Theta_t)_{t\in\mathbb{Z}}$ in $(\mathbb{R}^d)^{\mathbb{Z}}$ called the *spectral tail process* of $(\mathbf{X}_t)$.

The regular variation property, say $\mathbf{RV}_\alpha$, of $(\mathbf{X}_t)$ is determined by the (tail)-index $\alpha > 0$ and the spectral tail process.

Extending vague convergence to M_0-convergence, Hult and Lindskog [89] introduced regular variation for random elements assuming values in a general complete separable metric space; see also Lindskog *et al.* [113]. Segers *et al.* [152] proved regular variation of random elements with values in star-shaped metric spaces. Their results are based on weak convergence in the spirit of (2.2.1). Our focus will be on a special star-shaped space: the sequence space ℓ^p, $p \in (0,\infty]$ equipped with the metric

$$d_p(\mathbf{x},\mathbf{y}) := \begin{cases} \|\mathbf{x}-\mathbf{y}\|_p = \Big(\sum_{t\in\mathbb{Z}} |\mathbf{x}_t - \mathbf{y}_t|^p\Big)^{1/p}, & p \geq 1\,, \\ \|\mathbf{x}-\mathbf{y}\|_p^p\,, & p \in (0,1)\,, \end{cases} \qquad \mathbf{x},\mathbf{y} \in \ell^p\,,$$

with the usual convention in the case $p = \infty$. We know that d_p makes ℓ^p a separable Banach space for $p \geq 1$, and a separable complete metric space for $p \in (0,1)$. Using the *p-modulus function* $\|\cdot\|_p$, the ℓ^p-valued stationary process $(\mathbf{X}_t)$ has the property $\mathbf{RV}_\alpha$ if and only if relation (2.2.1) holds with $|\mathbf{X}_0|$ replaced by $\|\mathbf{X}_{[0,h]}\|_p$. Equivalently, (Proposition 3.1 in Segers *et al.* [152]), for any $h \geq 0$,

$$\mathbb{P}\big(x^{-1}\mathbf{X}_{[0,h]} \in \cdot \mid \|\mathbf{X}_{[0,h]}\|_p > x\big) \xrightarrow{w} \mathbb{P}\big(Y\,\mathbf{Q}^{(p)}(h) \in \cdot\big)\,, \qquad x \to \infty\,, \tag{2.2.2}$$

where for $a \leq b$, $\mathbf{X}_{[a,b]} = (\mathbf{X}_a,\ldots,\mathbf{X}_b)$, and the Pareto($\alpha$) variable Y is independent of $\mathbf{Q}^{(p)}(h) \in \mathbb{R}^{d\,(h+1)}$, and $\|\mathbf{Q}^{(p)}(h)\|_p = 1$ a.s. We call $\mathbf{Q}^{(p)}(h)$ the *spectral component* of $\mathbf{X}_{[0,h]}$.

2.2.2 Notation

For integers i and $a < b$ we write $i + [a,b] = \{i+a,\ldots,i+b\}$. It is convenient to embed the vectors $\mathbf{x}_{[a,b]} \in \mathbb{R}^{d(b-a+1)}$ in $(\mathbb{R}^d)^{\mathbb{Z}}$ by assigning zeros to indices $i \notin [a,b]$, and we then also write $\mathbf{x}_{[a,b]} \in (\mathbb{R}^d)^{\mathbb{Z}}$. We write $\mathbf{x} := (\mathbf{x}_t) = (\mathbf{x}_t)_{t\in\mathbb{Z}}$, and define truncation at level $\varepsilon > 0$ from above and below by $\underline{\mathbf{x}}_\epsilon = (\underline{\mathbf{x}_t}_\epsilon)_{t\in\mathbb{Z}}$, $\overline{\mathbf{x}}^\varepsilon = (\overline{\mathbf{x}_t}^\epsilon)_{t\in\mathbb{Z}}$, where $\underline{\mathbf{x}_t}_\epsilon = \mathbf{x}_t\,\mathbb{1}(|\mathbf{x}_t| > \epsilon)$, $\overline{\mathbf{x}}_t^\epsilon = \mathbf{x}_t\,\mathbb{1}(|\mathbf{x}_t| \leq \epsilon)$.

Recall the *backshift operator* acting on $\mathbf{x} \in (\mathbb{R}^d)^{\mathbb{Z}}$: $B^k\mathbf{x} = (\mathbf{x}_{t-k})_{t\in\mathbb{Z}}$, $k \in \mathbb{Z}$. Let $\widetilde{\ell}^p = \ell^p/\sim$ be the quotient space with respect to the equivalence relation $\sim$ in ℓ^p: $\mathbf{x} \sim \mathbf{y}$ holds if there exists $k \in \mathbb{Z}$ such that $B^k\mathbf{x} = \mathbf{y}$. An element of $\widetilde{\ell}^p$ is denoted by $[\mathbf{x}] = \{B^k\mathbf{x} : k \in \mathbb{Z}\}$. For ease of notation, we often write $\mathbf{x}$ instead of $[\mathbf{x}]$, and we notice that any element in ℓ^p can be embedded in $\widetilde{\ell}^p$ by using the equivalence relation. We define for $[\mathbf{x}],[\mathbf{y}] \in \widetilde{\ell}^p$,

$$\widetilde{d}_p([\mathbf{x}],[\mathbf{y}]) := \inf_{k\in\mathbb{Z}} \big\{d_p(B^k\mathbf{a},\mathbf{b}) : \mathbf{a} \in [\mathbf{x}]\,, \mathbf{b} \in [\mathbf{y}]\big\}\,.$$

For $p > 0$, $\widetilde{d}_p$ is a metric on $\widetilde{\ell}^p$ and turns it into a complete metric space; see Basrak *et al.* [9].

2.2.3 Main result

We start by giving our main result on large deviations of the sequence $\mathbf{X}_{[0,n]}$, that we embed in the space $(\tilde{\ell}^p, \tilde{d}_p)$. The proof is postponed to Section 2.7.1.

Theorem 2.2.1. *Consider an $\mathbb{R}^d$-valued stationary time series $(\mathbf{X}_t)$ satisfying $\mathbf{RV}_\alpha$ for some $\alpha > 0$. For a given $p > 0$, assume that there exist a sequence (x_n) such that $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ and some $\kappa > 0$ such that $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$ as $n \to +\infty$. Furthermore, assume that for all $\delta > 0$,*

AC : $\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\|\mathbf{X}_{[k,n]}\|_\infty > \delta\, x_n \mid |\mathbf{X}_0| > \delta\, x_n\big) = 0,$

$\mathbf{CS}_p$: $\lim_{\epsilon\to 0} \limsup_{n\to\infty} \dfrac{\mathbb{P}(\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p > \delta)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = 0.$

Then, there exists $c(p) > 0$ such that

$$\lim_{n\to+\infty} \frac{\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} \quad = \quad c(p)\,. \tag{2.2.3}$$

Moreover, $c(p) < \infty$ if $p \geq \alpha$, in particular, $c(\infty) \leq c(p) \leq c(\alpha) = 1$. If $c(p) < \infty$ there exists $\mathbf{Q}^{(p)} \in (\ell^p, d^p)$ such that $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. and

$$\mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_p > x_n) \xrightarrow{w} \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot), \quad n \to \infty\,, \tag{2.2.4}$$

in the space $(\tilde{\ell}^p, \tilde{d}_p)$ where Y is Pareto(α) distributed, independent of $\mathbf{Q}^{(p)}$.

Since $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., $\mathbf{Q}^{(p)}$ has interpretation as the spectral component of a regularly varying ℓ^p-valued random element. From (2.2.3) we infer that $c(p)$ is a non-decreasing function of p. Letting $p = \infty$, one recovers the classical definition of clusters (of exceedances) in (2.2.4). From Theorem 2.2.1 we also have $c(\alpha) = 1 < \infty$ which motivates the study of extremal ℓ^α–blocks and, for the purposes of statistical inference, suggests $\mathbf{Q}^{(\alpha)}$ as a potential competitor of $\mathbf{Q}^{(\infty)}$.

We refer to a relation of the type (2.2.3) as *large deviation probabilities* motivated by the following observation. Write $S_k^{(p)} = \sum_{t=1}^k |\mathbf{X}_t|^p$ for $k \geq 1$. Then $|\mathbf{X}|^p$ is regularly varying with index α/p. Relation (2.2.3) implies that

$$\begin{aligned}\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \quad &= \quad \mathbb{P}\big(S_n^{(p)} > x_n^p\big)\\ &\sim \quad c(p)\, n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0\,, \quad n \to \infty\,.\end{aligned}$$

Thus the left-hand probability describes the rare event that the sum process $S_n^{(p)}$ exceeds the extreme threshold x_n^p.

Equation (2.2.4) extends the large deviation result for $\|\mathbf{X}_{[0,n]}\|_p$ in (2.2.3) to one for the process $\mathbf{X}_{[0,n]}$ in the sequence space ℓ^p. Motivated by inference for the spectral cluster process $\mathbf{Q}^{(p)}$, we establish (2.2.4) employing weak convergence in the spirit of the polar decomposition given in Segers *et al.* [152].

We work under the one-sided anti-clustering condition **AC** together with a telescoping sum argument to compensate for the classical two-sided condition (2.5.1) used in Kulik and Soulier [108]. Conditions similar to $\mathbf{CS}_p$ are standard when dealing with sum functionals acting on $(\mathbf{X}_t)$ and hold for any time series satisfying $\mathbf{RV}_\alpha$ for $p > \alpha$ by a Karamata–type argument. In Section 2.5 we further discuss the assumptions of Theorem 2.2.1.

Remark 2.2.2. *Equation (2.2.4) provides a family of Borel sets in $(\tilde{\ell}^p, \tilde{d}_p)$ for which the weak limit of the self-normalized blocks $\mathbf{X}_{[0,n]}/\|\mathbf{X}_{[0,n]}\|_p$ exists. This result implies that the sequence of measures*

$$\begin{aligned} &\mu_n(\cdot) := \mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot)/\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \\ \to \qquad &\mu(\cdot) := \int_0^\infty \mathbb{P}(y\,\mathbf{Q}^{(p)} \in \cdot)\, d(-y^{-\alpha})\,, \quad n \to \infty\,, \end{aligned}$$

in the M_0-sense in $(\tilde{\ell}^p, \tilde{d}_p)$. By the portmanteau theorem for measures (Theorem 2.4. in Hult and Lindskog [89])

$$\mu_n(A) = \mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in A)/\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \to \mu(A)\,,$$

for all Borel sets A in $(\tilde{\ell}^p, \tilde{d}_p)$ satisfying $\mu(\partial A) = 0$ and $\mathbf{0} \notin \overline{A}$. This approach is discussed in Kulik and Soulier [108] where similar conditions are stated for obtaining limit results for $\tilde{\ell}^1$-functionals.

2.3 Spectral cluster process representation

2.3.1 The spectral cluster process in ℓ^p

From (2.2.1) recall the spectral tail process Θ of a stationary sequence $(\mathbf{X}_t)$ satisfying $\mathbf{RV}_\alpha$. We start with a representation of the spectral cluster process $\mathbf{Q}^{(p)}$ from (2.2.4) in terms of Θ. The proof is deferred to Section 2.7.3.

Proposition 2.3.1. *For $p, \alpha > 0$ assume $\|\Theta\|_p + \|\Theta\|_\alpha < \infty$ a.s. Under the assumptions of Theorem 2.2.1, the distribution of the spectral cluster process $\mathbf{Q}^{(p)}$ in the space (ℓ^p, d_p) is given by*

$$\mathbb{P}(\mathbf{Q}^{(p)} \in \ \cdot) \quad = \quad c(p)^{-1}\mathbb{E}\big[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha\ \mathbb{1}(\Theta/\|\Theta\|_p \in \ \cdot)\big]\,, \tag{2.3.1}$$

where the constant $c(p)$ defined in (2.2.3) admits the representation

$$c(p) \quad = \quad \mathbb{E}\big[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha\big]\,. \tag{2.3.2}$$

This result provides a new representation of the distribution of $\mathbf{Q}^{(p)}$ for fixed p and relates distinct spectral cluster processes to each other by the change of norms transform in (2.3.1). In the next sections we consider the cases $p \in \{\alpha, \infty\}$ in detail.

The spectral cluster process in ℓ^α

In view of (2.3.1) the process $\Theta/\|\Theta\|_\alpha$ is the candidate for the ℓ^α-spectral cluster process $\mathbf{Q}^{(\alpha)}$ introduced in (2.2.4), and it plays a key role for characterising $\mathbf{Q}^{(p)}$ in general. The following result shows that $\Theta/\|\Theta\|_\alpha$ is well defined under **AC**.

Proposition 2.3.2. *Let $(\mathbf{X}_t)$ be a stationary sequence satisfying $\mathbf{RV}_\alpha$ with spectral tail process (Θ_t). Then the following statements are equivalent:*

i) $\|\Theta\|_\alpha < \infty$ a.s. *and $\Theta/\|\Theta\|_\alpha$ is well defined in ℓ^α.*

ii) $|\Theta_t| \to 0$ a.s. *as $t \to \infty$.*

iii) *The time of the largest record $T^* := \inf\{s : s \in \mathbb{Z}$ such that $|\Theta_s| = \sup_{t\in\mathbb{Z}} |\Theta_t|\}$ is finite a.s.*

Moreover, these statements hold under **AC**.

A proof of Proposition 2.3.2 is given in Lemma 3.6 of Buriticá *et al.* [25], appealing to results by Janssen [96].

From (2.2.2) recall the sequence of spectral components $(\mathbf{Q}^{(\alpha)}(h))_{h\geq 0}$ of the vectors $(\mathbf{X}_{[0,h]})_{h\geq 0}$ with the property $\|\mathbf{Q}^{(\alpha)}(h)\|_\alpha = 1$ a.s. Our next result relates the sequence $(\mathbf{Q}^{(\alpha)}(h))_{h\geq 0}$ to $\Theta/\|\Theta\|_\alpha$, the candidate process for $\mathbf{Q}^{(\alpha)}$.

Proposition 2.3.3. *Let $(\mathbf{X}_t)$ be a stationary time series satisfying $\mathbf{RV}_\alpha$ and $\lim_{t\to\infty} |\Theta_t| = 0$ a.s. Then $\mathbf{Q}^{(\alpha)}(h) \xrightarrow{d} \mathbf{Q}^{(\alpha)}(\infty)$ as $h \to \infty$ in $(\widetilde{\ell}^\alpha, \widetilde{d}_\alpha)$ with $\mathbf{Q}^{(\alpha)}(\infty) \stackrel{d}{=} \Theta/\|\Theta\|_\alpha$.*

This result gives raise to the interpretation of $\Theta/\|\Theta\|_\alpha$ as *the spectral component of $(\mathbf{X}_t)$ in ℓ^α*. The proof is given in Section 2.7.3. Under the assumptions of Theorem 2.2.1 we also have the a.s. representations $\mathbf{Q}^{(\alpha)} = \Theta/\|\Theta\|_\alpha$ and $\Theta = \mathbf{Q}^{(\alpha)}/|\mathbf{Q}^{(\alpha)}_0|$ in (ℓ^α, d^α).

The spectral cluster process in ℓ^∞

Letting $p = \infty$ in Theorem 2.2.1, we retrieve the classical definition of the cluster (of exceedances) by embedding $\mathbf{Q}^{(\infty)}$ from (2.2.4) in $(\widetilde{\ell}^\infty, \widetilde{d}_\infty)$. From Proposition 2.3.2, assuming **AC**, we can relate the cluster (of exceedances) to $\Theta/\|\Theta\|_\alpha$ using (2.3.1) by

$$\mathbb{P}(Y\mathbf{Q}^{(\infty)} \in \cdot\,) \quad = \quad \mathbb{P}(Y\Theta/\|\Theta\|_\infty \in \cdot \mid \|\,Y\Theta/\|\Theta\|_\alpha\|_\infty > 1)\,, \tag{2.3.3}$$

where Y is a Pareto(α) random variable independent of $\mathbf{Q}^{(\infty)}$ and Θ. In particular, $\|\mathbf{Q}^{(\infty)}\|_\infty = 1$ a.s. and $\mathbb{P}(\|\,Y\Theta/\|\Theta\|_\alpha\|_\infty > 1) = \mathbb{E}[\|\,\Theta/\|\Theta\|_\alpha\|_\infty^\alpha] = c(\infty)$.

Assuming **AC** and additional mixing assumptions (see e.g. Theorem 2.3. in [25]), the extremal index $\theta_{|\mathbf{X}|}$ of $(|\mathbf{X}_t|)$ exists and equals $c(\infty)$. Under **AC**, Basrak and Segers [10] proved that $|\Theta_t| \to 0$ as $|t| \to \infty$ and $c(\infty)$ is given by

$$c(\infty) = \mathbb{E}\big[\|(\Theta_t)_{t\geq 0}\|_\infty^\alpha - \|(\Theta_t)_{t\geq 1}\|_\infty^\alpha\big]\,. \tag{2.3.4}$$

Following Planinić and Soulier [138], the spectral tail process (Θ_t) satisfies the *time-change formula:* for any measurable bounded function $f : (\widetilde{\ell}^p, \widetilde{d}_p) \to \mathbb{R}$ such that $f(\lambda\mathbf{x}) = f(\mathbf{x})$ for all $\lambda > 0$, we have for all $t, s \in \mathbb{Z}$,

$$\mathbb{E}[f(B^s(\Theta_t))\mathbb{1}(\Theta_{-s} \neq \mathbf{0})] \quad = \quad \mathbb{E}\big[|\Theta_s|^\alpha f\big((\Theta_t)\big)\big]\,. \tag{2.3.5}$$

An application of this formula and a telescoping sum argument yield

$$\begin{aligned}
c(p) &= \mathbb{E}[\|\Theta\|_p^\alpha / \|\Theta\|_\alpha^\alpha] \\
&= \sum_{s\in\mathbb{Z}} \mathbb{E}\Big[\big(\|(\Theta_t)_{t\geq -s}/\|\Theta\|_\alpha\|_p^\alpha - \|(\Theta_t)_{t\geq -s+1}/\|\Theta\|_\alpha\|_p^\alpha\big)\Big] \\
&= \sum_{s\in\mathbb{Z}} \mathbb{E}\Big[|\Theta_s|^\alpha\big(\|(\Theta_t)_{t\geq 0}/\|\Theta\|_\alpha\|_p^\alpha - \|(\Theta_t)_{t\geq 1}/\|\Theta\|_\alpha\|_p^\alpha\big)\Big] \\
&= \mathbb{E}\Big[\|\Theta\|_\alpha^\alpha\big(\|(\Theta_t)_{t\geq 0}/\|\Theta\|_\alpha\|_p^\alpha - \|(\Theta_t)_{t\geq 1}/\|\Theta\|_\alpha\|_p^\alpha\big)\Big] \\
&= \mathbb{E}\Big[\|(\Theta_t)_{t\geq 0}\|_p^\alpha - \|(\Theta_t)_{t\geq 1}\|_p^\alpha\Big]\,.
\end{aligned} \tag{2.3.6}$$

Thus we deduce the representation in (2.3.4) for $c(\infty)$ by letting $p = \infty$ in (2.3.6).

Furthermore the representation in (2.3.6) extends from $p = \infty$ to $0 < p \leq \infty$ under **AC**. We notice that under **AC** it is in general not obvious whether $c(p)$ is finite or not for $0 < p < \alpha$. A Taylor expansion shows that (2.3.6) is finite if $\mathbb{E}[\|(\Theta_t)_{t\geq 0}\|_p^{\alpha-p}] < \infty$ and thus $c(p) < \infty$, which is a necessary condition for defining $\mathbf{Q}^{(p)}$ as in (2.2.4).

2.4 Consistent cluster inference based on spectral cluster processes

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a sample from a stationary sequence $(\mathbf{X}_t)$ satisfying $\mathbf{RV}_\alpha$ for some $\alpha > 0$ and choose $p > 0$. We split the sample into disjoint blocks $\mathbf{B}_t := \mathbf{X}_{(t-1)b+[1,b]}$, $t = 1, \dots, m_n$, where $b = b_n \to \infty$ and $m = m_n := [n/b_n] \to \infty$. Throughout we assume the conditions of Theorem 2.2.1 for p. Then, in particular, $\mathbb{P}(\|\mathbf{B}_1\|_p > x_b) \to 0$ and $k = k_n := [m_n\mathbb{P}(\|\mathbf{B}\|_p > x_{b_n})] \to \infty$.

The key message from Proposition 2.3.1 is that the same quantity, say e.g., $c(q)$, for $q > 0$, can be obtained from letting different functionals act on $\mathbf{Q}^{(p)}$ for different p, in particular for $p < \infty$. In this context, we want to compare the inference procedures of $p < \infty$ and $p = \infty$, and promote the use of the order statistics of $\|\mathbf{B}_t\|_p$, $t = 1, \dots, m_n$, i.e., $\|\mathbf{B}\|_{p,(1)} \geq \|\mathbf{B}\|_{p,(2)} \geq \cdots \geq \|\mathbf{B}\|_{p,(m)}$.

2.4.1 Cluster functionals and mixing

The real-valued function g on $\tilde{\ell}^p$ is a *cluster functional* if it vanishes in some neighborhood of the origin and $\mathbb{P}(Y\mathbf{Q}^{(p)} \in D(g)) = 0$ where $D(g)$ denotes the set of discontinuity points of g. In what follows, it will be convenient to write $\mathcal{G}_+(\tilde{\ell}^p)$ for the class of non-negative functions on $\tilde{\ell}^p$ which vanish in some neighborhood of the origin.

For asymptotic theory we will need the following *mixing condition.*

Condition $\mathbf{MX}_p$. There exists an integer sequence $b = b_n \to \infty$ such that $m = m_n \to \infty$, and for any Lipschitz-continuous $f \in \mathcal{G}_+(\tilde{\ell^p})$, the sequence (x_n) satisfies

$$\mathbb{E}\big[\mathrm{e}^{-\frac{1}{k}\sum_{t=1}^{m} f(x_b^{-1}\mathbf{B}_t)}\big] = \big(\mathbb{E}\big[\mathrm{e}^{-\frac{1}{k}\sum_{t=1}^{\lfloor m/k \rfloor} f(x_b^{-1}\mathbf{B}_t)}\big]\big)^k + o(1)\,, \quad n \to \infty\,. \tag{2.4.1}$$

If $\mathbf{MX}_p$ is required in the sequel we will refer to the sequences (b_n), (m_n) and (k_n) chosen in this condition. $\mathbf{MX}_p$ is similar to the mixing conditions $\mathcal{A}$, $\mathcal{A}'$ in Davis and Hsing [40], Basrak *et al.* [8], respectively. These are defined in terms of sequences $(f(\mathbf{X}_t))$ while our functionals f act on blocks. $\mathbf{MX}_p$ holds under mild conditions, for example, under strong mixing with quite general mixing rate; cf. Lemma 6.2. in Basrak *et al.* [9].

2.4.2 Consistent cluster inference

The following result is the basis for an empirical procedure for spectral cluster inference built on disjoint blocks. The proof is given in Section 2.7.4.

Theorem 2.4.1. *Assume the conditions of Theorem 2.2.1 hold for $p > 0$ with $c(p) < \infty$ together with $\mathbf{MX}_p$. Then $\|\mathbf{B}\|_{p,(k)}/x_b \xrightarrow{\mathbb{P}} 1$ and for all $g \in \mathcal{G}_+(\tilde{\ell^p})$,*

$$\frac{1}{k}\sum_{t=1}^{m} g\big(\|\mathbf{B}\|_{p,(k)}^{-1}\mathbf{B}_t\big) \xrightarrow{\mathbb{P}} \int_0^\infty \mathbb{E}\big[g(y\mathbf{Q}^{(p)})\big]d(-y^{-\alpha}), \quad n \to \infty\,. \tag{2.4.2}$$

The sequence (k_n) in (2.4.2) depends on p, but the choice of p is flexible due to Proposition 2.3.1 relating spectral cluster processes by a change of norms functional. The constant at the right-hand side of (2.4.2) can be written in different ways by combining the choice of p with the choice of the functional g. The choice of p, g are left to the practitioner. Theorem 2.4.1 points to the following two advantages.

Notice Theorem 2.4.1 promotes the use of order statistics of the sample of ℓ^p–norms where thresholds adapt automatically to block lengths. This allows us to exploit the link between the block length and the threshold through the sequence (k_n) that must satisfy

$$k_n = [m_n\mathbb{P}(\|\mathbf{B}\|_p > x_{b_n})] \;\sim\; c(p)\, n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \;=\; o(n/b_n^{(\alpha/p)\vee 1}),$$
$$n \to +\infty. \tag{2.4.3}$$

The asymptotic equivalence follows from the restriction on the thresholds in Theorem 2.2.1 where for $p < \alpha$, $n/x_n^p \to 0$, and for $p = \alpha$, $n/x_n^{\alpha-\kappa} \to 0$ for some $\kappa > 0$.

Also from (2.4.2) we see that k_n corresponds to the total number of extreme blocks used for inference. Then, for inference through $\mathbf{Q}^{(p)}$, equation (2.4.3) justifies taking k_n larger if $p < \infty$ than if $p = \infty$ since $c(\cdot)$ is a non-increasing function of p.

In the next section we illustrate these two points.

2.4.3 Applications

In this section we apply Theorem 2.4.1 to inference on some indices related to the extremes in a time-dependent sample and focus on cluster inference using $\mathbf{Q}^{(\alpha)}$ and $\mathbf{Q}^{(1)}$.

The extremal index

The extremal index of a regularly varying stationary time series has interpretation as a measure of clustering of serial exceedances, and was originally introduced in Leadbetter [109] and Leadbetter *et al.* [110]. If $(\mathbf{X}'_t)$ is iid with the same marginal distribution as $(\mathbf{X}_t)$ then the extremal index $\theta_{|\mathbf{X}|}$ relates the expected number of serial exceedances of $(|\mathbf{X}_t|)$ with the serial exceedances of $(|\mathbf{X}'_t|)$.

We aim to apply Theorem 2.4.1 with $p = \alpha$. In this setting, the change of norms formula in (2.3.2) and the discussion in Section 2.3.1 lead to the identities

$$\theta_{|\mathbf{X}|} = c(\infty) = \mathbb{E}\Big[\frac{\|\Theta\|_\infty^\alpha}{\|\Theta\|_\alpha^\alpha}\Big] = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha].$$

Then letting $p = \alpha$ and $g(\mathbf{x}) = \big(\|\mathbf{x}\|_\infty^\alpha/\|\mathbf{x}\|_\alpha^\alpha\big)\mathbb{1}(\|\mathbf{x}\|_\alpha > 1)$ on the right-hand side of (2.4.2), we obtain

$$\begin{aligned}\int_0^\infty \mathbb{E}\big[g(y\mathbf{Q}^{(\alpha)})\big]d(-y^{-\alpha}) &= \int_0^\infty \mathbb{E}\Big[\frac{\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha}{\|\mathbf{Q}^{(\alpha)}\|_\alpha^\alpha}\mathbb{1}(\|\mathbf{Q}^{(\alpha)}\|_\alpha^\alpha > y^{-\alpha})\Big]d(-y^{-\alpha}) \\ &= \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha] = c(\infty)\,.\end{aligned}$$

Next we introduce a new consistent disjoint blocks estimator of the extremal index defined from exceedances of ℓ^α-norm blocks.

Corollary 2.4.2. *Assume the conditions of Theorem 2.4.1 for $p = \alpha$. Then*

$$\frac{1}{k}\sum_{t=1}^{m}\frac{\|\mathbf{B}_t\|_\infty^\alpha}{\|\mathbf{B}_t\|_\alpha^\alpha}\mathbb{1}(\|\mathbf{B}_t\|_\alpha > \|\mathbf{B}\|_{\alpha,(k)}) \xrightarrow{\mathbb{P}} c(\infty)\,,\ n \to \infty\,. \tag{2.4.4}$$

To motivate this estimator of $c(\infty)$ we compare it to one based on the cluster (of exceedances) following a more classical approach. Motivated by the blocks estimator in Hsing [88] for the extremal index, we let $g(\mathbf{x}) := \sum_{j\in\mathbb{Z}}\mathbb{1}(|\mathbf{x}_t| > 1)$ act on large ℓ^∞–blocks. Choosing $p = \infty$ and fixing g on the right-hand side of (2.4.2), we can find an integer sequence $k' = k'_n \to \infty$ such that

$$\Big(\frac{1}{k'}\sum_{t=1}^{n}\mathbb{1}(|\mathbf{X}_t| \geq \|\mathbf{B}\|_{\infty,(k')})\Big)^{-1} \xrightarrow{\mathbb{P}} c(\infty)\,,\quad n \to \infty\,. \tag{2.4.5}$$

Arguing as for (2.4.3), we obtain $k'_n \sim m_n\mathbb{P}(\|\mathbf{B}\|_\infty > x_b) \sim c(\infty)\,n\mathbb{P}(|\mathbf{X}_0| > x_b)$. Thus, the proportion of extreme blocks used in (2.4.5) depends on $c(\infty)$ whereas the number of extreme blocks used in (2.4.4) is constant regardless of the time dependence structure.

A cluster index for sums

In this section we assume $\alpha \in (0,2)$ and $\mathbb{E}[\mathbf{X}] = \mathbf{0}$ for $\alpha \in (1,2)$. We study the partial sums $\mathbf{S}_n := \sum_{t=1}^n \mathbf{X}_t$, $n \geq 1$, and introduce a normalizing sequence (a_n) such that $n\,\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$. Starting with Davis and Hsing [40], α-stable central limit theory for $(\mathbf{S}_n/a_n)$ was proved under suitable anti-clustering and mixing conditions.

In this setting, the quantity $c(1)$ appears naturally and was coined cluster index in Mikosch and Wintenberger [122]. For $d = 1$ it can be interpreted as an equivalent of the extremal index for partial sums rather than maxima. Indeed, consider a real-valued regularly varying stationary sequence (X_t) with index of regular variation $\alpha \in (0,2)$ satisfying $\mathbb{P}(X \leq -x) = o\big(\mathbb{P}(X > x)\big)$ or $X \stackrel{d}{=} -X$. Consider an iid sequence (X'_t) with $X \stackrel{d}{=} X'$ and partial sums (S'_n). Then $a_n^{-1} S_n \xrightarrow{d} \xi_\alpha$ and $a_n^{-1} S'_n \xrightarrow{d} \xi'_\alpha$, both ξ_α and ξ'_α are α-stable and

$$\mathbb{E}[\mathrm{e}^{\,iu\xi_\alpha}] \quad = \quad \big(\mathbb{E}[\mathrm{e}^{\,iu\xi'_\alpha}]\big)^{c(1)} .$$

Letting $p = 1$ and $g(\mathbf{x}) = \big(\|\mathbf{x}\|_\alpha^\alpha / \|\mathbf{x}\|_1^\alpha\big) \mathbb{1}(\|\mathbf{x}\|_1 > 1)$ on the right-hand side of (2.4.2), we obtain

$$\begin{aligned} \int_0^\infty \mathbb{E}\big[g(y\mathbf{Q}^{(1)})\big] d(-y^{-\alpha}) &= \int_0^\infty \mathbb{E}\Big[\frac{\|\mathbf{Q}^{(1)}\|_\alpha^\alpha}{\|\mathbf{Q}^{(1)}\|_1^\alpha} \mathbb{1}(\|\mathbf{Q}^{(1)}\|_1^\alpha > y^{-\alpha})\Big] d(-y^{-\alpha}) \\ &= \mathbb{E}[\|\mathbf{Q}^{(1)}\|_\alpha^\alpha] = (c(1))^{-1} , \end{aligned}$$

where the last identity follows from Proposition 2.3.1. Then Theorem 2.4.1 for $p = 1$ and g as mentioned yields a consistent estimator of $c(1)$.

Corollary 2.4.3. *Assume the conditions of Theorem 2.4.1 for $p = 1$. Then*

$$\Big(\frac{1}{k}\sum_{t=1}^m \frac{\|\mathbf{B}_t\|_\alpha^\alpha}{\|\mathbf{B}_t\|_1^\alpha} \mathbb{1}(\|\mathbf{B}_t\|_1 > \|\mathbf{B}\|_{1,(k)})\Big)^{-1} \xrightarrow{\mathbb{P}} c(1) , \quad n \to \infty . \tag{2.4.6}$$

As we have argued above, such an estimator is appealing since it is based on large ℓ^1-norms of blocks instead of ℓ^∞-norms. Moreover, arguing as in Cissokho and Kulik [28], Kulik and Soulier [108], we can extend Theorem 2.4.1 for $p = \infty$ to hold for ℓ^1-functionals under $\mathbf{CS}_1$. Then, for $g(\mathbf{x}) := \mathbb{1}(\|\mathbf{x}\|_1 > 1)$ and $p = \infty$ on the right-hand side of (2.4.2), we find $k' = k'_n \to \infty$ such that

$$\frac{\sum_{t=1}^m \mathbb{1}(\|\mathbf{B}_t\|_1 > \|\mathbf{B}\|_{\infty,(k')})}{\sum_{t=1}^n \mathbb{1}(|\mathbf{X}_t| > \|\mathbf{B}\|_{\infty,(k')})} \xrightarrow{\mathbb{P}} c(1) , \quad n \to \infty , \tag{2.4.7}$$

where as in (2.4.3), $k'_n \sim c(\infty)\, n\mathbb{P}(|\mathbf{X}_0| > x_b)$ then this value gets shrunken by a factor $c(\infty) \leq 1$ compared to the i.i.d. case. This is not the case for ℓ^1-cluster inference since $k_n \sim c(1) n\mathbb{P}(|\mathbf{X}_0| > x_b)$ with $c(1) \geq 1$ which suggest the choice of k_n in (2.4.6) is more robust. For $\alpha \in (0,1)$ we can argue similarly and improve inference by proposing an estimator based on ℓ^α-norm order statistics.

As in the extremal index example, we promote inference based on extremal ℓ^p–blocks with $p < \infty$ for a better control of the tuning parameter k_n compared to inference with extremal ℓ^∞–blocks.

Furthermore, Theorem 2.4.1 yields estimators of the parameters of the α-stable limit for $(\mathbf{S}_n/a_n)$ denoted $\boldsymbol{\xi}_\alpha$. Indeed, following the theory in Bartkiewicz *et al.* [5], we characterize the α-stable limit in terms of $\mathbf{Q}^{(1)}$; the proof is given Section 2.7.4.

Proposition 2.4.4. *Assume that $(\mathbf{X}_t)$ is a regularly varying stationary sequence with $\alpha \in (0,1) \cup (1,2)$ together with the mixing condition:*

$$\mathbb{E}[\mathrm{e}^{\,i\mathbf{u}^\top \mathbf{S}_n/a_n}] = (\mathbb{E}[\mathrm{e}^{\,i\mathbf{u}^\top \mathbf{S}_{b_n}/a_n}])^{m_n} + o(1)\,, \qquad n \to \infty\,, \quad \mathbf{u} \in \mathbb{R}^d\,,$$

and the anti-clustering condition: for every $\delta > 0$,

$$\lim_{l\to\infty} \limsup_{n\to\infty} n \textstyle\sum_{t=l}^{b_n} \mathbb{E}[(|\mathbf{X}_t/a_n| \wedge \delta)\,(|\mathbf{X}_0/a_n| \wedge \delta)] = 0\,. \tag{2.4.8}$$

Then $\mathbf{S}_n/a_n \xrightarrow{d} \boldsymbol{\xi}_\alpha$ for an α-stable random vector $\boldsymbol{\xi}_\alpha$ with characteristic function $\mathbb{E}[\exp(i\mathbf{u}^\top \boldsymbol{\xi}_\alpha)] = \exp(-c_\alpha\, \sigma_\alpha(\mathbf{u})\,(1 - i\,\beta(\mathbf{u})\tan(\alpha\pi/2)))$, $\mathbf{u} \in \mathbb{R}^d$, where $c_\alpha := (\Gamma(2-\alpha)/|1-\alpha|)(1\wedge\alpha)\cos(\alpha\pi/2)$, and the scale and skewness parameters have representation

$$\begin{aligned} \sigma_\alpha(\mathbf{u}) &:= c(1)\,\mathbb{E}[|\mathbf{u}^\top \textstyle\sum_{t\in\mathbb{Z}} \mathbf{Q}_t^{(1)}|^\alpha]\,, \\ \beta(\mathbf{u}) &:= \big(\mathbb{E}[(\mathbf{u}^\top \textstyle\sum_{t\in\mathbb{Z}} \mathbf{Q}_t^{(1)})_+^\alpha - (\mathbf{u}^\top \textstyle\sum_{t\in\mathbb{Z}} \mathbf{Q}_t^{(1)})_-^\alpha]\big)/\mathbb{E}[|\mathbf{u}^\top \textstyle\sum_{t\in\mathbb{Z}} \mathbf{Q}_t^{(1)}|^\alpha]\,. \end{aligned}$$

As for $c(1)$, an application of Theorem 2.4.1 with $p = 1$ for $\alpha \in (1,2)$ and $p = \alpha$ for $\alpha \in (0,1)$ yields natural estimators of the parameters $(\sigma_\alpha(\mathbf{u}), \beta(\mathbf{u}))$ in the central limit theorem of Proposition 2.4.4

An example: a regularly varying linear process

We illustrate the index estimators of Corollaries 2.4.2 and 2.4.3 for a regularly varying linear process $X_t := \sum_{j\in\mathbb{Z}} \varphi_j Z_{t-j}$, $t \in \mathbb{Z}$, where (Z_t) is an iid real-valued regularly varying sequence (Z_t) with (tail)-index $\alpha > 0$, and (φ_j) are real coefficients such that $\sum_{j\in\mathbb{Z}} |\varphi_j|^{1\wedge(\alpha-\varepsilon)} < \infty$ for some $\varepsilon > 0$.

In this setting, (X_t) is regularly varying with the same (tail)-index $\alpha > 0$, and the distributions of Z_t and X_t are tail-equivalent; see Davis and Resnick [46]. The spectral cluster process of (X_t) is given by $Q_t^{(\alpha)} = (\varphi_{t+J}/\|(\varphi_t)\|_\alpha)\,\Theta_0^Z$, $t \in \mathbb{Z}$, where $\lim_{x\to\infty} \mathbb{P}(\pm Z_0 > x)/\mathbb{P}(|Z_0| > x) = \mathbb{P}(\Theta_0^Z = \pm 1)$, Θ_0^Z is independent of a random shift J with distribution $\mathbb{P}(J = j) = |\varphi_j|^\alpha/\|(\varphi_t)\|_\alpha^\alpha$; see Kulik and Soulier [108], (15.3.9). Then

$$c(\infty) \;=\; \max_{t\in\mathbb{Z}} |\varphi_t|^\alpha/\|\varphi\|_\alpha^\alpha\,, \qquad c(1) \;=\; \big(\textstyle\sum_{t\in\mathbb{Z}} |\varphi_t|\big)^\alpha/\|\varphi\|_\alpha^\alpha\,.$$

For the causal AR(1) model given by $X_t = \varphi\, X_{t-1} + Z_t$, $t \in \mathbb{Z}$, $|\varphi| < 1$, one retrieves $\theta_{|X|} = c(\infty) = 1 - |\varphi|^\alpha$ and $c(1) = (1 - |\varphi|^\alpha)/(1 - |\varphi|)^\alpha$.

We aim at comparing estimators of $\theta_{|X|}$ and $c(1)$ built on extremal ℓ^p–blocks, $p < \infty$, against inference over ℓ^∞–blocks, for the causal AR(1) model with student(α) noise for $\alpha = 1.3$. Guided by

(2.4.3), we take $k = k_n := \max\{2, \lfloor n/b_n^{(1+\kappa)} \rfloor\}$ with $\kappa = 1$. For estimation of α, we follow the bias-correction procedure in de Haan *et al.* [83]. This estimator is plugged into (2.4.4), (2.4.6), resulting in the estimators $\widehat{\theta}_{|\mathbf{X}|}$, $1/\widehat{c}(1)$, based on ℓ^α- and ℓ^1-extremal blocks, respectively. Figures 2.4.5 and 2.4.6 present boxplots of these estimators as a function of b_n and for different sample sizes n in blue. For comparison, we also show boxplots of the estimators in (2.4.5) and (2.4.7) based on extremal ℓ^∞–blocks in white. Estimation based on ℓ^p–block for $p < \infty$ outperforms in simulation the ℓ^∞–blocks approach in terms of bias for fixed k. Also, notice the bias for large block lengths decreases as n increases. Indeed, if we fix n, the relation $\lfloor n/b^{(1+\kappa)} \rfloor \to 0$ as $b \to \infty$ restricts the block length for small sample sizes. We have fixed the tuning parameter $\kappa = 1$, though improvement can be achieved by fine-tuning $\kappa > 0$. We also refer to Buriticá *et al.* [25] for further simulation experiences showing that the estimator of the extremal index in (2.4.4) compares favorably with various classical estimators as regards bias.

2.5 A discussion of the assumptions of the large deviation principle in Theorem 2.2.1

Consider a stationary sequence $(\mathbf{X}_t)$ satisfying $\mathbf{RV}_\alpha$ and let (x_n) be a threshold sequences such that $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$. In the conditions $\mathbf{AC}$ and $\mathbf{CS}_p$ below we refer to the same sequence (x_n). We will discuss the conditions of Theorem 2.2.1.

2.5.1 Anti-clustering condition AC.

For any $\delta > 0$,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\|\mathbf{X}_{[k,n]}\|_\infty > \delta\, x_n \mid |\mathbf{X}_0| > \delta\, x_n\big) = 0.$$

Condition $\mathbf{AC}$ ensures that a large value at present time does not persist indefinitely in the extreme future of the time series. This anti-clustering is weaker than the more common two-sided one:

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(\max_{k\le |t|\le n} |\mathbf{X}_t| > \delta\, x_n \mid |\mathbf{X}_0| > \delta\, x_n\big) = 0. \tag{2.5.1}$$

A simple sufficient condition, which breaks block-wise extremal dependence into pair-wise, is given by

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{t=k}^{n} \mathbb{P}\big(|\mathbf{X}_t| > \delta\, x_n \mid |\mathbf{X}_0| > \delta\, x_n\big)\,.$$

For m-dependent $(\mathbf{X}_t)$ the latter condition turns into $n\,\mathbb{P}(|\mathbf{X}_0| > \delta\, x_n) \to 0$ which is always satisfied.

If $p \le \alpha$ an extra assumption is required for controlling the accumulation of moderate extremes within a block.

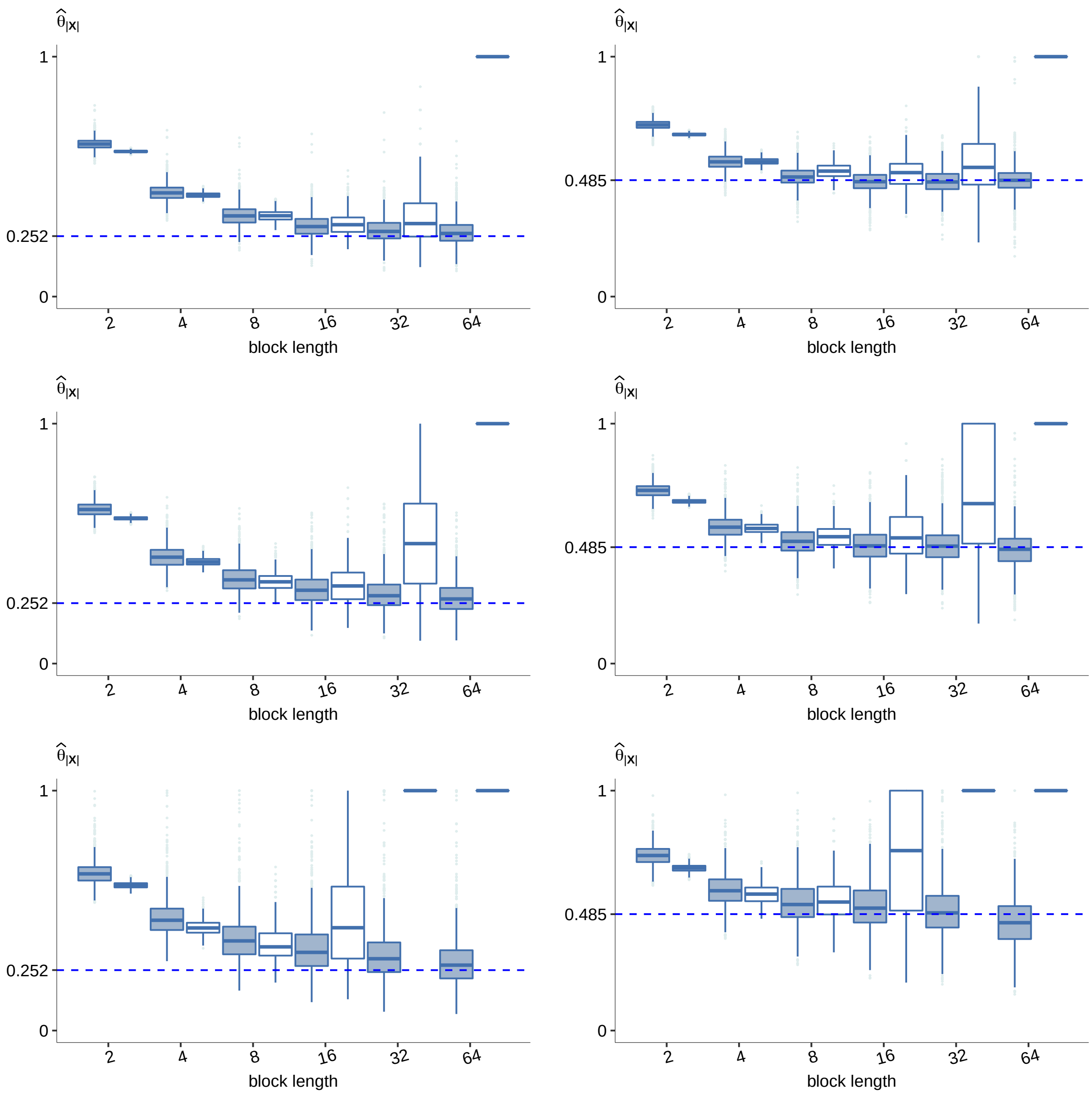


Figure 2.4.5. Boxplot of estimates $\widehat{\theta}_{|X|}$ as a function of b_n from (2.4.4) for inference through $\mathbf{Q}^{(\alpha)}$ (in blue) and from (2.4.5) through $\mathbf{Q}^{(\infty)}$ (in white). 1 000 simulated samples $(X_t)_{t=1,\dots,n}$ from a causal AR(1) model with student(α) noise with $\alpha = 1.3$ and $\varphi = 0.8$ (left column), $\varphi = 0.6$ (right column) were considered. Rows correspond to results for $n = 8\,000, 3\,000, 1\,000$ from top to bottom.

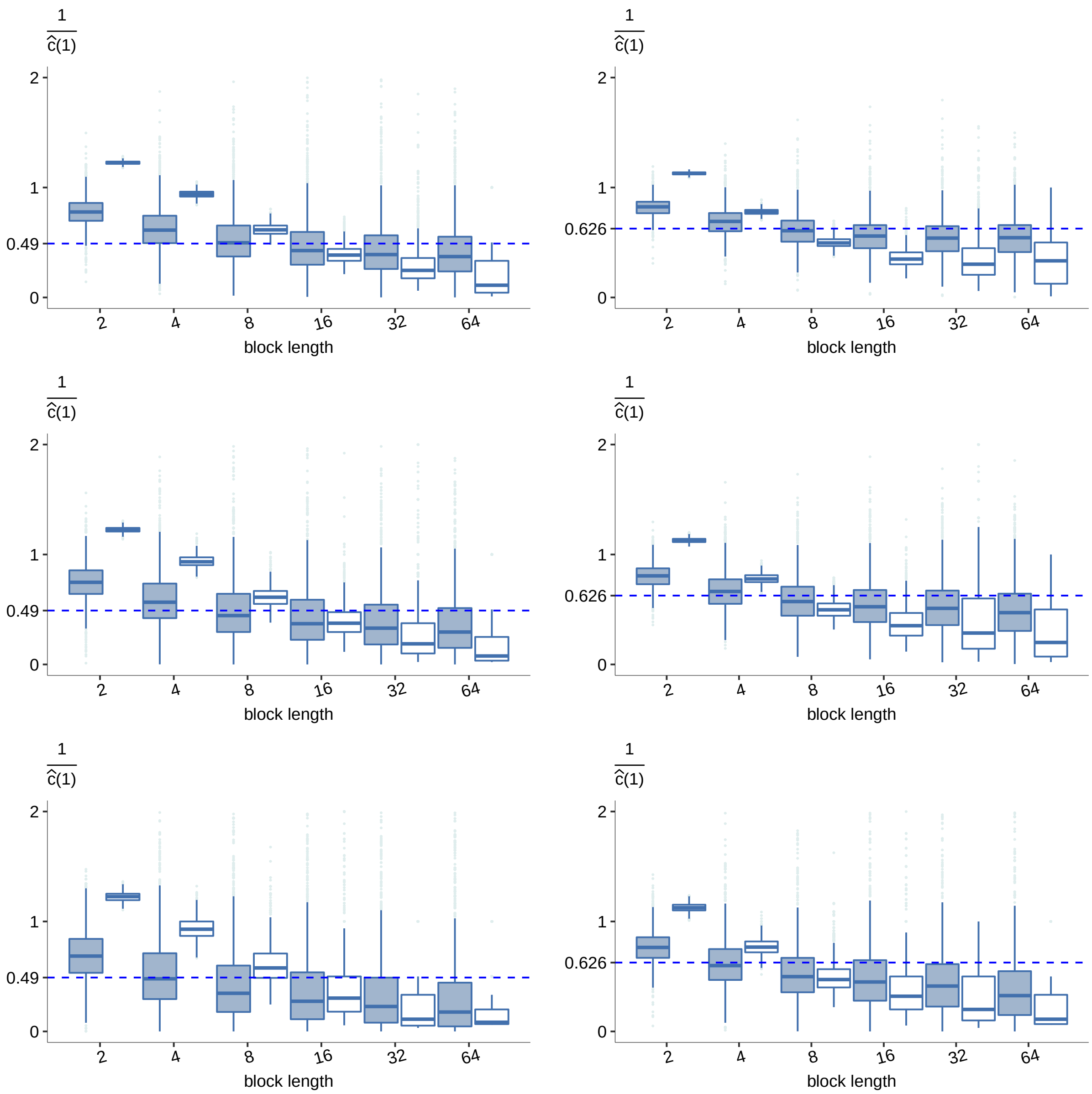


Figure 2.4.6. Boxplot of estimates $1/\widehat{c}(1)$ as a function of b_n from (2.4.6) for inference through $\mathbf{Q}^{(1)}$ (in blue) and from (2.4.7) for inference through $\mathbf{Q}^{(\infty)}$ (in white). We simulate 1 000 samples $(X_t)_{t=1,\dots,n}$ from an AR(1) model with student(α) noise; models are the same as in Figure 2.4.5. Rows correspond to $n = 8\,000, 3\,000, 1\,000$ from top to bottom.

2.5.2 Vanishing-small-values condition $\mathbf{CS}_p$.

For $p \in (0,\alpha]$ we assume that for a sequence (x_n) satisfying $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ and for any $\delta > 0$, we have

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} \frac{\mathbb{P}\big(\big|\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p - \mathbb{E}\big[\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p\big]\big| > \delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = 0. \tag{2.5.2}$$

We refer to (2.5.2) as condition $\mathbf{CS}_p$ in what follows. Indeed, if $\alpha < p < \infty$ then by Karamata's theorem (see Bingham *et al.* [13]) and since $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$,

$$\mathbb{E}[\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}|_p^p] = n\,\mathbb{E}[|\overline{x_n^{-1}\mathbf{X}}^{\epsilon}|^p] = o(1)\,, \qquad n\to\infty\,.$$

Also, if $p < \alpha$, then $\mathbb{E}[|\mathbf{X}|^p] < \infty$. If we also have $n/x_n^p \to 0$ then

$$\mathbb{E}[\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p] \le n\,x_n^{-p}\mathbb{E}[|\mathbf{X}|^p] \to 0\,, \qquad n\to\infty\,.$$

If $p = \alpha$, $\mathbb{E}[|\mathbf{X}|^\alpha] < \infty$ and $n/x_n^\alpha \to 0$ then the latter relation remains valid. If $\mathbb{E}[|\mathbf{X}|^\alpha] = \infty$ then $\mathbb{E}[|\overline{x_n^{-1}\mathbf{X}}^{\epsilon}|^\alpha] = x_n^{-\alpha}\,\ell(x_n)$ for some slowly varying function ℓ depending on ϵ, hence for any small $\kappa > 0$ and large n, $\ell(x_n) \le x_n^\kappa$. Then the condition $nx_n^{-\alpha+\kappa} \to 0$ also implies that $\mathbb{E}[\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_\alpha^\alpha] = o(1)$. Thus we retrieve $\mathbf{CS}_p$ as used in Theorem 2.2.1. In sum, under the aforementioned additional growth conditions on (x_n), if (2.5.2) holds then it holds without the centering term holds too as is the case for condition $\mathbf{CS}_p$ written in Theorem 2.2.1.

We mentioned that conditions of a similar type as $\mathbf{CS}_p$ are standard when dealing with sum functionals acting on $(\mathbf{X}_t)$ (see for example Davis and Hsing [40], Bartkiewicz *et al.* [5], Mikosch and Wintenberger [121, 122, 123]), and are also discussed in Kulik and Soulier [108].

Remark 2.5.1. *Assume $\alpha < p < \infty$. Then by an applications of Markov's inequality of order 1 and Karamata's theorem yield for $\delta > 0$, as $n \to \infty$,*

$$\begin{aligned}
\frac{\mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p > \delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} &= \frac{\mathbb{P}\big(\sum_{t=1}^n |\overline{x_n^{-1}\mathbf{X}_t}^{\epsilon}|^p > \delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\
&\le \frac{\mathbb{E}[|\overline{x_n^{-1}\mathbf{X}_0}^{\epsilon}|^p]}{\delta\,\mathbb{P}(|\mathbf{X}_0| > \epsilon\,x_n)}\,\frac{\mathbb{P}(|\mathbf{X}_0| > \epsilon\,x_n)}{\mathbb{P}(|\mathbf{X}_0| > x_n)} \to c\,\epsilon^{p-\alpha}\,.
\end{aligned}$$

The right-hand side converges to zero as $\epsilon \to 0$. Here and in what follows, c denotes any positive constant whose value is not of interest. We conclude that (2.5.2) is automatic for $p > \alpha$.

Remark 2.5.2. *Condition $\mathbf{CS}_p$ is challenging to check for $p \le \alpha$. For $p/\alpha \in (1/2, 1]$, by Čebyshev's* *inequality,*

$$
\begin{aligned}
&\mathbb{P}\big(\big|\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p - \mathbb{E}[\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p]\big| > \delta\big)/[n\,\mathbb{P}(|\mathbf{X}_0| > x_n)] \\
\leq\;& \delta^{-2}\mathrm{var}\big(\|\overline{x_n^{-1}\mathbf{X}_{[1,n]}}^{\epsilon}\|_p^p\big)/[n\,\mathbb{P}(|\mathbf{X}_0| > x_n)] \\
\leq\;& \delta^{-2}\frac{\mathbb{E}[|\overline{x_n^{-1}\mathbf{X}_0}^{\epsilon}|^{2p}]}{\mathbb{P}(|\mathbf{X}_0| > x_n)}\Big[1 + 2\sum_{h=1}^{n-1}|\operatorname{corr}\big(|\overline{x_n^{-1}\mathbf{X}_0}^{\epsilon}|^p, |\overline{x_n^{-1}\mathbf{X}_h}^{\epsilon}|^p\big)|\Big].
\end{aligned}
$$

Now assume that $(\mathbf{X}_t)$ is ρ-mixing with summable rate function (ρ_h); cf. Bradley [14]. Then the right-hand side is bounded by

$$
\delta^{-2}\,\frac{\mathbb{E}[|\overline{x_n^{-1}\mathbf{X}}^{\epsilon}|^{2p}]}{\mathbb{P}(|\mathbf{X}_0| > x_n)}\Big[1 + 2\sum_{h=1}^{\infty}\rho_h\Big] \sim \delta^{-2}\epsilon^{2p-\alpha}\Big[1 + 2\sum_{h=1}^{\infty}\rho_h\Big], \qquad \epsilon \to 0,
$$

*where we applied Karamata's theorem in the last step, and $\mathbf{CS}_p$ follows. For Markov chains weaker assumptions such as the drift condition (**DC**) in Mikosch and Wintenberger [122, 123] can be used for checking $\mathbf{CS}_p$.*

Remark 2.5.3. *Condition $\mathbf{CS}_p$ not only restricts the serial dependence of the time series $(\mathbf{X}_t)$ but also the level of thresholds (x_n). Indeed for $p/\alpha < 1/2$ and $(\mathbf{X}'_t)$ iid, since $\big(\|n^{-1/2}\mathbf{X}'_{[1,n]}\|_p^p - \mathbb{E}[\|n^{-1/2}\mathbf{X}'_{[1,n]}\|_p^p]\big)$ converges in distribution to a Gaussian limit by virtue of the central limit theorem, $\mathbf{CS}_p$ implies necessarily that $x_n/\sqrt{n} \to \infty$ as $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$.*

2.5.3 Threshold condition

In Theorem 2.2.1 we assume growth conditions on (x_n): $n/x_n^p \to 0$ if $p < \alpha$ and $n/x_n^{\alpha-\kappa} \to 0$ for some $\kappa > 0$ if $p = \alpha$.

For inference purposes it is tempting to decrease the threshold level x_n such that more exceedances are included in the estimators. Indeed, the assumptions on (x_n) can be relaxed, justified by results such as Nagaev's large deviation principle in [124], by adding a centering term as we will show in Lemma 2.5.4. However, in this section we aim at pointing at the difficulties that might arise while doing so in practice.

To motivate the results of this section we start by considering an iid sequence $(\mathbf{X}_t)$ satisfying $\mathbf{RV}_\alpha$ for some $\alpha > 0$. Then, for $p > \alpha$, (2.2.3) holds with limit $c(p) = 1$ and $S_n^{(p)} = \sum_{t=1}^n |\mathbf{X}_i|^p$ has infinite expectation. If $p < \alpha$ the process $(S_n^{(p)})$ has finite expectation and by the law of large numbers, for $n/x_n^p \to 0$,

$$
\begin{aligned}
&\mathbb{P}\big(\|\mathbf{X}_{[0,n]}\|_p > x_n\,(n\,x_n^{-p}\,\mathbb{E}[|\mathbf{X}|^p] + 1)^{1/p}\big) \\
=\;& \mathbb{P}\big(S_n^{(p)} - \mathbb{E}[S_n^{(p)}] > x_n^p(1 + o(1))\big) \to 0\,.
\end{aligned} \tag{2.5.3}
$$

Following Nagaev [124], a large deviation result for the centered process holds:

$$\mathbb{P}\Big(S_n^{(p)} - \mathbb{E}[S_n^{(p)}] > x_n^p\Big) \sim n\,\mathbb{P}(|\mathbf{X}_0| > x_n)\,, \qquad n \to \infty\,,$$

provided $n/x_n^{\alpha-\kappa} \to 0$ for $p/\alpha \in (1/2, 1)$ and some $\kappa > 0$, and $\sqrt{n\,\log n}/x_n^p \to 0$ for $p/\alpha < 1/2$. These conditions are satisfied for extreme thresholds satisfying $n/x_n^p \to 0$: in this case the centering term $\mathbb{E}[S_n^{(p)}]$ in (2.5.3) is always negligible which allows us to derive (2.2.3). Next, we extend the previous ideas to regularly varying time series.

Lemma 2.5.4. *Consider an $\mathbb{R}^d$-valued stationary process $(\mathbf{X}_t)$ satisfying the conditions $\mathbf{RV}_\alpha$, $\mathbf{AC}$, $\mathbf{CS}_p$ and $c(p) < \infty$ for some $p > 0$. If $p < \alpha$ then*

$$\lim_{n\to\infty} \frac{\mathbb{P}\big(\|\mathbf{X}_{[0,n]}\|_p > x_n\,\big(n\,x_n^{-p}\,\mathbb{E}[|\mathbf{X}|^p] + 1\big)^{1/p}\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = c(p)\,. \tag{2.5.4}$$

If $p = \alpha$ then

$$\lim_{n\to\infty} \frac{\mathbb{P}\big(\|\mathbf{X}_{[0,n]}\|_\alpha > x_n\,\big(n\mathbb{E}[|\overline{\mathbf{X}/x_n}^1|^\alpha] + 1\big)^{1/\alpha}\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = c(\alpha) = 1\,. \tag{2.5.5}$$

Moreover, if also $\mathbb{E}[|\mathbf{X}|^\alpha] < \infty$ then equation (2.5.4) holds for $p = \alpha$.

The proof is given in Section 2.7.2. Now the restrictions on the level of the thresholds (x_n) are the ones implicitly implied by condition $\mathbf{CS}_p$ in (2.5.2); see Remark 2.5.3.

We define an auxiliary sequence of levels:

$$z_n := z_n(p) = \begin{cases} x_n\,\big(n\,x_n^{-p}\mathbb{E}[|\mathbf{X}|^p] + 1\big)^{1/p} & \text{if } p < \alpha, \\ x_n\,\big(n\mathbb{E}[|\overline{\mathbf{X}/x_n}^1|^\alpha] + 1\big)^{1/\alpha} & \text{if } p = \alpha\,, \\ x_n & \text{if } p > \alpha\,. \end{cases}$$

For thresholds satisfying the growth conditions $n/x_n^p \to 0$ we have $z_n \sim x_n$, while for moderate thresholds verifying $\mathbf{CS}_p$ and $n/x_n^p \to \infty$ this is no longer the case.

For the purposes of inference Lemma 2.5.4 is not as satisfactory as (2.2.3) in Theorem 2.2.1. Indeed, the level z_n in the selection of the exceedances is not the original threshold x_n. For any moderate threshold x_n with $z_n/x_n \to \infty$ the use of x_n instead of z_n might include a bias. As a toy example, consider the problem of inferring the constant $c(q)/c(p)$ for $p < \alpha$, $q > p$. Then an application of Lemma 2.5.4 ensures that

$$\begin{aligned}&\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_q > z_n(q) \mid \|\mathbf{X}_{[0,n]}\|_p > z_n(p)) \\ &\qquad \to \mathbb{P}(\|Y\mathbf{Q}^{(p)}\|_q > 1) = \mathbb{E}[\|\mathbf{Q}^{(p)}\|_q^\alpha] = c(q)/c(p), \qquad n \to \infty\,.\end{aligned}$$

However, choosing moderate thresholds (x_n), we would have

$$\begin{aligned}\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_q > x_n \mid \|\mathbf{X}_{[0,n]}\|_p > x_n) &\sim \frac{\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_q > x_n)}{\mathbb{P}(n^{1/p}\mathbb{E}[|\mathbf{X}|^p]^{1/p} > x_n)} \\ &\to \begin{cases} 1 & \text{if } q < \alpha, \\ 0 & \text{if } q > \alpha, \quad n \to \infty\,. \end{cases}\end{aligned}$$

By this argument, the assumption $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$ is justified to simplify inference procedures. Otherwise, the choice of the threshold sequence becomes complicated.

2.6 Inference beyond shift-invariant functionals

So far we only considered inference for shift-invariant functionals acting on $(\tilde{\ell}^p, \tilde{d}_p)$ such as maxima and sums. Following the shift-projection ideas in Janssen [96], jointly with continuous mapping arguments, we extend inference to functionals on (ℓ^p, d_p).

2.6.1 Inference for cluster functionals in (ℓ^p, d_p)

Let $g : (\ell^p, d_p) \to \mathbb{R}$ be a bounded measurable function. We define the functional $\psi_g : (\widetilde{\ell^p}, \widetilde{d_p}) \to \mathbb{R}$ by

$$[\mathbf{z}] \mapsto \psi_g([\mathbf{z}]) \;:=\; \sum_{j\in\mathbb{Z}} |\mathbf{z}^*_{-j}|^\alpha g\big((B^j \mathbf{z}^*_t)_{t\in\mathbb{Z}}\big), \tag{2.6.1}$$

where $\mathbf{z}^*_t := \mathbf{z}_{t-T^*(\mathbf{z})}$, for $t \in \mathbb{Z}$, such that $T^*(\mathbf{z}) := \inf\{s \in \mathbb{Z} : |\mathbf{z}_s| = \|\mathbf{z}\|_\infty\}$ and $B : \ell^p \to \ell^p$ is the backward-shift map.

We link the distribution of the spectral cluster process $\mathbf{Q}^{(\alpha)}$ and the distribution of the class $[\mathbf{Q}^{(\alpha)}]$ through the mappings (2.6.1) in the next proposition whose proof is given in Section 2.7.5.

Proposition 2.6.1. *The following relation holds for any real-valued bounded measurable function g on ℓ^α*

$$\mathbb{E}[g(\mathbf{Q}^{(\alpha)})] \;=\; \mathbb{E}[\psi_g([\mathbf{Q}^{(\alpha)}])]\,.$$

where ψ_g is as in (2.6.1). This relation remains valid if α is replaced by p, whenever the spectral cluster process in ℓ^p is well defined.

For $p \le \alpha$ the mappings in (2.6.1) are continuous functionals on $(\tilde{\ell}^p, \tilde{d}_p)$ and we can extend Theorem 2.4.1 to continuous functionals on (ℓ^p, d_p) evaluated at the spectral cluster process $\mathbf{Q}^{(p)}$.

Theorem 2.6.2. *Assume the conditions of Theorem 2.4.1 for $p \le \alpha$. Then for any continuous* *bounded function* $g : \ell^p \cap \{\mathbf{x} : \|\mathbf{x}\|_p = 1, |\mathbf{x}_0| > 0\} \to \mathbb{R}$,

$$\begin{aligned} \widehat{g}^{(p)} \quad &:= \quad \frac{1}{k}\sum_{t=1}^{m} \underbrace{\sum_{j=1}^{b} W_{j,t}(p)\, g\Big(\frac{B^{j-1}\mathbf{B}_t}{\|\mathbf{B}_t\|_p}\Big)}_{=:\psi_g(\mathbf{B}_t/\|\mathbf{B}_t\|_p)}\, \mathbb{1}(\|\mathbf{B}_t\|_p > \|\mathbf{B}\|_{p,(k)}) \\ &\xrightarrow{\mathbb{P}} \quad \mathbb{E}[g(\mathbf{Q}^{(p)})]\,, \qquad n \to \infty\,, \end{aligned} \tag{2.6.2}$$

where $W_{j,t}(p) = |\mathbf{X}_{(t-1)b+j}|^{\alpha}/\|\mathbf{B}_t\|_p^{\alpha}$ *for all* $1 \le j \le b$.

The proof is given in Section 2.7.5.

2.6.2 Applications

Examples of non-shift-invariant functionals on (ℓ^p, d_p) are measures of serial dependence, probabilities of large deviations such as the supremum of a random walk and ruin probabilities, and functionals of the spectral tail process Θ. We study these examples in the remainder of this section.

Measures of serial dependence

Define $g_h(\mathbf{x}_t) = |\mathbf{x}_h|^{\alpha}\frac{\mathbf{x}_0^{\top}}{|\mathbf{x}_0|}\frac{\mathbf{x}_h}{|\mathbf{x}_h|}$. Then the following result is straightforward from Theorem 2.6.2.

Corollary 2.6.3. *Assume the conditions of Theorem 2.6.2 for* $p = \alpha$. *Then*

$$\begin{aligned} \widehat{g}_h^{(\alpha)} &:= \frac{1}{k}\sum_{t=1}^{m} \underbrace{\sum_{j=1}^{b-h} W_{j,t}W_{j+h,t}\,\frac{\mathbf{X}_{j,t}^{\top}}{|\mathbf{X}_{j,t}|}\,\frac{\mathbf{X}_{j+h,t}}{|\mathbf{X}_{j+h,t}|}}_{=:\psi_{g_h}(\mathbf{B}_t)}\, \mathbb{1}(\|\mathbf{B}_t\|_{\alpha} > \|\mathbf{B}\|_{\alpha,(k)}). \\ &\xrightarrow{\mathbb{P}} \mathbb{E}[g_h(\mathbf{Q}^{(a)})], \qquad n \to +\infty, \end{aligned}$$

where the weights $W_{j,t} = W_{j,t}(\alpha)$ *are defined in Theorem 2.6.2, satisfying* $\sum_{j=1}^{b} W_{j,t} = 1$, *and* $\mathbf{X}_{j,t} := \mathbf{X}_{(t-1)b+j}$ *for* $1 \le j \le b$.

The function g_h gives a summary of the magnitude and direction of the time series h lags after recording a high-level exceedance of the norm, and satisfies the relation $\sum_{h\in\mathbb{Z}} \mathbb{E}[g_h(\mathbf{Q}^{(\alpha)})] = 1$.

Example 2.6.4. Let (X_t) be a linear process satisfying the assumptions in Example 2.4.3, then

$$\mathbb{E}[g_h(Q^{(\alpha)})] = \frac{\sum_{t\in\mathbb{Z}} |\varphi_t|^{\alpha}|\varphi_{t+h}|^{\alpha}\mathrm{sign}(\varphi_t)\mathrm{sign}(\varphi_{t+h})}{\big(\|\varphi\|_{\alpha}^{\alpha}\big)^2}, \qquad h \in \mathbb{Z}\,.$$

This function is proportional to the autocovariance function of a finite variance linear process with coefficients $(|\varphi_t|^{\alpha}\,\mathrm{sign}(\varphi_t))$. In particular, for $\alpha = 1$ it is proportional to the autocovariance function of a finite variance linear process with coefficients (φ_t).

Large deviations for the supremum of a random walk

We start by reviewing Theorem 4.5 in Mikosch and Wintenberger [123]; the proof is given in Section 2.7.5.

Proposition 2.6.5. *Consider a univariate stationary sequence (X_t) satisfying $\mathbf{RV}_\alpha$ for some $\alpha \geq 1$, $\mathbf{AC}$, $\mathbf{CS}_1$, and $c(1) < \infty$. Then for all $p \geq 1$,*

$$\Big| \frac{\mathbb{P}(\sup_{1\leq t\leq n} S_t > x_n)}{n\mathbb{P}(|X| > x_n)} \\ - c(p)\,\mathbb{E}\Big[\lim_{s\to\infty}\Big(\sup_{t\geq -s}\sum_{i=-s}^{t} Q_i^{(p)}\Big)_+^{\alpha}\Big]\Big| \to 0, \quad n\to\infty. \tag{2.6.3}$$

If $\alpha \geq 1$, then $\|Q^{(1)}\|_\alpha^\alpha \leq \|Q^{(1)}\|_1^\alpha = 1$ and a consistent estimator of $c(1) = 1/\mathbb{E}[\|Q^{(1)}\|_\alpha^\alpha]$ was suggested in Corollary 2.4.3. A consistent estimator of the term in (2.6.3) is given next.

Corollary 2.6.6. *Assume the conditions of Theorem 2.6.2 for $p=1$. Then*

$$\Big| \frac{\sum_{t=1}^m \big(\sup_{1\leq j\leq b}\frac{X_{t,j}}{\|\mathbf{B}_t\|_1}\big)_+^\alpha \mathbb{1}(\|\mathbf{B}_t\|_1 > \|\mathbf{B}\|_{1,(k)})}{\sum_{t=1}^m \frac{\|\mathbf{B}_t\|_\alpha^\alpha}{\|\mathbf{B}_t\|_1^\alpha}\mathbb{1}(\|\mathbf{B}_t\|_1 > \|\mathbf{B}\|_{1,(k)})} \\ - c(1)\,\mathbb{E}\Big[\lim_{s\to\infty}\Big(\sup_{t\geq -s}\sum_{i=-s}^{t} Q_i^{(1)}\Big)_+^{\alpha}\Big]\Big| \xrightarrow{\mathbb{P}} 0\,, \qquad n\to\infty\,,$$

where $X_{t,j} := X_{(t-1)b+j}$, for $1\leq j\leq b$, $1\leq t\leq m$.

Following the same ideas and using Theorem 4.9 in [123], one can also derive a consistent estimator for the constant in the related ruin problem.

Application: a cluster-based method for inferring on (Θ_t)

Exploiting the relation $(\mathbf{Q}_t^{(\alpha)})/|\mathbf{Q}_0^{(\alpha)}| \overset{d}{=} (\Theta_t)$ discussed in Section 2.3.1, we propose cluster-based estimation methods for the spectral tail process.

Cluster-based approaches with the goal to improve inference on Θ_1 for Markov chains were considered in Drees *et al.* [53]; see also Davis *et al.* [39], Drees *et al.* [50] for related cluster-based procedures on $(\Theta_t)_{|t|\leq h}$ for fixed $h\geq 0$. Our approach can be seen as an extension for inference on the ℓ^α-valued sequence (Θ_t).

Consider the continuous re-normalization function $\zeta(\mathbf{x}) = \mathbf{x}/|\mathbf{x}_0|$ on $\{\mathbf{x}\in\ell^\alpha : |\mathbf{x}_0| > 0\}$. We derive the following result from Theorem 2.6.2; the proof is given in Section 2.7.5.

Proposition 2.6.7. *Assume the conditions of Theorem 2.6.2 for $p=\alpha$. Let $\rho : (\ell^\alpha, d_\alpha)\to\mathbb{R}$ be a homogeneous continuous function and $\rho_\zeta(\mathbf{x}) := (\rho^\alpha\wedge 1)\circ\zeta(\mathbf{x})$. Then*

$$\widehat{\rho_\zeta}^{(\alpha)} := \frac{1}{k}\sum_{t=1}^m \psi_{\rho_\zeta}(\mathbf{B}_t)\,\mathbb{1}(\|\mathbf{B}_t\|_p > \|\mathbf{B}\|_{p,(k)}) \xrightarrow{\mathbb{P}} \mathbb{P}\big(\rho(Y\,\Theta) > 1\big)\,, \qquad n\to\infty\,,$$

where $\psi_{\rho_\zeta}(\mathbf{B}_t)$ is defined in (2.6.2) and the Pareto(α) random variable Y is independent of Θ.

Classical examples of such functionals are $\rho(\mathbf{x}) = \max_{i\geq 0, j\geq i}(x_i - x_j)_+$, functionals related to large deviations such as $\rho(\mathbf{x}) = \sup_{t\geq 0}(\sum_{i=0}^t x_i)_+$, or measures of serial dependence such as $\rho(\mathbf{x}) = |\mathbf{x}_h|$.

2.7 Proofs

2.7.1 Proof of Theorem 2.2.1

Recall the properties of the sequence (x_n) from Section 2.5, in particular $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$. The main result in Theorem 2.2.1 follows by applications of Lemma 2.7.1 and Proposition 2.7.2 below; their proofs are given at the end of this section.

Lemma 2.7.1. *Consider an $\mathbb{R}^d$-valued stationary time series $(\mathbf{X}_t)$ satisfying the conditions $\mathbf{RV}_\alpha$, $\mathbf{AC}$, $\mathbf{CS}_p$. If $p < \alpha$, assume also also $n/x_n^p \to 0$ and if $p = \alpha$ also $n/x_n^{\alpha-\kappa} \to 0$ for some $\kappa > 0$. Then the following relation holds*

$$\lim_{n\to\infty} \frac{\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = c(p)\,, \tag{2.7.1}$$

where $c(p)$ is given in (2.3.2).

We recall from Remark 2.5.1 that (2.5.2) in $\mathbf{CS}_p$ is always satisfied for $p > \alpha$. Moreover, for $p \leq \alpha$, under the growth conditions on (x_n) in Theorem 2.2.1, centering with the expectation in (2.5.2) is not necessary.

Proposition 2.7.2. *Assume the conditions of Lemma 2.7.1. Then,*

$$\mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_p > x_n) \xrightarrow{w} \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot), \quad n \to \infty\,, \tag{2.7.2}$$

in the space $(\tilde{\ell}^p, \tilde{d}_p)$ where the Pareto(α) random variable Y and $\mathbf{Q}^{(p)}$ are independent.

Proof of Lemma 2.7.1

Choose some $\epsilon > 0$, $\delta \in (0,1)$. Since $\|a_n^{-1}\mathbf{X}_{[0,n]}\|_p^p$ is a sum of non-negative random variables we have the following bounds via truncation

$$\begin{aligned} \mathbb{P}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > 1\big) &\leq \mathbb{P}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p^p > 1\big) \\ &\leq \mathbb{P}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > (1-\delta^p)\big) + \mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p > \delta^p\big)\,. \end{aligned} \tag{2.7.3}$$

By $\mathbf{CS}_p$ and in view of Remark 2.5.1 we have

$$\lim_{\epsilon\downarrow 0}\limsup_{n\to\infty} \mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p > \delta^p\big)/(n\mathbb{P}(|\mathbf{X}_0| > x_n)) = 0\,. \tag{2.7.4}$$

Now, for any choice of $u > 0$, it remains to determine the limits of the terms $\mathbb{P}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > u)/(n\mathbb{P}(|\mathbf{X}_0| > x_n))$. We start with a telescoping sum representation

$$\begin{aligned}
&\mathbb{P}(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > u) - \mathbb{P}(|\underline{x_n^{-1}\mathbf{X}_0}_\epsilon|_p^p > u) \\
&= \textstyle\sum_{i=1}^n \big(\mathbb{P}(\|\underline{x_n^{-1}\mathbf{X}_{[0,i]}}_\epsilon\|_p^p > u) - \mathbb{P}(\|\underline{x_n^{-1}\mathbf{X}_{[0,i-1]}}_\epsilon\|_p^p > u)\big) \\
&= \textstyle\sum_{i=1}^n \mathbb{E}\big[\big(\mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[0,i]}}_\epsilon\|_p^p > u) - \mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[1,i]}}_\epsilon\|_p^p > u)\big)\,\mathbb{1}(|\mathbf{X}_0| > \epsilon x_n)\big]\,,
\end{aligned}$$

where we used stationarity in the last step and the fact that the difference of the indicator functions vanishes on $\{|\mathbf{X}_0| \le \epsilon x_n\}$. We also observe that the second term on the left-hand side is of the order $o(n\,\mathbb{P}(|\mathbf{X}_0| > x_n))$. For any fixed k write $A_k = \{\max_{k\le t\le n}|\mathbf{X}_t| > \epsilon x_n\}$. Regular variation of $(\mathbf{X}_t)$ ensures that, as $n\to\infty$,

$$\begin{aligned}
&\frac{\mathbb{P}(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > u)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\
&\sim \epsilon^{-\alpha}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[\mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[0,i]}}_\epsilon\|_p^p > u) - \mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[1,i]}}_\epsilon\|_p^p > u)\;\big|\;|\mathbf{X}_0| > \epsilon x_n\big] \\
&\sim \epsilon^{-\alpha}\mathbb{E}\big[\big(\mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[0,k-1]}}_\epsilon\|_p^p > u) - \mathbb{1}(\|\underline{x_n^{-1}\mathbf{X}_{[1,k-1]}}_\epsilon\|_p^p > u)\big) \\
&\qquad \times\mathbb{1}(A_k^c)\;\big|\;|\mathbf{X}_0| > \epsilon x_n\big] + \epsilon^{-\alpha}O\big(\mathbb{P}(A_k \mid |\mathbf{X}_0| > \epsilon x_n)\big)\,,
\end{aligned}$$

where the second term vanishes, first letting $n\to\infty$ and then $k\to\infty$, by virtue of **AC**. Now the regular variation property of $(\mathbf{X}_t)$ implies that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{\mathbb{P}(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p > u)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} \\
&= \lim_{k\to\infty}\epsilon^{-\alpha}\big(\mathbb{P}\big(\textstyle\sum_{t=0}^{k-1}|\epsilon Y\Theta_t|^p\mathbb{1}(|Y\Theta_t| > 1) > u\big) \\
&\qquad -\mathbb{P}\big(\textstyle\sum_{t=1}^{k-1}|\epsilon Y\Theta_t|^p\mathbb{1}(|Y\Theta_t| > 1) > u\big)\big),
\end{aligned}$$

by a change of variable this term equals

$$\begin{aligned}
&= \lim_{k\to\infty}\mathbb{E}\big[\textstyle\int_\epsilon^\infty\big(\mathbb{1}\Big(\sum_{t=0}^{k-1}|y\Theta_t|^p\mathbb{1}(|y\Theta_t| > \epsilon) > u\big) \\
&\qquad -\mathbb{1}\big(\textstyle\sum_{t=1}^{k-1}|y\Theta_t|^p\mathbb{1}(|y\Theta_t| > \epsilon) > u\big)\big)\,d(-y^{-\alpha})\big] \\
&= \lim_{k\to\infty}\mathbb{E}\big[\textstyle\int_0^\infty\big(\mathbb{1}\big(\|\underline{y\Theta_{[0,k-1]}}_\epsilon\|_p^p > u\big) - \mathbb{1}\big(\|\underline{y\Theta_{[1,k-1]}}_\epsilon\|_p^p > u\big)\big)\,d(-y^{-\alpha})\big]\,.
\end{aligned}$$

In the last step we used the fact that the integrand vanishes for $0 \le y \le \epsilon$. The integrand is non-negative and bounded by $\mathbb{1}(y > \epsilon)$ which is integrable. Thus we may take the limit as $k\to\infty$ inside the integral to derive the quantity

$$\mathbb{E}\big[\textstyle\int_0^\infty\Big(\mathbb{1}\big(\|\underline{y\Theta_{[0,\infty]}}_\epsilon\|_p^p > u\big) - \mathbb{1}\big(\|\underline{y\Theta_{[1,\infty]}}_\epsilon\|_p^p > u\big)\Big)\,d(-y^{-\alpha})\big]\,.$$

By monotone convergence as $\epsilon \downarrow 0$ we get the limit

$$u^{-\alpha/p}\mathbb{E}\left[\|\Theta_{[0,\infty]}\|_p^\alpha - \|\Theta_{[1,\infty]}\|_p^\alpha\right] = u^{-\alpha/p}\, c(p)\,; \tag{2.7.5}$$

see (2.3.6). Now an appeal to (2.7.3) with $u = 1$ and $u = 1-\delta^p$ yields

$$\begin{aligned} c(p) &\le \liminf_{n\to\infty} \frac{\mathbb{P}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p^p > 1\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\ &\le \limsup_{n\to\infty} \frac{\mathbb{P}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p^p > 1\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \le (1-\delta^p)^{-\alpha/p}\, c(p)\,. \end{aligned}$$

The limit relation (2.7.1) follows as $\delta \downarrow 0$.

Proof of Proposition 2.7.2

Consider any bounded Lipschitz-continuous function $f : (\tilde{\ell}^p, \widetilde{d}_p) \to \mathbb{R}$. The statement is proved if we can show that

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{E}[f(x_n^{-1}\mathbf{X}_{[0,n]}) \mid \|\mathbf{X}_{[0,n]}\|_p > x_n] \\ &\quad = \quad c(p)^{-1}\mathbb{E}\big[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha f\big(Y\,\Theta/\|\Theta\|_p\big)\big]\,. \end{aligned}$$

In view of Lemma 2.7.1 it suffices to show

$$\begin{aligned} &\lim_{n\to\infty} \frac{\mathbb{E}[f(x_n^{-1}\mathbf{X}_{[0,n]})\,\mathbb{1}(\|\mathbf{X}_{[0,n]}\|_p > x_n)]}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\ &\quad = \quad \mathbb{E}\big[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha f\big(Y\,\Theta/\|\Theta\|_p\big)\big]\,. \end{aligned}$$

In these limit relations we may replace $f(x_n^{-1}\mathbf{X}_{[0,n]})$ by $f(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon)$ since by (2.7.4) for any $\delta > 0$, some $K_f > 0$,

$$\begin{aligned} &\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty} \frac{\mathbb{P}\big(|f(x_n^{-1}\mathbf{X}_{[0,n]}) - f(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon)| > \delta\,, \|\mathbf{X}_{[0,n]}\|_p > x_n\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\ &\quad \le \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty} \frac{\mathbb{P}\big(K_f\, d_p(\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon, \mathbf{0}) > \delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\ &\quad \le \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty} \frac{\mathbb{P}\big(K_f\,\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p > \delta(p)\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} = 0\,, \end{aligned}$$

where $\delta(p)=\delta^p$ for $p\ge 1$ and $=\delta$ for $p\in(0,1)$. We also have for $\delta\in(0,1)$,

$$
\begin{aligned}
G_{n,\epsilon} &= \frac{\mathbb{E}\big[f\big(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\big)\,\big|\mathbb{1}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p>1\big)-\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1\big)\big|\big]}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}\\
&\le c\,\frac{\mathbb{P}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p>1\ge\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p\big)}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}\\
&\le c\,\frac{\mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^{\epsilon}\|_p>\delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}+c\,\frac{\mathbb{P}\big(1\ge\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1-\delta\big)}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}\\
&= G^{(1)}_{n,\epsilon,\delta}+G^{(2)}_{n,\epsilon,\delta}\,.
\end{aligned}
$$

Applying (2.7.4) to $G^{(1)}_{n,\epsilon,\delta}$ and using the calculations in the proof of Lemma 2.7.1 leading to (2.7.5) for $G^{(2)}_{n,\epsilon,\delta}$, we conclude that

$$
\begin{aligned}
\lim_{\epsilon\downarrow 0}\limsup_{n\to\infty}G_{n,\epsilon} &\le \lim_{\epsilon\downarrow 0}\limsup_{n\to\infty}G^{(1)}_{n,\epsilon,\delta}+\lim_{\epsilon\downarrow 0}\limsup_{n\to\infty}G^{(2)}_{n,\epsilon,\delta}\\
&= 0+c\,\big((1-\delta)^{-\alpha/p}-1\big)\downarrow 0\,,\qquad \delta\downarrow 0\,.
\end{aligned}
$$

Thus it suffices to show

$$
\begin{aligned}
\lim_{\epsilon\downarrow 0}\lim_{n\to\infty}&\frac{\mathbb{E}\big[f\big(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1\big)\big]}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}\\
&\qquad= \mathbb{E}\big[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha f\big(Y\,\Theta/\|\Theta\|_p\big)\big]\,. \qquad (2.7.6)
\end{aligned}
$$

This is the goal of the remaining proof.

Choose any $\epsilon>0$. Noticing that $\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}=\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon}$ on $\{|x_n^{-1}\mathbf{X}_0|\le\epsilon\}$, we have

$$
\begin{aligned}
I &:= \mathbb{E}\big[f\big(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1\big)\big]\\
&= \mathbb{E}\Big[\Big(f\big(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1\big)\\
&\qquad -f\big((0,\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon})\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon}\|_p>1\big)\Big)\,\mathbb{1}(|\mathbf{X}_0|>\epsilon\,x_n)\Big]\\
&\quad +\mathbb{E}\big[f\big((0,\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon})\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon}\|_p>1\big)\big]\\
&= \mathbb{E}\Big[\Big(f\big(\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_{\epsilon}\|_p>1\big)\\
&\qquad -f\big((0,\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon})\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,n]}}_{\epsilon}\|_p>1\big)\Big)\,\mathbb{1}(|\mathbf{X}_0|>\epsilon\,x_n)\Big]\\
&\quad +\mathbb{E}\big[f\big((0,\underline{x_n^{-1}\mathbf{X}_{[0,n-1]}}_{\epsilon})\big)\,\mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n-1]}}_{\epsilon}\|_p>1\big)\big]\,.
\end{aligned}
$$

where we used the stationarity in the last step. Using the same idea recursively, we obtain

$$
\begin{aligned}
I &= \textstyle\sum_{j=1}^n \mathbb{E}\big[\big(f\big((\mathbf{0}^{n-j}, \underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon)\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon\|_p > 1\big) \\
&\qquad -f\big((\mathbf{0}^{n-j+1}, \underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon)\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon\|_p > 1\big)\Big)\, \mathbb{1}(|\mathbf{X}_0| > \epsilon\, x_n)\big] \\
&\quad + \mathbb{E}\big[f\big((\mathbf{0}^n, \underline{x_n^{-1}\mathbf{X}_0}_\epsilon)\big)\, \mathbb{1}\big(|\mathbf{X}_0| > \epsilon\, x_n\big)\big]\,,
\end{aligned}
$$

where $\mathbf{0}^k := \{0\}^k$ for $k \geq 1$. By regular variation of $\mathbf{X}_0$ the last right-hand term is $o(n\,\mathbb{P}(|\mathbf{X}_0| > x_n))$. Therefore by regular variation of $(\mathbf{X}_t)$ we obtain as $n \to \infty$,

$$
\begin{aligned}
&I/\big(n\mathbb{P}(|\mathbf{X}_0| > x_n)\big) \\
&\sim \frac{\epsilon^{-\alpha}}{n} \textstyle\sum_{j=1}^n \mathbb{E}\big[f\big((\mathbf{0}^{n-j}, \underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon)\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon\|_p > 1\big) \\
&\qquad -f\big((\mathbf{0}^{n-j+1}, \underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon)\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon\|_p > 1\big) \,\big|\, |\mathbf{X}_0| > \epsilon\, x_n\big] \\
&=: II\,.
\end{aligned}
$$

Write $A_k = \{\|\mathbf{X}_{[k,n]}\|_\infty > \epsilon\, x_n\}$ for fixed $k \geq 1$. By **AC**, $\mathbb{P}(A_k \mid |\mathbf{X}_0| > \epsilon\, x_n)$ vanishes by first letting $n \to \infty$ then $k \to \infty$. Since each of the summands in II is uniformly bounded in absolute value we may restrict the summation to $j \in \{k-1, \ldots, n\}$ for any fixed $k \geq 1$. Therefore we have as $n \to \infty$,

$$
\begin{aligned}
&II - O\big(\mathbb{P}(A_k \mid |\mathbf{X}_0| > \epsilon\, x_n)\big) \\
&\sim \frac{\epsilon^{-\alpha}}{n} \sum_{j=k-1}^n \mathbb{E}\big[\big(f\big((\mathbf{0}^{n-j}, \underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon)\big)\, \mathbb{1}\big(\{\|\underline{x_n^{-1}\mathbf{X}_{[0,j]}}_\epsilon\|_p > 1\}\big) \\
&\quad -f\big((\mathbf{0}^{n-j+1}, \underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon)\big)\, \mathbb{1}\big(\{\|\underline{x_n^{-1}\mathbf{X}_{[1,j]}}_\epsilon\|_p > 1\}\big)\Big)\, \mathbb{1}(A_k^c) \big| \, |\mathbf{X}_0| > \epsilon\, x_n\big] \\
&= \frac{\epsilon^{-\alpha}}{n} \sum_{j=k-1}^n \mathbb{E}\big[f\big((\mathbf{0}^{n-j}, \underline{x_n^{-1}\mathbf{X}_{[0,k-1]}}_\epsilon, \mathbf{0}^{j-k})\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,k-1]}}_\epsilon\|_p > 1\big) \\
&\quad -f\big((\mathbf{0}^{n-j+1}, \underline{x_n^{-1}\mathbf{X}_{[1,k-1]}}_\epsilon, \mathbf{0}^{j-k})\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,k-1]}}_\epsilon\|_p > 1\big) \,\big|\, |\mathbf{X}_0| > \epsilon\, x_n\big]\,.
\end{aligned}
$$

Next we apply shift-invariance and regular variation in $\tilde{\ell}^p$:

$$
\begin{aligned}
&\sim \epsilon^{-\alpha}\mathbb{E}\big[f\big(\underline{x_n^{-1}\mathbf{X}_{[0,k-1]}}_\epsilon\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,k-1]}}_\epsilon\|_p > 1\big) \\
&\qquad -f\big(\underline{x_n^{-1}\mathbf{X}_{[1,k-1]}}_\epsilon\big)\, \mathbb{1}\big(\|\underline{x_n^{-1}\mathbf{X}_{[1,k-1]}}_\epsilon\|_p > 1\big) \,\big|\, |\mathbf{X}_0| > \epsilon\, x_n\big] \\
&\to \epsilon^{-\alpha}\mathbb{E}\big[f\big(\underline{\epsilon\, Y\, \Theta_{[0,k-1]}}_\epsilon\big)\, \mathbb{1}\big(\|\underline{\epsilon Y\, \Theta_{[0,k-1]}}_\epsilon\|_p > 1\big) \\
&\qquad -f\big(\underline{\epsilon\, Y \Theta_{[1,k-1]}}_\epsilon\big)\, \mathbb{1}\big(\|\underline{\epsilon \mathbf{Y}_{[1,k-1]}}_\epsilon\|_p > 1\big)\big] \\
&= \mathbb{E}\big[f\big(\underline{\epsilon\, Y\, \Theta_{[0,k-1]}}_1\big)\, \mathbb{1}\big(\|\underline{\epsilon\, \mathbf{Y}_{[0,k-1]}}_1\|_p > 1\big) \\
&\qquad -f\big(\underline{\epsilon\, \mathbf{Y}_{[1,k-1]}}_1\big)\, \mathbb{1}\big(\epsilon\, \|\underline{Y \Theta_{[1,k-1]}}_1\|_p > 1\big)\big] =: J_{k,\epsilon}\,.
\end{aligned}
$$

By Proposition 2.3.2 we have $\|\Theta\|_\alpha < \infty$ a.s., $|\Theta_t| \xrightarrow{\text{a.s.}} 0$ as $|t| \to \infty$, hence $T := \inf_{t \geq 0}\{t : Y\,|\Theta_t| < 1\} < \infty$ a.s. Then by monotone convergence as $k \to \infty$,

$$\begin{aligned}
J_{k,\epsilon} &= \epsilon^{-\alpha}\mathbb{E}\big[\big(f\big(\epsilon\,\underline{Y\Theta_{[0,\infty]}}_1\big)\,\mathbb{1}\big(\|\epsilon\,\underline{Y\,\Theta_{[0,\infty]}}_1\|_p > 1\big) \\
&\qquad -f\big(\epsilon\,\underline{Y\,\Theta_{[1,\infty]}}_1\big)\,\mathbb{1}\big(\|\epsilon\,\underline{Y\,\Theta_{[1,\infty]}}_1\|_p > 1\big)\big)\mathbb{1}(T<k)\big] \\
&\quad +O(\mathbb{P}(T \geq k)) \\
&\to \epsilon^{-\alpha}\mathbb{E}\big[f\big(\epsilon\,\underline{Y\Theta_{[0,\infty]}}_1\big)\,\mathbb{1}\big(\|\epsilon\,\underline{Y\,\Theta_{[0,\infty]}}_1\|_p > 1\big) \\
&\qquad -f\big(\epsilon\,\underline{Y\,\Theta_{[1,\infty]}}_1\big)\,\mathbb{1}\big(\|\epsilon\,\underline{Y\,\Theta_{[1,\infty]}}_1\|_p > 1\big)\big] \\
&= \int_0^\infty \mathbb{E}\big[f\big(\underline{y\,\Theta_{[0,\infty]}}_\epsilon\big)\,\mathbb{1}\big(\|\underline{y\,\Theta_{[0,\infty]}}_\epsilon\|_p > 1\big) \\
&\qquad -f\big(\underline{y\,\Theta_{[1,\infty]}}_\epsilon\big)\,\mathbb{1}\big(\|\underline{y\,\Theta_{[1,\infty]}}_\epsilon\|_p > 1\big)\big]\,d(-y^{-\alpha}) =: J_\epsilon\,.
\end{aligned}$$

In the last step we changed variables, $u = \epsilon y$, and observed that the integrand vanishes for $y < \epsilon$.

Finally, we want to let $\epsilon \downarrow 0$. We start by inter-changing expectation and integral in J_ϵ, and change variables, $u = y\|\Theta_{[0,\infty]}\|_\alpha$, in the first term of the integrand and then proceed similarly for the second term with the convention that it is zero on $\{\|\Theta_{[1,\infty]}\|_\alpha = 0\}$:

$$\begin{aligned}
J_\epsilon &= \mathbb{E}\Big[\int_0^\infty \Big(\|\Theta_{[0,\infty]}\|_\alpha^\alpha\, f\Big(\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big) \\
&\qquad -\|\Theta_{[1,\infty]}\|_\alpha^\alpha\, f\Big(\underline{\frac{y\,\Theta_{[1,\infty]}}{\|\Theta_{[1,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\,\Theta_{[1,\infty]}}{\|\Theta_{[1,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big)\Big)\,d(-y^{-\alpha})\Big] \\
&= \sum_{t=1}^\infty \int_0^\infty \mathbb{E}\Big[|\Theta_t|^\alpha\Big(f\Big(\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big) \\
&\qquad -f\Big(\underline{\frac{y\,\Theta_{[1,\infty]}}{\|\Theta_{[1,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\,\Theta_{[1,\infty]}}{\|\Theta_{[1,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big)\Big)\Big]d(-y^{-\alpha}) \\
&\qquad +\mathbb{E}\Big[\int_0^\infty f\left(\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\right)\,\mathbb{1}\left(\Big\|\underline{\frac{y\,\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\right)d(-y^{-\alpha})\Big]\,.
\end{aligned}$$

Next we apply the time-change formula (2.3.5) to each summand.

$$\begin{aligned}
J_\epsilon &= \sum_{t=1}^\infty \int_0^\infty \mathbb{E}\Big[f\Big(\underline{\frac{y\Theta_{[-t,\infty]}}{\|\Theta_{[-t,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\Theta_{[-t,\infty]}}{\|\Theta_{[-t,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big) \\
&\qquad -f\Big(\underline{\frac{y\Theta_{[1-t,\infty]}}{\|\Theta_{[1-t,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\Theta_{[1-t,\infty]}}{\|\Theta_{[1-t,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big)\Big]d(-y^{-\alpha}) \\
&\qquad +\mathbb{E}\Big[\int_0^\infty f\Big(\underline{\frac{y\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\Theta_{[0,\infty]}}{\|\Theta_{[0,\infty]}\|_\alpha}}_\epsilon\Big\|_p > 1\Big)d(-y^{-\alpha})\Big]\,.
\end{aligned}$$

This is a telescoping sum in t with value

$$J_\epsilon = \mathbb{E}\Big[\int_0^\infty f\Big(\underline{\frac{y\Theta}{\|\Theta\|_\alpha}}_\epsilon\Big)\,\mathbb{1}\Big(\Big\|\underline{\frac{y\Theta}{\|\Theta\|_\alpha}}_\epsilon\Big\|_p > 1\Big)d(-y^{-\alpha})\Big]\,.$$

By monotone convergence we have

$$\lim_{\epsilon\downarrow 0} J_\epsilon = \mathbb{E}\Big[\int_0^\infty f\big(y\frac{\Theta}{\|\Theta\|_\alpha}\big)\,\mathbb{1}\Big(\big\|y\frac{\Theta}{\|\Theta\|_\alpha}\big\|_p > 1\Big)d(-y^{-\alpha})\Big]\,. \tag{2.7.7}$$

Combining the arguments above, we proved (2.7.6) as desired. □

2.7.2 Proof of Lemma 2.5.4

The case $p<\alpha$. Choose some $\epsilon>0$, $\delta\in(0,1)$. We have the following bounds via truncation

$$\begin{aligned} I_1 - I_2 \;&:=\; \mathbb{P}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p - \mathbb{E}\big[\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p\big] > 1+\delta^p\big) \\ &\quad -\mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p - \mathbb{E}\big[\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p\big] \le -\delta^p\big) \\ &\le\; \mathbb{P}\big(\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p^p - \mathbb{E}\big[\|x_n^{-1}\mathbf{X}_{[0,n]}\|_p^p\big] > 1\big) \\ &\le\; \mathbb{P}\big(\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p - \mathbb{E}\big[\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p\big] > 1-\delta^p\big) \\ &\quad +\mathbb{P}\big(\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p - \mathbb{E}\big[\|\overline{x_n^{-1}\mathbf{X}_{[0,n]}}^\epsilon\|_p^p\big] > \delta^p\big) =: I_3 + I_4\,. \end{aligned}$$

Taking into account $\mathbf{CS}_p$ for $p<\alpha$, we have

$$\lim_{\epsilon\downarrow 0}\limsup_{n\to\infty}(I_2+I_4)/(n\,\mathbb{P}(|\mathbf{X}_0|>x_n)) = 0.$$

Moreover, we observe that by Karamata's theorem for $p<\alpha$

$$\begin{aligned} \mathbb{E}\big[\|\underline{x_n^{-1}\mathbf{X}_{[0,n]}}_\epsilon\|_p^p\big] \;&=\; n\mathbb{E}\big[|\mathbf{X}_0/x_n|^p\,\mathbb{1}(|\mathbf{X}_0|>\epsilon x_n)\big] \\ &=\; O\big(n\,\mathbb{P}(|\mathbf{X}_0|>\epsilon x_n)\big) = o(1)\,. \end{aligned}$$

Thus centering in I_1 and I_3 is not needed, and one can follow the lines of the proof of Lemma 2.7.1 to conclude.

The case $p=\alpha$. It requires only slight changes; we omit details. □

2.7.3 Proofs of the results of Section 2.3

Proof of Proposition 2.3.1

The representation (2.3.1) follows by identifying $\lim_{\epsilon\downarrow 0} J_\epsilon$ as on the right-hand side of (2.7.7). In particular, taking f as the constant map $(\mathbf{x}_t)\mapsto 1$ in (2.3.1) we obtain the representation of the constant $c(p)$ in (2.3.2).

Proof of Proposition 2.3.3

Our goal is first to relate the sequence of spectral components $(\mathbf{Q}^{(p)}(h))_{h\ge 0}$ to (Θ_t). We start with two auxiliary results whose proofs are given at the end of this section.

Lemma 2.7.3. *Let $(\mathbf{X}_t)$ be a stationary time series satisfying $\mathbf{RV}_\alpha$. Then for $h \geq 0$,*

$$\mathbb{P}(\mathbf{Q}^{(p)}(h) \in \cdot) = \frac{1}{c(p,h)} \sum_{k=0}^{h} \mathbb{E}\Big[\frac{\|\Theta_{-k+[0,h]}\|_p^\alpha}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha} \mathbb{1}\Big(\frac{\Theta_{-k+[0,h]}}{\|\Theta_{-k+[0,h]}\|_p} \in \cdot\Big)\Big]\,, \tag{2.7.8}$$

where $c(p,h) := \sum_{k=0}^h \mathbb{E}[\|\Theta_{-k+[0,h]}\|_p^\alpha / \|\Theta_{-k+[0,h]}\|_\alpha^\alpha]$. In particular, $c(\alpha,h) = h+1$ and

$$\mathbb{P}(\mathbf{Q}^{(\alpha)}(h) \in \cdot) = \mathbb{P}\big(\Theta_{-U^{(h)}+[0,h]} / \|\Theta_{-U^{(h)}+[0,h]}\|_\alpha \in \cdot\big)\,, \tag{2.7.9}$$

where $U^{(h)}$ is uniformly distributed on $\{0,\dots,h\}$ and independent of Θ.

Lemma 2.7.4. *Assume $|\Theta_t| \to 0$ as $t \to \infty$ and let $f : \tilde{\ell}^\alpha \cap \{\mathbf{x} : \|\mathbf{x}\|_p = 1\} \to (0,\infty)$ be any bounded Lipschitz-continuous function in $(\tilde{\ell}^\alpha, \tilde{d}_\alpha)$. Then, for every $p \geq \alpha$,*

$$\tfrac{c(p,h)}{h+1}\, \mathbb{E}[f(\mathbf{Q}^{(p)}(h))] \to \mathbb{E}[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha f(\Theta/\|\Theta\|_p)]\,, \tag{2.7.10}$$

as $h \to +\infty$

We conclude from (2.7.10) for $f(\mathbf{x}) \equiv 1$ that $\lim_{h\to\infty} c(p,h)/(h+1) = c(p)$. If $0 < c(p) < \infty$,

$$\lim_{h\to\infty} \mathbb{E}[f(\mathbf{Q}^{(p)}(h))] \quad = \quad c(p)^{-1}\, \mathbb{E}[\|\Theta/\|\Theta\|_\alpha\|_p^\alpha f(\Theta/\|\Theta\|_p)]. \tag{2.7.11}$$

Finally, the portmanteau theorem yields $\mathbf{Q}^{(p)}(h) \xrightarrow{d} \mathbf{Q}^{(p)}(\infty)$ in $(\tilde{\ell}^p \cap \{\mathbf{x} : \|\mathbf{x}\|_p = 1\}, \tilde{d}_p)$ where $\mathbf{Q}^{(p)}(\infty)$ is well defined in view of the right-hand side of (2.7.11). This finishes the proof of the proposition. □

Proof of Lemma 2.7.3. If $\|\mathbf{X}_{[0,h]}/x\|_p > 1$ then for sufficiently small $\epsilon > 0$, $\|\mathbf{X}_{[0,h]}/x\|_\infty > \epsilon$. Therefore, on $\{\|\mathbf{X}_{[0,h]}/x\|_p > 1\}$,

$$\sum_{i=0}^{h} |\mathbf{X}_i/x|^\alpha\, \mathbb{1}\big(|\mathbf{X}_i/x| > \epsilon\big) > 0\,.$$

Using stationarity, we obtain

$$\begin{aligned}
&\mathbb{P}(\|\mathbf{X}_{[0,h]}/x\|_p > 1) \\
&= \sum_{i=0}^{h} \mathbb{E}\Big[\frac{|\mathbf{X}_i/x|^\alpha \mathbb{1}\big(|\mathbf{X}_i/x| > \epsilon\big)}{\sum_{t=0}^h |\mathbf{X}_t/x|^\alpha\, \mathbb{1}\big(|\mathbf{X}_t/x| > \epsilon\big)} \mathbb{1}(\|\mathbf{X}_{[0,h]}/x\|_p > 1)\Big] \\
&= \sum_{i=0}^{h} \mathbb{E}\Big[\frac{|\mathbf{X}_0/x|^\alpha \mathbb{1}\big(|\mathbf{X}_0/x| > \epsilon\big)}{\sum_{t=-i}^{h-i} |\mathbf{X}_t/x|^\alpha\, \mathbb{1}\big(|\mathbf{X}_t/x| > \epsilon\big)} \mathbb{1}(\|\mathbf{X}_{[-i,h-i]}/x\|_p > 1)\Big] \\
&= \mathbb{P}(|\mathbf{X}_0| > x\epsilon) \sum_{i=0}^{h} \mathbb{E}\Big[\frac{|\mathbf{X}_0/x|^\alpha \mathbb{1}(\|\mathbf{X}_{[-i,h-i]}/x\|_p > 1)}{\sum_{t=-i}^{h-i} |\mathbf{X}_t/x|^\alpha\, \mathbb{1}\big(|\mathbf{X}_t/x| > \epsilon\big)} \;\Big|\; |\mathbf{X}_0| > x\epsilon\Big]\,.
\end{aligned}$$

Applying the definition (2.2.1) of regular variation and dominated convergence, we obtain as $x \to \infty$,

$$\begin{aligned}\frac{\mathbb{P}(\|\mathbf{X}_{[0,h]}/x\|_p > 1)}{\mathbb{P}(|\mathbf{X}_0| > x)} &\to \epsilon^{-\alpha} \sum_{i=0}^{h} \mathbb{E}\Big[\frac{|\epsilon Y\,\Theta_0|^\alpha \mathbb{1}(\|\epsilon Y\,\Theta_{[-i,h-i]}\|_p > 1)}{\sum_{t=-i}^{h-i} |\epsilon Y\,\Theta_t|^\alpha\, \mathbb{1}\big(|Y\,\Theta_t| > 1\big)}\Big] \\ &= \sum_{i=0}^{h} \int_\epsilon^\infty \mathbb{E}\Big[\frac{\mathbb{1}(y\,\|\Theta_{[-i,h-i]}\|_p > 1)}{\sum_{t=-i}^{h-i} |\Theta_t|^\alpha\, \mathbb{1}\big(y\,|\Theta_t| > \epsilon\big)}\Big]\, d(-y^{-\alpha})\,.\end{aligned}$$

The left-hand side does not depend on ϵ. Therefore, letting $\epsilon \downarrow 0$, we arrive at

$$\begin{aligned}\lim_{x\to\infty} \frac{\mathbb{P}(\|\mathbf{X}_{[0,h]}/x\|_p > 1)}{\mathbb{P}(|\mathbf{X}_0| > x)} &= \sum_{i=0}^{h} \int_0^\infty \mathbb{E}\Big[\frac{\mathbb{1}(y\,\|\Theta_{[-i,h-i]}\|_p > 1)}{\|\Theta_{[-i,h-i]}\|_\alpha^\alpha}\Big]\, d(-y^{-\alpha}) \\ &= \sum_{i=0}^{h} \mathbb{E}\Big[\frac{\|\Theta_{-i+[0,h]}\|_p^\alpha}{\|\Theta_{-i+[0,h]}\|_\alpha^\alpha}\Big] = c(p,h)\,. \end{aligned} \tag{2.7.12}$$

This constant is finite since $\|\Theta_{i+[0,h]}\|_p \le (h+1)\|\Theta_{i+[0,h]}\|_\infty$.

Next we prove (2.7.8). For this reason, let A be a continuity set with respect to the limit law in (2.7.8). An appeal to (2.7.12) yields

$$\begin{aligned}&c(p,h)\,\mathbb{P}(x^{-1}\mathbf{X}_{[0,h]} \in A \mid \|\mathbf{X}_{[0,h]}\|_p > x) \\ &\quad\sim \frac{\mathbb{P}(x^{-1}\mathbf{X}_{[0,h]} \in A\,, \|\mathbf{X}_{[0,h]}\|_p > x)}{\mathbb{P}(|\mathbf{X}_0| > x)} =: I(x)\,.\end{aligned}$$

Proceeding as for the derivation of (2.7.12), we obtain

$$\begin{aligned}I(x) &\sim \int_0^\infty \sum_{i=0}^{h} \mathbb{E}\Big[\frac{\mathbb{1}(y\,\|\Theta_{[-i,h-i]}\|_p > 1)}{\|\Theta_{[-i,h-i]}\|_\alpha^\alpha}\, \mathbb{1}(y\,\Theta_{[-i,h-i]} \in A)\Big]\, d(-y^{-\alpha}) \\ &= \int_1^\infty \sum_{i=0}^{h} \mathbb{E}\Big[\frac{\|\Theta_{[-i,h-i]}\|_p^\alpha}{\|\Theta_{[-i,h-i]}\|_\alpha^\alpha}\, \mathbb{1}\Big(y\frac{\Theta_{[-i,h-i]}}{\|\Theta_{[-i,h-i]}\|_p} \in A\Big)\Big] d(-y^{-\alpha})\,.\end{aligned}$$

In the last step we changed the variable, $u = y\,\|\Theta_{[-i,h-i]}\|_p > 0$ a.s., observing that $\|\Theta_{[-i,h-i]}\|_p \ge |\Theta_0| = 1$. This proves (2.7.8) and the lemma. $\square$

Proof of Lemma 2.7.4. Assume $f : \tilde{\ell}^\alpha \cap \{\mathbf{x} : \|\mathbf{x}\|_p = 1\} \to (0,\infty)$ is any bounded Lipschitz-continuous function in $(\tilde{\ell}^\alpha, \tilde{d}_\alpha)$. By Lemma 2.7.3 we have for all $p \geq \alpha$,

$$\begin{aligned}
&\frac{c(p,h)}{h+1}\mathbb{E}[f(\mathbf{Q}^{(p)}(h))] - c(p)\mathbb{E}[f(\mathbf{Q}^{(p)})] = \\
&= \frac{1}{h+1}\sum_{k=0}^{h}\mathbb{E}\Big[\Big(\frac{\|\Theta_{-k+[0,h]}\|_p^\alpha}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha} - \frac{\|\Theta\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\Big) f(\Theta_{-k+[0,h]}/\|\Theta_{-k+[0,h]}\|_p)\Big] + \\
&\qquad + \mathbb{E}\Big[\frac{\|\Theta\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\Big(f(\Theta_{-k+[0,h]}/\|\Theta_{[-k+[0,h]}\|_p) - f(\Theta/\|\Theta\|_p)\Big)\Big] \\
&=: I + II\,.
\end{aligned}$$

We will prove that I and II vanish as $h \to \infty$. Since $p \geq \alpha$ subadditivity yields for $k \in [0,h]$,

$$\begin{aligned}
\Big|\frac{\|\Theta_{-k+[0,h]}\|_p^\alpha}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha} - \frac{\|\Theta\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\Big| &= \frac{\big|\big\|\|\Theta\|_\alpha\Theta_{-k+[0,h]}\big\|_p^\alpha - \big\|\|\Theta_{-k+[0,h]}\|_\alpha\Theta\big\|_p^\alpha\big|}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha\|\Theta\|_\alpha^\alpha} \\
&\leq \frac{\big|\big\|\|\Theta\|_\alpha\Theta_{-k+[0,h]}\big\|_p^p - \big\|\|\Theta_{-k+[0,h]}\|_\alpha\Theta\big\|_p^p\big|^{\alpha/p}}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha\|\Theta\|_\alpha^\alpha}\,.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\big|\big\|\|\Theta\|_\alpha\Theta_{-k+[0,h]}\big\|_p^p - \big\|\|\Theta_{-k+[0,h]}\|_\alpha\Theta\big\|_p^p\big| \\
&\leq \quad \|\Theta_{-k+[0,h]}\|_\alpha^p\Big(\sum_{t=-\infty}^{-k-1}|\Theta_t|^p + \sum_{t=-k+h+1}^{+\infty}|\Theta|^p\Big) \\
&\qquad + \big|\|\Theta\|_\alpha - \|\Theta_{-k+[0,h]}\|_\alpha\big|^p\sum_{t=-k}^{-k+h}|\Theta_t|^p\,.
\end{aligned}$$

Thus, $|I|$ is bounded from above by

$$\begin{aligned}
&\frac{1}{h+1}\|f\|_\infty\sum_{k=0}^{h}\Big(\mathbb{E}\Big[\frac{\|\Theta_{[-\infty,-k-1]}\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\Big] + \mathbb{E}\Big[\frac{\|\Theta_{[-k+h+1,\infty]}\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\Big] \\
&\qquad + \mathbb{E}\Big[\frac{\|\Theta_{[-\infty,-k-1]}\|_\alpha^\alpha + \|\Theta_{[-k+h+1,+\infty]}\|_\alpha^\alpha}{\|\Theta\|_\alpha^\alpha}\frac{\|\Theta_{-k+[0,h]}\|_p^\alpha}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha}\Big]\Big) \\
&\leq \quad \frac{1}{h+1}\|f\|_\infty\sum_{k=0}^{h}\Big(\mathbb{E}\Big[\frac{\|\Theta_{[-\infty,-(k+1)]}\|_p^\alpha + \|\Theta_{[-\infty,-(k+1)]}\|_\alpha^\alpha}{\|\Theta\|_\alpha^\alpha}\Big] \\
&\qquad + \mathbb{E}\Big[\frac{\|\Theta_{[k+1,+\infty]}\|_p^\alpha + \|\Theta_{[k+1,+\infty]}\|_\alpha^\alpha}{\|\Theta\|_\alpha^\alpha}\Big]\Big)\,.
\end{aligned}$$

Taking the limit as $h \to \infty$, the Cèsaro limit on the right-hand side converges to zero.

We use the Lipschitz-continuity of f to obtain an upper bound of $|II|$:

$$|II| \le \frac{1}{h+1} c \sum_{k=0}^{h} \mathbb{E}\Big[\frac{\|\Theta_{-k+[0,h]}\|_p^\alpha}{\|\Theta_{-k+[0,h]}\|_\alpha^\alpha} \tilde{d}_\alpha\Big(\frac{\Theta_{-k+[0,h]}}{\|\Theta_{-k+[0,h]}\|_p}, \frac{\Theta}{\|\Theta\|_p}\Big)\Big] .$$

Similar arguments as for $|I| \to 0$ show that $|II| \to 0$. ☐

2.7.4 Proofs of the results of Section 2.4

Proof of Theorem 2.4.1

We start with a version of Theorem 2.4.1 for deterministic thresholds (x_b).

Lemma 2.7.5. *Assume the conditions of Theorem 2.4.1. Then for every $g \in \mathcal{G}_+(\tilde{\ell}^p)$,*

$$\frac{1}{k} \sum_{t=1}^{m} g(x_b^{-1} \mathbf{B}_t) \xrightarrow{\mathbb{P}} \int_0^\infty \mathbb{E}\big[g(y\,\mathbf{Q}^{(p)})\big]\, d(-y^{-\alpha})\,, \qquad n \to \infty\,, \tag{2.7.13}$$

holds for sequences $k_n \to \infty$ and $m_n := [n/b_n] \to \infty$ as in $\mathbf{MX}_p$.

Proof. If $\mathbf{MX}_p$ holds for Lipschitz-continuous $f \in \mathcal{G}_+(\tilde{\ell}^p)$, then it holds for functions $g \in \mathcal{G}_+(\tilde{\ell}^p)$ of the form $g(\mathbf{x}_t) = \mathbb{1}(\mathbf{x}_t \in A)$ where A is a continuity-set of $\tilde{\ell}^p$ and $\mathbf{0} \notin \overline{A}$. It suffices to prove that

$$\big(\mathbb{E}\big[\mathrm{e}^{-\frac{1}{k}\sum_{t=1}^{\lfloor m/k \rfloor} g(x_b^{-1}\mathbf{B}_t)}\big]\big)^k \quad \to \quad \mathrm{e}^{-\mathbb{E}\big[\int_0^\infty g(y\,\mathbf{Q}^{(p)}) d(-y^{-\alpha})\big]}\,. \tag{2.7.14}$$

By stationarity,

$$\mathbb{E}\big[1 - \mathrm{e}^{-\frac{1}{k}\sum_{t=1}^{\lfloor m/k \rfloor} g(x_b^{-1}\mathbf{B}_t)}\big] = O\big(k^{-2}\, m\, \mathbb{E}[g(x_b^{-1}\mathbf{B}_1)]\big)\,. \tag{2.7.15}$$

Since g vanishes in some neighborhood of the origin there exists $c_g > 0$ such that $g(\mathbf{x}) = g(\mathbf{x})\,\mathbb{1}(\|\mathbf{x}\|_p > c_g)$. Therefore and by virtue of Proposition 2.7.2 the right-hand side of (2.7.15) vanishes as $n \to \infty$. Now a Taylor expansion argument shows that the left-hand side of (2.7.14) is of the asymptotic order $\sim \exp\{-(m/k)\mathbb{E}[g(x_b^{-1}\mathbf{B}_t)]\}$, and another application of Proposition 2.7.2 yields (2.7.14). We conclude by the portmanteau theorem for $M_0(\tilde{\ell}^p)$–convergence in Hult and Lindskog [89], Theorem 2.4. that (2.7.13) holds. ☐

We continue with the proof of Theorem 2.4.1. Lemma 2.7.5 implies convergence of the empirical measures in $M_0(\tilde{\ell}^p)$:

$$P_n(\cdot) := \frac{1}{k}\sum_{t=1}^{m} \mathbb{1}(x_b^{-1}\mathbf{B}_t \in \,\cdot\,) \xrightarrow{\mathbb{P}} P(\cdot) := \int_0^\infty \mathbb{P}(y\mathbf{Q}^{(p)} \in \cdot) d(-y^{-\alpha})\,.$$

Using the argument in Resnick [143], p. 81, we may conclude $\|\mathbf{B}\|_{p,(k)}/x_b \xrightarrow{\mathbb{P}} 1$, and thus the joint convergence in $(P_n, \|\mathbf{B}\|_{p,(k)}/x_b) \xrightarrow{\mathbb{P}} (P, 1)$ in $M_0(\tilde{\ell}^p) \times \mathbb{R}_+$ follows. Now (2.4.2) follows by an application of the continuous mapping theorem to the scaling function $s(P(\cdot),t) = P(t\,\cdot)$. To prove continuity of s we use again the portmanteau theorem for $M_0(\tilde{\ell}^p)$–convergence in Hult and Lindskog [89], Theorem 2.4. Thus it suffices to check whether the limit $P_n\,f(\cdot/t) \xrightarrow{\mathbb{P}} P\,f$ holds as $(n,t) \to (\infty,1)$ for Lipschitz-continuous $f \in \mathcal{G}_+(\tilde{\ell}^p)$. But we have with Lemma 2.7.5

$$\begin{aligned}|P_n\,f(\cdot/t) - P\,f| &\leq |P_n\,f(\cdot/t) - P_n\,f| + |P_n\,f - P\,f| \\ &= |P_n\,f(\cdot/t) - P_n\,f| + o_{\mathbb{P}}(1), \quad n \to \infty.\end{aligned}$$

Then, for all $0 < t_0 \leq t < 2$, for $t_0 \leq 1$, setting $g(\mathbf{x}) = (\|\mathbf{x}\|_p \wedge \|f\|_\infty)\,\mathbb{1}(\{\mathbf{x} : \|\mathbf{x}\|_p > c_f/t_0\})$, we have

$$\begin{aligned}|P_n\,f(\cdot/t) - P\,f| &\leq \big|t^{-1} - 1\big|\,P_n\,g + o_{\mathbb{P}}(1) \\ &\leq \big|t^{-1} - 1\big|\,\big(c + o_{\mathbb{P}}(1)\big) + o_{\mathbb{P}}(1)\,,\end{aligned}$$

for some $c > 0$, $c_f > 0$ as above. Letting $t \to 1$, continuity of s follows. □

Proof of Proposition 2.4.4

The result follows by a direct application of Theorem 3.1 in Bartkiewicz *et al.* [5] on $\mathbf{u}^\top \mathbf{S}_n$ for any $\mathbf{u} \in \mathbb{R}^d$ such that $|\mathbf{u}| = 1$ by checking their conditions **(AC)**, **(TB)**. Condition (2.4.8) implies that for all $\delta > 0$,

$$\lim_{l\to\infty} \limsup_{n\to\infty} n \sum_{t=l}^{b_n} \mathbb{P}(|\mathbf{X}_t| > \delta\,a_n\,, |\mathbf{X}_0| > \delta\,a_n)\,,$$

from which **(AC)** is immediate. This condition also implies **(TB)**. We show this in two steps. First, we identify the coefficients $b(d)$ in **(TB)** in terms of the spectral tail process. Mikosch and Wintenberger [121] showed that

$$b_\pm(d) - b_\pm(d-1) = \mathbb{E}\Big[\Big(\sum_{j=0}^{d} \mathbf{u}^\top \Theta_j\Big)_\pm^\alpha\Big] - \mathbb{E}\Big[\Big(\sum_{j=1}^{d} \mathbf{u}^\top \Theta_j\Big)_\pm^\alpha\Big]\,,$$

where we suppress in the notation the dependence of the left-hand side on $\mathbf{u}$ in what follows. **(TB)** amounts to verifying that $b_\pm(d) - b_\pm(d-1)$ converges as $d \to \infty$. For $\alpha \in (0,1)$ this follows by concavity since $\|\Theta\|_\alpha < \infty$ a.s. For $1 < \alpha < 2$ this will follow by a convexity argument if $\mathbb{E}[(\sum_{j\geq 0} |\Theta_j|)^{\alpha-1}] < \infty$. By subadditivity and Jensen's inequality, it is enough to check

$$\sum_{j=0}^{\infty} \big(\mathbb{E}[|\Theta_j|^{\alpha-1}\mathbb{1}(|\Theta_j| > 1)] + \mathbb{E}[|\Theta_j| \wedge 1]\big) < +\infty\,. \tag{2.7.16}$$

We start by showing

$$\sum_{j=0}^{\infty} \mathbb{E}\big[|\Theta_j| \wedge 1\big] < \infty\,. \tag{2.7.17}$$

Condition (2.4.8) implies

$$\lim_{l\to\infty} \limsup_{n\to\infty} n \sum_{j=l}^{b_n} \mathbb{E}\big[(|a_n^{-1}\mathbf{X}_j| \wedge 1)\,\mathbb{1}(|\mathbf{X}_0| > a_n)\big] = 0\,,$$

which yields the following Cauchy criterion: for any $\varepsilon > 0$ there exists K sufficiently large such that for $l \geq K, h \geq 0$,

$$\limsup_{n\to\infty} n \sum_{j=l}^{l+h} \mathbb{E}\big[(|a_n^{-1}\mathbf{X}_j| \wedge 1)\,\mathbb{1}(|\mathbf{X}_0| > a_n)\big] = \sum_{j=l}^{l+h} \mathbb{E}\big[|Y\,\Theta_j| \wedge 1\big] \leq \varepsilon\,,$$

where we used regular variation of $(\mathbf{X}_t)$ in the last step. Then, we conclude (2.7.17) holds. By stationarity we can show similarly

$$\sum_{j=0}^{\infty} \mathbb{E}[|\Theta_{-j}| \wedge 1] < +\infty. \tag{2.7.18}$$

Then, by the time-change formula in (2.3.5) we deduce

$$\begin{aligned} \infty &> \sum_{j=0}^{\infty} \mathbb{E}[|\Theta_{-j}| \wedge 1] = \sum_{j=0}^{\infty} \mathbb{E}[|\Theta_j|^{\alpha}\,(|\Theta_j|^{-1} \wedge 1)] \\ &= \sum_{j=0}^{\infty} \mathbb{E}[|\Theta_j|^{\alpha-1} \wedge |\Theta_j|^{\alpha}] > \sum_{j=0}^{\infty} \mathbb{E}[|\Theta_j|^{\alpha-1}\mathbb{1}(|\Theta_j| > 1)], \end{aligned}$$

and (2.7.16) holds. This finishes the proof of the fact that $\mathbb{E}[(\sum_{j\geq 0} |\Theta_t|)^{\alpha-1}] < +\infty$, in particular $c(1) < \infty$. Applying the mean value theorem and dominated convergence we arrive at the relation

$$b_{\pm}(d) - b_{\pm}(d-1) \quad\to\quad \mathbb{E}\Big[\Big(\sum_{j=0}^{\infty} \mathbf{u}^{\top}\Theta_j\Big)_{\pm}^{\alpha} - \Big(\sum_{j=1}^{\infty} \mathbf{u}^{\top}\Theta_j\Big)_{\pm}^{\alpha}\Big]\,, \qquad d \to \infty\,.$$

Reasoning for the limit as for (2.3.6) and recalling that $c(1) < \infty$, we identify

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{j=0}^{\infty}\mathbf{u}^\top\Theta_j\Big)_{\pm}^{\alpha} - \Big(\sum_{j=1}^{\infty}\mathbf{u}^\top\Theta_j\Big)_{\pm}^{\alpha}\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{j=-\infty}^{\infty}\mathbf{u}^\top\Theta_j\Big)_{\pm}^{\alpha}/\|\Theta\|_\alpha^\alpha\Big] = \mathbb{E}\Big[\Big(\sum_{j=-\infty}^{\infty}\mathbf{u}^\top\mathbf{Q}_j^{(\alpha)}\Big)_{\pm}^{\alpha}\Big] \\
&= c(1)\mathbb{E}\Big[\Big(\sum_{j=-\infty}^{\infty}\mathbf{u}^\top\mathbf{Q}_j^{(1)}\Big)_{\pm}^{\alpha}\Big].
\end{aligned}$$

2.7.5 Proofs of the results of Section 2.6

Proof of Proposition 2.6.1

Notice that ψ_g is bounded and measurable. For $p = \alpha$ we have $\mathbf{Q}^{(\alpha)} \stackrel{d}{=} \Theta/\|\Theta\|_\alpha$. Then the result follows from Proposition 3.6 in Janssen [96]. For $p > 0$, assuming the spectral cluster process $\mathbf{Q}^{(p)}$ is well defined we have $\|\Theta\|_p < \infty$ a.s. and $c(p) < \infty$. Then, we introduce the Radon-Nikodym derivative of $\mathcal{L}(\mathbf{Q}^{(p)})$ with respect to $\mathcal{L}(\Theta/\|\Theta\|_p)$ which by (2.3.1) is the function $h : \ell^p \cap \{\mathbf{x} : \|\mathbf{x}\|_p = 1\} \to \mathbb{R}_{\geq 0}$ defined by $h(\mathbf{y}/\|\mathbf{y}\|_p) := \|\mathbf{y}\|_\alpha/\|\mathbf{y}\|_p$. Finally, the result follows by another application of Proposition 3.6 in Janssen [96]. □

Proof of Theorem 2.6.2

The proof is given for $p = \alpha$ only; the case $p \leq \alpha$ extends in a natural way. Let $g : \ell^\alpha \to \mathbb{R}$ be a continuous bounded function. We start by proving that ψ_g defined in (2.6.1) is a continuous bounded function on $\tilde{\ell^\alpha}$. Fix $\epsilon > 0$ and $[\mathbf{z}] \in \tilde{\ell^\alpha} \cap \{[\mathbf{y}] : \|\mathbf{y}\|_\alpha = 1\}$. Then for all $[\mathbf{x}] \in \tilde{\ell^\alpha} \cap \{[\mathbf{y}] : \|\mathbf{y}\|_\alpha = 1\}$, $k \in \mathbb{Z}$ and $N \in \mathbb{N}$, we have

$$\begin{aligned}
&|\psi_g(\mathbf{x}) - \psi_g(\mathbf{z})| = \Big|\sum_{j\in\mathbb{Z}}|\mathbf{x}_j^*|^\alpha g((\mathbf{x}_{j+t}^*)_t) - \sum_{j\in\mathbb{Z}}|\mathbf{z}_j^*|^\alpha g((\mathbf{z}_{j+t}^*)_t)\Big| \\
= \;& \Big|\sum_{j\in\mathbb{Z}}|\mathbf{x}_j^* - \mathbf{z}_{j-k}^*|^\alpha g((\mathbf{x}_{j+t}^*)_t) - \sum_{j\in\mathbb{Z}}|\mathbf{z}_j^*|^\alpha\big(g((\mathbf{z}_{j+t}^*)_t) - g((\mathbf{x}_{j+t+k}^*)_t)\big)\Big| \\
\leq \;& \|g\|_\infty d_\alpha^\alpha(B^{-k}\mathbf{z}^*, \mathbf{x}^*) + 2\|g\|_\infty\, d_\alpha^\alpha(\mathbf{z}^*, \mathbf{z}_{[-N,N]}^*) \\
&+ \sum_{|j|<N}|\mathbf{z}_j^*|^\alpha\big|g((\mathbf{z}_{j+t}^*)_t) - g((\mathbf{x}_{j+t+k}^*)_t)\big|.
\end{aligned}$$

If $[\mathbf{x}]$ satisfies $\tilde{d}_\alpha^\alpha(\mathbf{z}, \mathbf{x}) < \epsilon(3\|g\|_\infty)^{-1}$ then there exists $k_0 \in \mathbb{Z}$ such that

$$\tilde{d}_\alpha^\alpha(\mathbf{z}, \mathbf{x}) < d_\alpha^\alpha(B^{-k_0}\mathbf{z}^*, \mathbf{x}^*) < \epsilon(3\|g\|_\infty)^{-1}.$$

Furthermore, choose $N_0 \geq 0$ such that $d_\alpha^\alpha(\mathbf{z}^*, \mathbf{z}_{[-N_0,N_0]}^*) < \epsilon(2 \times 3\|g\|_\infty)^{-1}$ and consider the finite set $C_{[\mathbf{z}]} \subset \ell^\alpha \cap \{\mathbf{y} : \|\mathbf{y}\|_\alpha = 1\}$, defined by $C_{[\mathbf{z}]} := \{(\mathbf{z}_{j+t}^*)_t \in \ell^\alpha : |j| < N_0, |\mathbf{z}_j^*| > 0\}$. Notice that for every $\tilde{\mathbf{z}} \in C_{[\mathbf{z}]}$ there exists $\delta(\tilde{\mathbf{z}})$ such that if $d_\alpha^\alpha(\tilde{\mathbf{z}}, \mathbf{x}) < \delta(\tilde{\mathbf{z}})$ implies $|g(\tilde{\mathbf{z}}) - g(\mathbf{x})| < \epsilon/3$. Finally, define $\eta(\mathbf{z}) := \min\{\delta(\tilde{\mathbf{z}}) : \tilde{\mathbf{z}} \in C\} \wedge \epsilon(3\|g\|_\infty)^{-1}$. Then, noticing that $\sum_{|j|<N_0} |\mathbf{z}_j^*|^\alpha \leq \|\mathbf{z}\|_\alpha^\alpha = 1$, we also obtain a bound for the last term. Hence, for every $[\mathbf{x}] \in \tilde{\ell^\alpha}$ satisfying $d_\alpha^\alpha(\mathbf{z}, \mathbf{x}) < \eta(\mathbf{z})$ we have $|\psi_g(\mathbf{x}) - \psi_g(\mathbf{z})| < \epsilon$.

This finishes the proof of the continuity of the function ψ_g on $\tilde{\ell^\alpha} \cap \{\mathbf{y} : \|\mathbf{y}\|_\alpha = 1\}$. We conclude with applications of Lemma 2.7.5 and Proposition 2.6.1. $\square$

Proof of Proposition 2.6.5

Theorem 4.5 in Mikosch and Wintenberger [123] yields immediately

$$\Big|\frac{\mathbb{P}(\sup_{1\leq t\leq n} S_t > x_n)}{n\mathbb{P}(|X_1| > x_n)} - \mathbb{E}\Big[\big(\sup_{t\geq 0} \sum_{i=0}^t \Theta_i\big)_+^\alpha - \big(\sup_{t\geq 1} \sum_{i=1}^t \Theta_i\big)_+^\alpha\Big]\Big| \to 0, \quad n \to \infty\,, \tag{2.7.19}$$

and $n\,\mathbb{P}(|X_1| > x_n) \to 0$. We multiply the function inside the limiting expected value by the constant $1 = \|\Theta\|_\alpha^\alpha / \|\Theta\|_\alpha^\alpha$. Moreover, since $c(1) < \infty$, then $\mathbb{E}[(\sum_{t=1}^\infty |\Theta_t|)^{\alpha-1}] < \infty$; see Lemma 3.11 in Planinić and Soulier [138]. Then, by Fubini's theorem,

$$\begin{aligned}
&\mathbb{E}\Big[\big(\sup_{t\geq 0} \sum_{i=0}^t \Theta_i\big)_+^\alpha - \big(\sup_{t\geq 1} \sum_{i=1}^t \Theta_i\big)_+^\alpha\Big] \\
&= \sum_{j\in\mathbb{Z}} \mathbb{E}\Big[|\Theta_j|^\alpha \Big(\big(\sup_{t\geq 0} \sum_{i=0}^t \frac{\Theta_i}{\|\Theta\|_\alpha}\big)_+^\alpha - \big(\sup_{t\geq 1} \sum_{i=1}^t \frac{\Theta_i}{\|\Theta\|_\alpha}\big)_+^\alpha\Big)\Big]\,.
\end{aligned}$$

At this point we apply the time-change formula for positive measurable functions of Θ at every term of the sum in $j \in \mathbb{Z}$; see Corollary 2.8. in Dombry *et al.* [48]. By the same argument as in the proof of Proposition 2.7.2 we obtain the representation of the expectation in (2.7.19) in terms of the univariate spectral cluster process $Q^{(p)}$.

Now we apply Theorem 2.6.2 to $f(\mathbf{x}) := \lim_{k\to\infty} (\sup_{t\geq -k} \sum_{i=-k}^t x_i)_+^\alpha$ on ℓ^1. It is uniformly continuous and bounded by one on the sphere of ℓ^p, hence (2.6.2) holds for f. Similarly, the constant $c(1)^{-1}$ can be estimated by employing the function $g(\mathbf{x}) := \|\mathbf{x}\|_\alpha$ on ℓ^1 which is bounded by one on the unitary ℓ^1-sphere for $\alpha \geq 1$. $\square$

Proof of Proposition 2.6.7

The re-normalization function ζ is continuous on the unit sphere of (ℓ^α, d_α), except for sequences with $\mathbf{x}_0 = 0$. Then

$$\begin{aligned}
\mathbb{P}(\rho(Y\,\Theta) > 1) &= \mathbb{E}\big[\rho(\Theta_t)^\alpha \wedge 1\big] = \mathbb{E}\Big[\rho(\mathbf{Q}_t^{(\alpha)}/|\mathbf{Q}_0^{(\alpha)}|)^\alpha \wedge 1\Big] \\
&= \mathbb{E}\Big[(\rho^\alpha \wedge 1) \circ \zeta(\mathbf{Q}^{(\alpha)})\Big].
\end{aligned}$$

The proof is finished by an application of Theorem 2.6.2. $\square$

Chapter 3: Some variations on the extremal index

Abstract

We re-consider Leadbetter's extremal index for stationary sequences. It has interpretation as reciprocal of the expected size of an extremal cluster above high thresholds. We focus on heavy-tailed time series, in particular on regularly varying stationary sequences, and discuss recent research in extreme value theory for these models. A regularly varying time series has multivariate regularly varying finite-dimensional distributions. Thanks to results by Basrak and Segers [10] we have explicit representations of the limiting cluster structure of extremes, leading to explicit expressions of the limiting point process of exceedances and the extremal index as a summary measure of extremal clustering. The extremal index appears in various situations which do not seem to be directly related, like the convergence of maxima and point processes. We consider different representations of the extremal index which arise from the considered context. We discuss the theory and apply it to a regularly varying AR(1) process and the solution to an affine stochastic recurrence equation.

keywords[a]: *Extremal index, cluster Poisson process, extremal cluster, regularly varying time series, affine stochastic recurrence equation, autoregressive process*

[a] *Primary 60G70 Secondary 60G55 60F99 60J10 62M10 62G32*

Main contributions

I address (**Quest. 3**) in this Chapter. The following are my main contributions:

- (**Quest. 3**): This chapter reviews the extremal index from the original work in [109, 110], and highlights its relationship with the α-cluster. Mainly, the extremal index summarizes the temporal clustering effect due to successive observations with multiple large values. Recall the spectral tail process (Θ_t) in (1.5.2), and the α-cluster $\mathbf{Q} = \mathbf{Q}^{(\alpha)} \in \ell^\alpha$ introduced in Theorem 2.2.1, which admits the representation

$$\mathbf{Q} \stackrel{d}{=} \Theta/\|\Theta\|_\alpha.$$

  Theorem 3.3.8 reconsiders the asymptotics of the point process $N_n = \sum_{t=1}^n \epsilon_{X_t/a_n}$ in terms of the α-cluster, as $n \to +\infty$, where $n\mathbb{P}(|\mathbf{X}_0| > a_n) \to 1$. Then, a re-interpretation of the extremal index as a summary statistic of the α-cluster follows (see Proposition 3.3.10).

  Section 3.4 otlines state-of-the-art methods for inferring the extremal index and re-introduces the α-cluster-based estimator from extremal ℓ^α-blocks; see (2.4.4). A Monte-carlo study compares it against state-of-the-art estimators. The simulations illustrate Northrop's estimator outperforms in terms of variance. The new estimator is competitive in terms of bias, and improves classical cluster-based blocks and runs estimators in this aspect.

Contents

3.1 Leadbetter's approach to modeling the extremes of a stationary sequence

The paper by Leadbetter [109] and the book of Leadbetter, Lindgren and Rootzén [110] provided a first systematic approach to the extreme value theory of dependent stationary sequences. In particular, Leadbetter introduced mixing and anti-clustering conditions, the conditions D and D', which are tailored for the analysis of dependent extremal events. Moreover, [110] propagated the use of the *extremal index* as a measure for extremal clustering.

The idea of an extremal index originates from [127, 115, 132] who discovered that the maxima

$$M_n = \max_{t=1,\dots,n} X_t\,, \qquad n \geq 1\,,$$

of numerous examples of dependent stationary sequences (X_t) with common distribution F share the property that

$$\mathbb{P}(M_n \leq u_n) \approx \big[\mathbb{P}(X \leq u_n)\big]^{n\,\theta_X} = \big((F(u_n))^n\big)^{\theta_X}\,, \qquad n \to \infty\,,$$

for some number $\theta_X \in [0,1]$ provided (u_n) is a sequence of high thresholds converging sufficiently fast to the right endpoint x_F of F. Leadbetter [109] made this notion precise as the *expected size of an extremal cluster of exceedances above high-level thresholds.* Since $(F(u_n))^n$ is the distribution function of the maximum of n iid random variables with common distribution F at the threshold u_n, the quantity θ_X describes the shrinking effect that the appearance of dependent extremes may have on the distribution of M_n compared to $(F(u_n))^n$.

Leadbetter defined the extremal index θ_X as follows: assume that for every $\tau \in (0,\infty)$ there exists a sequence $(u_n(\tau))$ such that

$$n\,\overline{F}(u_n(\tau)) = n\,(1 - F(u_n(\tau))) \to \tau,$$

and there exists a number θ_X such that

$$\mathbb{P}(M_n \leq u_n(\tau)) \to \mathrm{e}^{-\tau\,\theta_X}\,, \qquad n \to \infty\,.$$

If such a number θ_X exists it belongs to the interval $[0,1]$ and is independent of the choice of the sequences (u_n).

An immediate application is to the convergence in distribution of the sequence (M_n). Assume that (X_t) belongs to the maximum domain of attraction of an extreme value distribution H, i.e., for iid copies $(\widetilde{X}_t)$ of X_1, $\widetilde{M}_n = \max(\widetilde{X}_1, \ldots, \widetilde{X}_n)$, there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that $c_n^{-1}(\widetilde{M}_n - d_n) \xrightarrow{d} \xi$ as $n \to \infty$ and ξ has distribution H. Then if (X_t) has an extremal index θ_X we have

$$n\,\overline{F}(\underbrace{c_n\,x + d_n}_{=:u_n(\tau)}) \to \underbrace{-\log H(x)}_{=:\tau}, \qquad n \to \infty\,, \qquad x \in \mathrm{supp}H\,.$$

and

$$\mathbb{P}\big(c_n^{-1}(M_n - d_n) \le x\big) \to H^{\theta_X}(x)\,, \qquad n \to \infty\,, \qquad x \in \mathrm{supp}H\,.$$

In the case of an iid sequence it is easily seen that $n\,\overline{F}(u_n(\tau)) \to \tau$ holds if and only if $\mathbb{P}(M_n \le u_n(\tau)) \to \mathrm{e}^{-\tau}$. Hence $\theta_X = 1$. The extremal index 1 is not exclusive to iid sequences. Indeed, in the book [110] various examples of strictly stationary sequences are considered for which $\theta_X = 1$. For example, if (X_t) is a Gaussian stationary sequence whose autocovariance function satisfies $\mathrm{cov}(X_0, X_h) = o(1/\log h)$ as $h \to \infty$, then $\theta_X = 1$.

3.2 Sufficient conditions for the existence of the extremal index

The extremal index is often interpreted as *the reciprocal of the expected size of an extremal cluster* for a stationary sequence (X_t). We will give some justification for this interpretation.

3.2.1 The method of block maxima

The key is the definition of an *extremal cluster in the sample* $X_1, \ldots, X_n$: split the sample into $k_n = [n/r_n]$ blocks of equal length r_n:

$$\underbrace{X_1, \ldots, X_{r_n}}_{\text{Block 1}}, \underbrace{X_{r_n+1}, \ldots, X_{2\,r_n}}_{\text{Block 2}}, \ldots, \underbrace{X_{(k_n-1)\,r_n+1}, \ldots, X_{k_n\,r_n}}_{\text{Block } k_n},$$

we ignore the last block of length less than r_n, and we simply call a block an *extremal cluster* relative to a high threshold $u = u_n$ (this means that $u_n \uparrow x_F$ as $n \to \infty$) if there is at least one exceedance of this threshold in this block. For an asymptotic theory it will be important that $r = r_n \to \infty$ such that r_n is small compared to n, i.e., $k_n \to \infty$.

In view of the stationarity of (X_t) the *expected cluster size of a block* is given by

$$\begin{aligned}
\mathbb{E}\Big[\sum_{t=1}^{r_n} \mathbb{1}(X_t > u_n)\,\Big|\, M_{r_n} > u_n\Big] &= \sum_{t=1}^{r_n} \frac{\mathbb{P}(X_t > u_n\,, M_{r_n} > u_n)}{\mathbb{P}(M_{r_n} > u_n)} \\
&= \sum_{t=1}^{r_n} \frac{\mathbb{P}(X_t > u_n)}{\mathbb{P}(M_{r_n} > u_n)} \\
&= \frac{r_n\,\mathbb{P}(X > u_n)}{\mathbb{P}(M_{r_n} > u_n)} =: \frac{1}{\theta_n}\,.
\end{aligned}$$

Obviously, θ_n is a number in $[0,1]$. Under mild regularity conditions the limit $\theta = \lim_{n\to\infty} \theta_n$ exists, assumes values in $(0,1]$ and coincides with Leadbetter's extremal index θ_X; see Theorem 3.2.3 below. For this reason, the extremal index θ_X is often referred to as *the reciprocal of the expected extremal cluster size above high thresholds.*

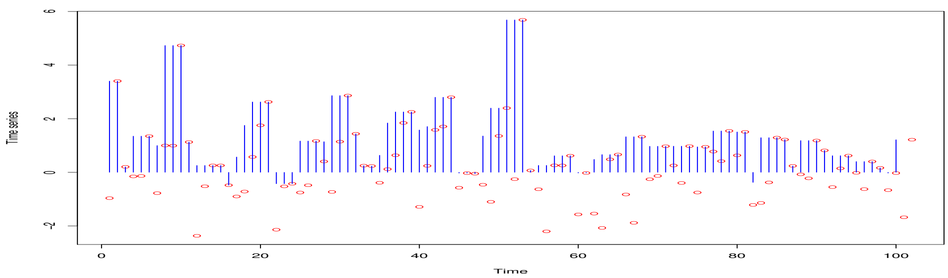


Figure 3.2.1. *Visualization of the max-moving average* $X_t = \max(Z_t, Z_{t+1}, Z_{t+2})$*,* $t = 1, \ldots, 100$*, (blue) for iid student noise* Z_t*,* $t = 1, \ldots, 102$*, with* $\alpha = 4$ *degrees of freedom (red dots). The values of* X_t *typically appear in clusters of size 3. The process* $(|X_t|)$ *has extremal index* $\theta_{|X|} = 1/3$.

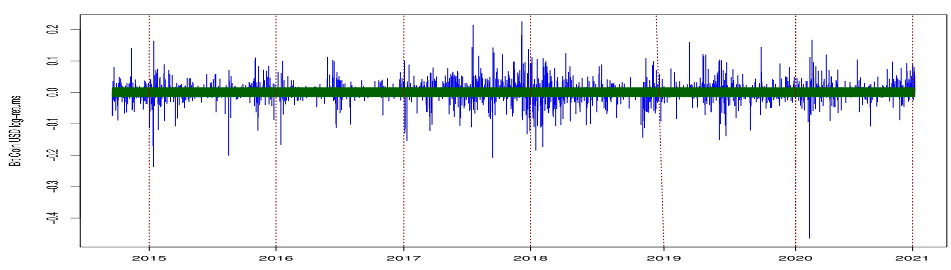


Figure 3.2.2. *The daily log-return series of the Bit Coin USD stock prices from 17 September 2014 until 8 January 2021. We only show the returns below -0.04 or above 0.04 which we interpret as extreme values. These limits roughly correspond to the 10% and 90% quantiles of the data. The extremes typically appear in clusters.*

3.2.2 Approximation of θ_X by θ_n

The following result can be found in slightly different forms in [40], proof of Lemma 2.8, [154, 10].

Theorem 3.2.3. *Consider the following conditions:*

(1) (X_t) *is a real-valued stationary sequence whose marginal distribution* F *does not have an atom at the right endpoint* x_F.

(2) *For a sequence* $u_n \uparrow x_F$ *and an integer sequence* $r = r_n \to \infty$ *such that* $k_n = [n/r_n] \to \infty$ *the following* anti-clustering condition *is satisfied:*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > u_n \mid X_0 > u_n\big) = 0\,. \tag{3.2.1}$$

Here $M_{a,b} = \max_{i=a,\dots,b} X_i$ *for* $a \le b$ *such that* $M_b = M_{a,b}$ *with* $a = 1$.
(3) *A* mixing condition *holds:*

$$\mathbb{P}(M_n \le u_n) - \big(\mathbb{P}(M_{r_n} \le u_n)\big)^{k_n} \to 0\,, \qquad n \to \infty\,, \tag{3.2.2}$$

where (u_n), (k_n) *and* (r_n) *are as in* (2).
(4) *For all positive* τ *there exists a sequence* $(u_n) = (u_n(\tau))$ *such that* $n\,\overline{F}(u_n) \to \tau$ *and* (2), (3) *are satisfied for these sequences* (u_n).

Then the following statements hold:

1. *If* (1) *and* (2) *are satisfied then*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \big|\theta_n - \mathbb{P}\big(M_k \le u_n \mid X_0 > u_n\big)\big| = 0\,, \tag{3.2.3}$$

and $\liminf_{n\to\infty} \theta_n > 0$.
2. *If* (1) *and* (4) *are satisfied and* $\theta = \lim_{n\to\infty} \theta_n$ *exists, then* $\theta_X \in (0,1]$ *exists and coincides with* θ.

Condition (3.2.2) is satisfied for strongly mixing (X_t) with mixing rate (α_h) if one can find integer sequences (ℓ_n) and (r_n) such that $\ell_n/r_n \to 0$, $r_n/n \to 0$ and $k_n \alpha_{\ell_n} \to 0$ as $n \to \infty$. Anti-clustering conditions are common in extreme value theory since Leadbetter introduced the D' condition which is much stronger than (3.2.1) but also easily verified on examples. The goal of such a condition is to avoid that the stationary sequence stays above a high threshold for too long.

Relation (3.2.3) is in agreement with O 'Brien's [133] characterization of the extremal index of (X_t) as the limit

$$\theta_X = \lim_{n\to\infty} \mathbb{P}(M_{\ell_n} \le u_n \mid X_0 > u_n), \tag{3.2.4}$$

for a sequence (ℓ_n) with $\ell_n/n \to 0$, thresholds $u_n \uparrow x_F$ such that $n\,\overline{F}(u_n) \to 1$ as $n \to \infty$, provided a mixing condition holds. O'Brien's condition (3.2.4) has the advantage that it avoids the definition of an extremal cluster.

Remark 3.2.4. *Relation* (3.2.3) *provides a constructive way of calculating* θ_X*: if we know that the limits* $f(k) := \lim_{n\to\infty} \mathbb{P}\big(M_k \le u_n \mid X_0 > u_n\big)$ *exist for every* $k \ge 1$ *then we can try to derive* $\theta_X = \lim_{k\to\infty} f(k)$*. In Section 3.3 we will follow this approach in the case of a* regularly varying sequence.

3.3 Regularly varying sequences

3.3.1 Definition and examples

As a matter of fact, clusters of extremes are more prominent in stationary sequences with heavy-tailed marginal distribution. To illustrate this fact, consider a stationary causal AR(1) process which solves the difference equation $X_t = \varphi\, X_{t-1} + Z_t$, $t \in \mathbb{Z}$, for an iid noise sequence (Z_t). Necessarily, $\varphi \in (-1, 1)$ and, if (Z_t) is iid standard normal then $(|X_t|)$ has extremal index $\theta_{|X|} = 1$ (see [110]), while for iid student noise (Z_t) with α degrees of freedom we have $\theta_{|X|} = 1 - |\varphi|^\alpha$; see Example 3.3.4 below. Thus, the smaller α (the heavier the tail) for given φ the closer $\theta_{|X|}$ to zero.

An AR(1) process with student noise is an example of a *regularly varying time series*. This class of heavy-tailed processes has been studied rather extensively in the last 15 years; see [143] for some basics about multivariate regular variation, and [108] for a recent textbook treatment. This class was considered in full generality first by [40]: they required that the finite-dimensional distributions of the process satisfy a multivariate regular variation condition; see [140, 143] for the definition of this notion. It is an extension of power-law tail behavior from the univariate to the multivariate case defined via the vague convergence of tail measures with infinite limit measures which have the homogeneity property.

Here we will follow an alternative approach by [10] tailored for stationary sequences, avoiding the vague convergence concept. They proved that a real-valued stationary sequence (X_t) is regularly varying with index $\alpha > 0$ in the sense of [40] if and only if there exists a sequence (Θ_t) and a Pareto(α) distributed random variable Y, i.e., $\mathbb{P}(Y > y) = y^{-\alpha}$, $y > 1$, such that (Θ_t) and Y are independent and, for all $h \geq 0$,

$$\mathbb{P}\big(x^{-1}(X_t)_{|t|\leq h} \in \cdot \,\big|\, |X_0| > x\big) \xrightarrow{w} \mathbb{P}\big(Y\,(\Theta_t)_{|t|\leq h} \in \cdot\big)\,, \qquad x \to \infty\,.$$

In the latter relation x can be replaced by any sequence (a_n) such that $n\,\mathbb{P}(|X| > a_n) \to 1$ as $n \to \infty$. Moreover, by definition, $|\Theta_0| = 1$ a.s. The sequence (Θ_t) is the *spectral tail process* of the regularly varying process (X_t); it describes the propagation of a value $|X_0| > x$ for large x through the stationary sequence (X_t) into its past and future.

Example 3.3.1. We consider a stationary AR(1) process given as the causal solution to the difference equation $X_t = \varphi\, X_{t-1} + Z_t$, $t \in \mathbb{Z}$, where (Z_t) is iid regularly varying with index α (e.g. Pareto(α) or student(α)). This means that a generic element Z satisfies $\lim_{x\to\infty} \mathbb{P}(\pm Z > x)/\mathbb{P}(|Z| > x) = p_\pm$ for non-negative values $p_\pm$ such that $p_+ + p_- = 1$, and $\mathbb{P}(|Z| > x) = L(x)x^{-\alpha}$, $x > 0$, for some slowly varying function L. Then a generic element X inherits the regularly varying tail behavior from Z (see [42]):

$$\frac{\mathbb{P}(\pm X > x)}{\mathbb{P}(|Z| > x)} \sim \sum_{j=0}^{\infty} \big[p_\pm\,(\varphi^j)_\pm^\alpha + p_\mp\,(\varphi^j)_\mp^\alpha\big] = \mathbb{P}(\Theta_0 = \pm 1)(1 - |\varphi|^\alpha)\,.$$

But even more is true: (X_t) is a regularly varying time series with spectral tail process

$$\Theta_t = \Theta_Z\,\mathrm{sign}(\varphi^{J+t})\,|\varphi|^t \mathbb{1}(J + t \geq 0) = \Theta_0\,\varphi^t\,\mathbb{1}(J + t \geq 0)\,, \qquad t \in \mathbb{Z}\,, \tag{3.3.1}$$

where $\mathbb{P}(\Theta_Z = \pm 1) = p_\pm$, Θ_Z is independent of J which has distribution

$$\mathbb{P}(J = j) = (1 - |\varphi|^\alpha)\, |\varphi|^{j\,\alpha}\,, \qquad j \geq 0\,.$$

In particular, the forward spectral tail process is given by $\Theta_t = \Theta_0\, \varphi^t$, $t \geq 0$.

Example 3.3.2. We consider the unique causal solution to the affine stochastic recurrence equation $X_t = A_t\, X_{t-1} + B_t$, $t \in \mathbb{Z}$, for an iid sequence $((A_t, B_t))_{t\in\mathbb{Z}}$ with generic element $(A, B) \in \mathbb{R}_+^2$. We assume that the distribution of (A, B) satisfies the conditions of the Kesten-Goldie theory; see [103, 74], cf. [22] for a textbook treatment. The most important condition in this context is the existence of a unique solution $\alpha > 0$ to the equation $\mathbb{E}[A^\alpha] = 1$ which yields the tail index α. Under these conditions for a generic element X, there exists a positive constant c_+ such that

$$\mathbb{P}(X > x) \sim c_+\, x^{-\alpha}\,, \qquad x \to \infty\,.$$

The forward spectral process is then given by

$$(\Theta_t)_{t\geq 0} = (\Pi_t)_{t\geq 0}\,,\ \text{where } \Pi_t = A_1 \cdots A_t\,,$$

while the backward spectral tail process $(\Theta_t)_{t\leq -1}$ has a rather complicated structure.

Writing $S_t = \log \Pi_t = \sum_{i=1}^t \log A_i$, $t \geq 1$, we observe that (S_t) constitutes a random walk with a negative drift. Indeed, by Jensen's inequality we have $\mathbb{E}[\log(A^\alpha)] < \log(\mathbb{E}[A^\alpha]) = 0$.

3.3.2 The extremal index

Following Remark 3.2.4, we will derive the extremal index θ_X of a stationary non-negative regularly varying sequence (X_t) in terms of its spectral tail process. First, we observe that by virtue of the continuous mapping theorem, as $n \to \infty$ for $k \geq 1$,

$$\begin{aligned}
&\mathbb{P}\big(a_n^{-1} M_k \leq 1 \,\big|\, X_0 > a_n\big) \\
&\quad \to \quad \mathbb{P}\big(Y \max_{1\leq t\leq k} \Theta_t \leq 1\big) = \mathbb{P}\big(\max_{1\leq t\leq k} \Theta_t^\alpha \leq Y^{-\alpha}\big) \\
&\quad = \quad \mathbb{E}\big[\big(1 - \max_{1\leq t\leq k} \Theta_t^\alpha\big)_+\big] = \mathbb{E}\big[\max_{0\leq t\leq k} \Theta_t^\alpha - \max_{1\leq t\leq k} \Theta_t^\alpha\big]\,.
\end{aligned}$$

Here we used the fact that $Y^{-\alpha}$ is $U(0,1)$ uniformly distributed and $\Theta_0 = 1$ a.s. Using dominated convergence as $k \to \infty$, we proved under the anti-clustering condition (3.2.1) that

$$\begin{aligned}
\lim_{n\to\infty} \theta_n &= \lim_{k\to\infty} \lim_{n\to\infty} \mathbb{P}\big(a_n^{-1} M_k \leq 1 \big|\, X_0 > a_n\big) \\
&= \lim_{k\to\infty} \mathbb{E}\big[\max_{0\leq t\leq k} \Theta_t^\alpha - \max_{1\leq t\leq k} \Theta_t^\alpha\big] \\
&= \mathbb{E}\big[\big(1 - \max_{t\geq 1} \Theta_t^\alpha\big)_+\big]\,.
\end{aligned}$$

From Theorem 3.2.3 we obtain the following result in [10].

Corollary 3.3.3. *Consider a non-negative stationary regularly varying process (X_t) with index $\alpha > 0$. Then the following statements hold:*

1. *If the anti-clustering condition (3.2.1) holds for $(u_n) = (x\, a_n)$ and some $x > 0$ then the limit $\theta = \lim_{n\to\infty} \theta_n$ exists, is positive and has the representations*

$$\theta \quad = \quad \mathbb{P}\big(Y \sup_{t\geq 1} \Theta_t \leq 1\big) = \mathbb{E}\big[\big(1 - \sup_{t\geq 1} \Theta_t^\alpha\big)_+\big] = \mathbb{E}\big[\sup_{t\geq 0} \Theta_t^\alpha - \sup_{t\geq 1} \Theta_t^\alpha\big]\,. \tag{3.3.2}$$

2. *If (3.2.1) and the mixing condition (3.2.2) hold for $(u_n) = (x\, a_n)$ and all $x > 0$ then the extremal index θ_X exists and coincides with θ.*

The representations of θ given in (3.3.2) only depend on the forward spectral process $(\Theta_t)_{t\geq 0}$. In Proposition 3.3.10 below we provide representations of the extremal index $\theta_{|X|}$ depending on the whole spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$.

Example 3.3.4. We consider the regularly varying AR(1) process from Example 3.3.1. It can be shown to satisfy the anti-clustering and mixing conditions of Theorem 3.2.3. We conclude from Corollary 3.3.3 and the form of the spectral tail process given in (3.3.1) that

$$\theta_{|X|} = \mathbb{E}\big[\big(1 - \max_{t\geq 1} \Theta_t^\alpha\big)_+\big] = 1 - \max_{t\geq 1} |\varphi|^{\alpha\, t} = 1 - |\varphi|^\alpha\,.$$

This formula was already achieved in [42] in the wider context of linear processes.

Example 3.3.5. We consider the regularly varying solution of an affine stochastic recurrence equation under the conditions and with the notation of Example 3.3.2. It can be shown to satisfy the anti-clustering and mixing conditions of Theorem 3.2.3; see [22]. We conclude from this result that (X_t) has extremal index

$$\theta_X = \mathbb{E}\big[\big(1 - \max_{t\geq 1} \Pi_t\big)_+\big] = \mathbb{E}\big[\big(1 - \exp\big(\max_{t\geq 1} S_t\big)\big)_+\big]\,,$$

where $S_t = \sum_{i=1}^t \log A_i$, $t \geq 1$, is a random walk with a negative drift. This value of θ_X was derived in [82]. In that paper a Monte Carlo simulation procedure for the evaluation of θ_X was proposed. Direct calculation of θ_X is difficult; see Example 3.3.12 for an exception.

3.3.3 The extremal index and point process convergence toward a cluster Poisson process

A useful auxiliary result

Lemma 3.3.6. *Consider a non-negative stationary regularly varying sequence (X_t) with index $\alpha > 0$ and assume that (3.2.1) holds for $(u_n) = (x\, a_n)$ and all $x > 0$. Then*

$$\|\Theta\|_\alpha^\alpha := \sum_{j\in\mathbb{Z}} \Theta_j^\alpha < \infty \qquad \text{a.s.}$$

In particular, $\Theta_t \to 0$ a.s. as $|t| \to \infty$, and the time T^* of the largest record of (Θ_t) is finite, i.e., $|T^*|$ is the smallest integer such that

$$\Theta_{T^*} = \max_{t \in \mathbb{Z}} \Theta_t \,.$$

Proof. Write $(Y_t) = Y\,(\Theta_t)$ where the Pareto(α) variable Y and the spectral tail process (Θ_t) are independent. We start by showing

$$Y_t \xrightarrow{\text{a.s.}} 0\,, \qquad t \to \infty\,. \tag{3.3.3}$$

Since (X_t) is regularly varying we have for all $x > 0$ and integers $k \geq 1$,

$$\begin{aligned} \lim_{h\to\infty}\lim_{n\to\infty} \mathbb{P}\big(M_{k,k+h} > x\,a_n \mid X_0 > a_n\big) &= \lim_{h\to\infty} \mathbb{P}\big(\max_{k\le t\le k+h} Y_t > x\big) \\ &= \mathbb{P}\big(\max_{t\ge k} Y_t > x\big)\,. \end{aligned}$$

On the other hand, using the anti-clustering condition (3.2.1) for all $x \in (0,1]$, we have for fixed $k, h \geq 1$,

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}\big(M_{k,k+h} > x\,a_n \mid X_0 > a_n\big) \\ &\quad\leq \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > x\,a_n \mid X_0 > x\,a_n\big)\frac{\mathbb{P}(X > x\,a_n)}{\mathbb{P}(X > a_n)} \\ &\quad= x^{-\alpha} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > x\,a_n \mid X_0 > x\,a_n\big) \\ &\quad= x^{-\alpha}\varepsilon_k\,, \end{aligned}$$

and the right-hand side term ε_k vanishes for large k. Hence, letting $h \to \infty$, we obtain for all $x > 0$,

$$\mathbb{P}\big(\max_{t\ge k} Y_t > x\big) \leq x^{-\alpha}\varepsilon_k\,,$$

and therefore

$$\lim_{k\to\infty} \mathbb{P}\big(\max_{t\ge k} Y_t > x\big) \leq \lim_{k\to\infty} x^{-\alpha}\varepsilon_k = 0\,,$$

implying $\max_{t\geq k} Y_t \xrightarrow{\mathbb{P}} 0$ as $k \to \infty$. Since $(Y_t) = Y\,(\Theta_t)$ a.s. and $Y > 0$ is independent of (Θ_t) this is only possible if $\max_{t\geq k} \Theta_t \xrightarrow{\mathbb{P}} 0$ as $k \to \infty$ but the latter relation is equivalent to $\Theta_t \xrightarrow{\text{a.s.}} 0$ as $t \to \infty$, implying (3.3.3).

Next we show that

$$Y_{-t} \xrightarrow{\text{a.s.}} 0\,, \qquad t \to \infty\,.$$

Since $Y_t \xrightarrow{\text{a.s.}} 0$ as $t \to \infty$ and $Y_0 > 1$ a.s. the following relation holds

$$\mathbb{P}\big(\bigcup_{i\geq 0}\{Y_i \geq 1 > \max_{t>i} Y_t\}\big) = \sum_{i\geq 0}\mathbb{P}\big(Y_i \geq 1 > \max_{t>i} Y_t\big) = 1\,.$$

Suppose that $\mathbb{P}(\sum_{j\leq 0}\mathbb{1}(Y_j > \varepsilon) = \infty) > 0$ for some $\varepsilon > 0$. Then there exists some $i \geq 0$ such that

$$\mathbb{P}\Big(\sum_{j\leq 0}\mathbb{1}(Y_j > \varepsilon) = \infty, Y_i \geq 1 > \max_{t>i} Y_t\Big) > 0\,.$$

We recall the time-change formula from [10]:

$$\mathbb{P}((\Theta_{-h},\dots,\Theta_h) \in \cdot \mid \Theta_{-t} \neq 0) \quad = \quad \mathbb{E}\big[\frac{\Theta_t^\alpha}{\mathbb{E}[\Theta_t^\alpha]}\mathbb{1}\big(\frac{(\Theta_{t-h},\dots,\Theta_{t+h})}{\Theta_t} \in \cdot\big)\big]\,. \tag{3.3.4}$$

In particular, $\mathbb{P}(\Theta_t \neq 0) = \mathbb{E}[\Theta_t^\alpha] = 1$ if and only if for all $h \geq 0$,

$$\mathbb{P}((\Theta_{-h},\dots,\Theta_h) \in \cdot) = \mathbb{E}\big[\frac{\Theta_t^\alpha}{\mathbb{E}[\Theta_t^\alpha]}\,\mathbb{1}\big(\frac{(\Theta_{t-h},\dots,\Theta_{t+h})}{\Theta_t} \in \cdot\big)\big]\,.$$

Therefore

$$\begin{aligned}
\infty &= \mathbb{E}\big[\sum_{j\leq 0}\mathbb{1}(Y_j > \varepsilon)\,\mathbb{1}\big(Y_i \geq 1 > \max_{t>i} Y_t\big)\big] \\
&= \sum_{j\leq 0}\mathbb{P}\big(Y_j > \varepsilon, Y_i \geq 1 > \max_{t>i} Y_t\big) \\
&= \sum_{j\leq 0}\int_1^\infty \mathbb{E}\big[\mathbb{1}\big(y\,\Theta_j > \varepsilon\,, y\,\Theta_i \geq 1 > y\,\max_{t>i}\Theta_t\big)\big]d\big(-y^{-\alpha}\big) \\
&= \sum_{j\leq 0}\int_1^\infty \mathbb{E}\big[\Theta_{-j}^\alpha\,\mathbb{1}\big(y > \varepsilon\,\Theta_{-j}\,, y\,\frac{\Theta_{i-j}}{\Theta_{-j}} \geq 1 > y\,\max_{t>i-j}\frac{\Theta_t}{\Theta_{-j}}\big)\big]d\big(-y^{-\alpha}\big) \\
&\leq \varepsilon^{-\alpha}\sum_{j\leq 0}\mathbb{E}\big[\int_1^\infty \mathbb{1}\big(z > 1\,, z\,\Theta_{i-j} \geq \varepsilon^{-1} > z\,\max_{t>i-j}\Theta_t\big)d\big(-z^{-\alpha}\big)\big] \\
&= \varepsilon^{-\alpha}\sum_{j\leq 0}\mathbb{P}\big(Y_{i-j} \geq \varepsilon^{-1} > \max_{t>i-j} Y_t\big) \\
&= \varepsilon^{-\alpha}\sum_{k\geq i}\mathbb{P}\big(Y_k \geq \varepsilon^{-1} > \max_{t>k} Y_t\big) \\
&\leq \varepsilon^{-\alpha}\,.
\end{aligned}$$

In the last step we used the fact that the events $\{Y_k \geq \varepsilon^{-1} > \max_{t>k} Y_t\}$, $k \geq i$, are disjoint. Thus we got a contradiction. This proves that for all $\varepsilon > 0$ there exist only finitely many $j \leq 0$ such that $Y_j > \varepsilon$, hence $Y_t \xrightarrow{\text{a.s.}} 0$ and also $\Theta_t \xrightarrow{\text{a.s.}} 0$ as $t \to -\infty$, as desired.

In particular, the time T^* of the largest record of the sequence (Θ_t) is finite a.s.

Now suppose that $\mathbb{P}(\sum_{j\in\mathbb{Z}}\Theta_j^\alpha=\infty)>0$. Then there exists an $i\in\mathbb{Z}$ such that

$$\mathbb{P}\big(\sum_{j\in\mathbb{Z}}\Theta_j^\alpha=\infty\,,T^*=i\big)>0\,,$$

and an application of the time-change formula (3.3.4) yields

$$\begin{aligned}\infty &= \mathbb{E}\Big[\sum_{j\in\mathbb{Z}}\Theta_j^\alpha\mathbb{1}(T^*=i)\Big]=\sum_{j\in\mathbb{Z}}\mathbb{E}\big[\Theta_j^\alpha\mathbb{1}(T^*=i)\big]\\ &= \sum_{j\in\mathbb{Z}}\mathbb{P}(T^*=i-j)=1\,,\end{aligned}$$

leading to a contradiction. Thus $\sum_{j\in\mathbb{Z}}\Theta_j^\alpha<\infty$ a.s. This proves the lemma. □

Point process convergence toward cluster Poisson processes

The following point process result was proved in [40] and re-proved in [10] by using the terminology of the spectral tail process.

We adapt the mixing condition in [40] tailored for point process convergence. It is expressed in terms of the Laplace functionals of point processes. Recall that a point process N with state space $E=\mathbb{R}_0=\mathbb{R}\backslash\{0\}$ has Laplace functional

$$\Psi_N(g)=\mathbb{E}\left[\exp\left(-\int_E g\,dN\right)\right]\qquad\text{for } g\in\mathbb{C}_K^+,$$

where the set $\mathbb{C}_K^+$ consists of the continuous functions on E with compact support. Since 0 is excluded from E this means that $g\in\mathbb{C}_K^+$ vanishes in some neighborhood of the origin. Moreover, we have the weak convergence of point processes $N_n\xrightarrow{d}N$ on E if and only if $\Psi_{N_n}\to\Psi_N$ pointwise; see [140, 143].

Mixing condition $\mathcal{A}(a_n)$ Consider integer sequences $r_n\to\infty$ and $k_n=[n/r_n]\to\infty$ and the point processes with state space $E=\mathbb{R}_0$,

$$N_n=\sum_{i=1}^n\varepsilon_{a_n^{-1}X_i}\qquad\text{and}\qquad\widetilde{N}_{r_n}=\sum_{i=1}^{r_n}\varepsilon_{a_n^{-1}X_i}\,,\qquad n\geq 1\,,$$

where ε_x denotes Dirac measure at x. The stationary regularly varying sequence (X_t) satisfies $\mathcal{A}(a_n)$ if there exist (r_n) and (k_n) such that

$$\Psi_{N_n}(g)-\big(\Psi_{\widetilde{N}_{r_n}}(g)\big)^{k_n}\to 0\,,\qquad n\to\infty\,,\quad g\in\mathbb{C}_K^+\,.\tag{3.3.5}$$

Remark 3.3.7. *This condition is satisfied for a strongly mixing sequence (X_t) with mixing rate (α_h) if one can find integer sequences (ℓ_n) and (r_n) such that $\ell_n/r_n\to 0$, $r_n/n\to 0$ and $k_n\alpha_{\ell_n}\to 0$. This is a very mild condition indeed. Relation (3.3.5) ensures that, if $N_n\xrightarrow{d}N$ on the state space E, then also $\sum_{i=1}^{k_n}\widetilde{N}_{r_n}^{(i)}\xrightarrow{d}N$ where $(\widetilde{N}_{r_n}^{(i)})_{i=1,\ldots,k_n}$ are iid copies of $\widetilde{N}_{r_n}$. This fact ensures that the* *limit processes considered are infinitely divisible; cf. [101].*

Theorem 3.3.8. *Consider a stationary regularly varying sequence (X_t) with index $\alpha > 0$. We assume the following conditions:*

(1) *The mixing condition $\mathcal{A}(a_n)$ for integer sequences $r_n \to \infty$ such that $k_n = [n/r_n] \to \infty$ as $n \to \infty$.*

(2) *The anti-clustering condition (3.2.2) for the same sequence (r_n).*

Then we have the point process convergence on the state space $\mathbb{R}_0$

$$N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1} X_i} \xrightarrow{d} N = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha} Q_{ij}}\,, \tag{3.3.6}$$

where

- $\sum_{j=-\infty}^{\infty} \varepsilon_{Q_{ij}}$, $i = 1, 2, \ldots,$ *is an iid sequence of point processes with state space $\mathbb{R}$. A generic element $Q = (Q_j)$ of the sequence $Q^{(i)} = (Q_{ij})_{j\in\mathbb{Z}}$, $i = 1, 2, \ldots,$ has the distribution of the* spectral cluster process

$$Q = \Big(\frac{\Theta_t}{\|\Theta\|_\alpha}\Big)_{t\in\mathbb{Z}}\,.$$

- (Γ_i) *are the points of a unit rate homogeneous Poisson process on $(0, \infty)$.*
- (Γ_i) *and $(Q^{(i)})_{i=1,2,\ldots}$ are independent.*

Remark 3.3.9. *In view of Lemma 3.3.6 we know that $\|\Theta\|_\alpha < \infty$ a.s. Hence the spectral cluster process Q is well defined.*

Since the Poisson points $(\Gamma_i^{-1/\alpha})$ and the sequence of iid point processes $\big(\sum_{j\in\mathbb{Z}} \varepsilon_{Q_{ij}}\big)$ are independent it is not difficult to calculate the Laplace functional of the limit process N:

$$\Psi_N(g) = \exp\Big(-\int_0^\infty \mathbb{E}\big[1 - \mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(y\,Q_j)}\big]\, d(-y^{-\alpha})\Big)\,, \qquad g \in \mathbb{C}_K^+\,.$$

Now we apply the change of variables $z = y\,|Q_{T^*}|$ in $\Psi_N(g)$ where

$$|Q_{T^*}| = \frac{|\Theta_{T^*}|}{\|\Theta\|_\alpha} = \frac{\max_{t\in\mathbb{Z}} |\Theta_t|}{\big(\sum_{j\in\mathbb{Z}} |\Theta_j|^\alpha\big)^{1/\alpha}}\,.$$

Then we obtain for $g \in \mathbb{C}_K^+$,

$$\begin{aligned}\Psi_N(g) &= \exp\Big(-\mathbb{E}[|Q_{T^*}|^\alpha] \\ &\quad \times \int_0^\infty \mathbb{E}\Big[\frac{|Q_{T^*}|^\alpha}{\mathbb{E}[|Q_{T^*}|^\alpha]}\big(1 - \mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(z\,Q_j/|Q_{T^*}|)}\big)\Big]\, d(-z^{-\alpha})\Big)\,.\end{aligned}$$

According to Proposition 3.3.10 below, $\theta_{|X|} = \mathbb{E}[|Q_{T^*}|^\alpha]$. Now, changing the measure with the density $|Q_{T^*}|^\alpha/\mathbb{E}[|Q_{T^*}|^\alpha]$ and writing $\widetilde{Q} = (\widetilde{Q}_j)_{j\in\mathbb{Z}}$ for the sequence $Q/|Q_{T^*}|$ under the new measure, we arrive at

$$\Psi_N(g) \quad = \quad \exp\Big(-\int_0^\infty \mathbb{E}\big[\big(1-\mathrm{e}^{-\sum_{j\in\mathbb{Z}} g(z\,\widetilde{Q}_j|)}\big)\big]\, d\big(-(z/\theta_{|X|}^{1/\alpha})^{-\alpha}\big)\Big)\,.$$

However, this alternative expression of the Laplace functional Ψ_N corresponds to another representation of the point process N:

$$N = \sum_{i=1}^{\infty}\sum_{j=-\infty}^{\infty} \varepsilon_{(\Gamma_i/\theta_{|X|})^{-1/\alpha}\widetilde{Q}_{ij}}\,, \tag{3.3.7}$$

where the Poisson points $(\Gamma_i^{-1/\alpha})$ are independent of the sequence $\big(\sum_{j\in\mathbb{Z}}\varepsilon_{\widetilde{Q}_{ij}}\big)$ of iid copies of $\sum_{j\in\mathbb{Z}}\varepsilon_{\widetilde{Q}_j}$.

We observe that $|\widetilde{Q}_j|\le 1$ a.s. and $|\widetilde{Q}_{T^*}|=1$ a.s. The extremal index $\theta_{|X|}$ plays an important role in representation (3.3.7). Each Poisson point $(\Gamma_i/\theta_{|X|})^{-1/\alpha}$ stands for the radius of a circle around the origin, and the points $(\widetilde{Q}_{ij})_{j\in\mathbb{Z}}$ are inside or on this circle. In this sense, each Poisson point $(\Gamma_i/\theta_{|X|})^{-1/\alpha}$ creates an extremal cluster. Therefore we refer to the process N as a *cluster Poisson process.*

Equivalent expressions for the extremal index

Based on the results in the previous subsection we can derive equivalent expressions of $\theta_{|X|}$ in terms of Q_{T^*} and T^*.

Proposition 3.3.10. *Assume the conditions of Theorem 3.3.8. Then the extremal index $\theta_{|X|}$ of $(|X_t|)$ coincides with the following quantities:*

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{P}(Y\,|Q_{T^*}|>1) = \mathbb{P}(T^*=0)\,. \tag{3.3.8}$$

Here Y is a Pareto(α) random variable independent of Q_{T^*} and T^* is the time of the largest record of $(|\Theta_t|)$.

Remark 3.3.11. *We observe that*

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{E}\Big[\frac{\max_{t\in\mathbb{Z}}|\Theta_t|^\alpha}{\sum_{j\in\mathbb{Z}}|\Theta_j|^\alpha}\Big] = \theta_{|X|}\,.$$

Since $\theta_{|X|}=\mathbb{P}(T^*=0)$ the extremal index $\theta_{|X|}$ has the intuitive interpretation as the probability that $(|\Theta_t|)$ assumes its largest value at time zero.

Example 3.3.12. We consider the regularly varying solution of an affine stochastic recurrence equation under the conditions and with the notation of Example 3.3.2. An exception where the extremal index has an explicit solution is the case $\log A_t = N_t - 0.5$ for an iid standard normal sequence (N_t). Then $\mathbb{E}[A_t]=1$ and the theory mentioned in Example 3.3.2 yields regular variation of (X_t) with index 1. Using the expression $\mathbb{P}(T^*=0)$ and applying some random walk theory (such as the results in [29]), one obtains an exact expression for θ_X in terms of the Riemann zeta function ζ; see Example 3.3.13. A first order approximation to this formula is given by

$$\theta_X \approx \frac{1}{2}\exp\Big(\frac{\zeta(0.5)}{\sqrt{2\pi}}\Big) \approx \frac{1}{2}\exp(-0.5826) \approx 0.2792. \tag{3.3.9}$$

Example 3.3.13. Let $B^{(i)} = (B_t)_{t\in\mathbb{R}}$ be iid standard Brownian motions independent of $\Gamma_1 < \Gamma_2 < \cdots$ which are the points of a unit-rate Poisson process on $(0,\infty)$. We consider the stationary max-stable *Brown-Resnick [20] process*

$$X_t = \sup_{i\geq 1}\Gamma_i^{-1}\,\mathrm{e}^{\sqrt{2}\,B_t^{(i)}-|t|}, \qquad t\in\mathbb{R}\,.$$

It has unit Fréchet marginals $\mathbb{P}(X_t \leq x) = \Phi_1(x) = \mathrm{e}^{-x^{-1}}$, $x>0$. Any discretization $X^{(\delta)} = (X_{\delta t})_{t\in\mathbb{Z}}$ for $\delta>0$ is regularly varying with index 1 and spectral tail process $\Theta_t^{(\delta)} = \mathrm{e}^{\sqrt{2}B_{\delta t}-\delta|t|}$, $t\in\mathbb{Z}$. Direct calculation of $-x\log\mathbb{P}(n^{-1}\max_{1\leq t\leq n}X_{\delta t}\leq x)$, $x>0$, yields the extremal index of $X^{(\delta)}$ as the limit

$$\theta_X^{(\delta)} = \lim_{n\to\infty} n^{-1}\mathbb{E}\Big[\sup_{0\leq t\leq n}\mathrm{e}^{\sqrt{2}B_{\delta t}-\delta t}\Big]. \tag{3.3.10}$$

We use the expression $\theta_X^{(\delta)} = \mathbb{P}(T^{*(\delta)}=0)$ where $T^{*(\delta)}$ is the first record time of $(\Theta_t^{(\delta)})_{t\in\mathbb{Z}}$; see (3.3.8). We consider the first ladder height epoch $\tau_+(\delta) = \inf\{t\geq 1 : \sqrt{2}\,B_{\delta t}+\delta t<0\}$. Using the symmetry of the Gaussian distribution, $(\Theta_t^{(\delta)})_{t\geq 1} \overset{d}{=} (1/\Theta_{-t}^{(\delta)})_{t\geq 1}$, we obtain $\theta_X^{(\delta)} = \mathbb{P}(T^{*(\delta)} = 0) = \mathbb{P}(\tau_+(\delta)=\infty)^2$. Combining this with the classical identity $\mathbb{P}(\tau_+(\delta)=\infty) = 1/\mathbb{E}[\tau_-(\delta)]$ for $\tau_-(\delta) = \inf\{t\geq 1 : \sqrt{2}\,B_{\delta t}-\delta t\leq 0\}$, from random walk theory (see [2]) we get

$$\theta_X^{(\delta)} = \Big(\frac{1}{\mathbb{E}[\tau_-(\delta)]}\Big)^2 = \Big(\frac{\mathbb{E}[B_\delta-\delta]}{\mathbb{E}[\sqrt{2}B_{\tau_-(\delta)}-\tau_-(\delta)]}\Big)^2 = \delta^2(\mathbb{E}[\sqrt{2}\,B_{\tau_+(\delta)}+\tau_+(\delta)])^{-2}\,,$$

where we used Wald's lemma and the symmetry of the Gaussian distribution. To be able to apply Theorem 1.1 in [29] we standardize the increments of the random walk $\sqrt{2}B_{\delta t}$ dividing them by $\sqrt{2\delta}$, turning the drift into $\sqrt{\delta/2}$, and we get

$$\mathbb{E}[\sqrt{2}\,B_{\tau_+(\delta)}+\tau_+(\delta)] = \sqrt{\delta}\exp\Big(-\frac{\sqrt{\delta}}{2\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{\zeta(1/2-n)}{n!(2n+1)}\Big(-\frac{\delta}{4}\Big)^n\Big)\,.$$

This implies that

$$\theta_X^{(\delta)} = \delta\exp\Big(\sqrt{\frac{\delta}{\pi}}\sum_{n=0}^{\infty}\frac{\zeta(1/2-n)}{n!(2n+1)}\Big(-\frac{\delta}{4}\Big)^n\Big).$$

We recover the *Pickands constant* of the Brown-Resnick process (see [135]) as the limit $\lim_{\delta\downarrow 0}\delta^{-1}\,\theta_X^{(\delta)}$:

$$\mathcal{H}_X^{(0)} = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\Big[\sup_{0\leq t\leq T}\mathrm{e}^{\sqrt{2}B_t-t}\Big] = 1.$$

Proof of Proposition 3.3.10. Consider the supremum of all points of the limit process N in Theorem 3.3.8:

$$M = \sup_{i\geq 1} \Gamma_i^{-1/\alpha} \sup_{j\in\mathbb{Z}} |Q_{ij}|\,.$$

The sequences (Γ_i) and $(Q^{(i)})$ are independent and $M = \sup_{i\geq 1} \Gamma_i^{-1/\alpha} V_i$ for the iid sequence $V_i := \sup_{j\in\mathbb{Z}} |Q_{ij}|$, $i = 1, 2, \ldots$, whose generic element V has the property $\mathbb{E}[V^\alpha] < \infty$. Indeed, $V \leq 1$ a.s. by construction. The points $(\Gamma_i^{-1/\alpha}, V_i)$ constitute a marked Poisson process $N_{\Gamma,V}$ with state space $E = (0,\infty)\times[0,\infty)$ and mean measure given by $\mu((x,\infty)\times[0,y]) = x^{-\alpha}\, F_V(y)$, $x > 0, y \geq 0$, where F_V is the distribution function of V. For $x > 0$ we consider $B_x = \{(y,v) \in E : y\, v > x\}$. We observe that

$$\mu(B_x) = \int_{v=0}^{\infty}\int_{y=x/v}^{\infty} \alpha\, y^{-\alpha-1}\, F_V(dv) = \int_0^\infty (x/v)^{-\alpha}\, F_V(dv) = x^{-\alpha}\, \mathbb{E}[V^\alpha]\,.$$

Therefore we have for $x > 0$,

$$\begin{aligned}
\mathbb{P}(M \leq x) &= \mathbb{P}\big(\Gamma_i^{-1/\alpha}\, V_i \leq x\,, i \geq 1\big)\\
&= P(N_{\Gamma,V}(B_x) = 0)\\
&= \mathrm{e}^{\,-\mu(B_x)} = \mathrm{e}^{\,-x^{-\alpha}\,\mathbb{E}[V^\alpha]}\,.
\end{aligned}$$

Thus M is a scaled version of the standard Fréchet distribution, $\Phi_\alpha(x) = \mathrm{e}^{\,-x^{-\alpha}}$, $x > 0$:

$$\mathbb{P}(M \leq x) \;=\; \Phi_\alpha^{\mathbb{E}[V^\alpha]}(x)\,, \qquad x > 0\,.$$

On the other hand, Theorem 3.3.8 and an application of the continuous mapping theorem yield as $n \to \infty$,

$$\begin{aligned}
\mathbb{P}\big(a_n^{-1} M_n \leq x\big) &= \mathbb{P}\big(N_n(x,\infty) = 0\big)\\
&\to \mathbb{P}\big(N(x,\infty) = 0\big)\\
&= \mathbb{P}(M \leq x)\,, \qquad x > 0\,.
\end{aligned}$$

In view of the definition of the extremal index of the sequence $(|X_t|)$ we can identify

$$\mathbb{E}[V^\alpha] \;=\; \mathbb{E}\Big[\sup_{j\in\mathbb{Z}} |Q_j|^\alpha\Big] = \mathbb{E}[|Q_{T^*}|^\alpha].$$

as the value $\theta_{|X|}$. This proves the first part of (3.3.8). The identity

$$\mathbb{E}[|Q_{T^*}|^\alpha] = \mathbb{P}(Y\, |Q_{T^*}| > 1) = \mathbb{P}(|Q_{T^*}|^\alpha > Y^{-\alpha}).$$

is immediate since Q and Y are independent, and $Y^{-\alpha}$ is $U(0,1)$ distributed.

Applying the time-change formula (3.3.4), shifting k to zero, we obtain

$$\begin{aligned}\theta_{|X|} &= \mathbb{E}[|Q_{T^*}|^\alpha] \\ &= \sum_{k\in\mathbb{Z}} \mathbb{E}\Big[\frac{|\Theta_k|^\alpha}{\sum_{j\in\mathbb{Z}}|\Theta_j|^\alpha}\,\mathbb{1}(T^*=k)\Big] \\ &= \sum_{k\in\mathbb{Z}} \mathbb{E}\Big[\frac{|\Theta_{-k}|^\alpha}{\sum_{j\in\mathbb{Z}}|\Theta_{j-k}|^\alpha}\,\mathbb{1}(T^*=0)\Big] \\ &= \mathbb{P}(T^*=0)\,.\end{aligned}$$

This proves the last identity in (3.3.8). □

3.4 Estimation of the extremal index - a short review and a new estimator based on the spectral cluster process

First approaches to the estimation of the extremal index are due to [88, 161]. Estimators based on exceedences of a threshold were proposed in [65, 155, 159, 150]. A modern approach to the maxima method was started in [129]; improvements and asymptotic limit theory can be found in [16, 21].

We will consider some standard estimators of θ_X. For the sake of argument we assume that (X_t) is a non-negative stationary process with marginal distribution F, $k_n = n/r_n$ is an integer sequence such that $r_n \to \infty$, $k_n \to \infty$, and (u_n) is a threshold sequence satisfying $u_n \uparrow x_F$.

3.4.1 Blocks estimator

Recall that θ_X has interpretation as the reciprocal of the expected size of extremal clusters. This idea is the basis for inference procedures from the early 1990s (see [160, 43]). Clusters are identified as blocks of length $r = r_n$ with at least one exceedance of a high threshold $u = u_n$. A blocks estimator $\widehat{\theta}^{\rm bl}$ is given by the ratio of the number $K_n(u)$ of such clusters and the total number of exceedences $N_n(u)$:

$$\widehat{\theta}^{\rm bl}_u(r) = \frac{K_n(u)}{N_n(u)} := \frac{\sum_{t=1}^{k_n}\mathbb{1}(M_{(t-1)r+1,t\,r} > u)}{\sum_{t=1}^{n}\mathbb{1}(X_t > u)}. \tag{3.4.11}$$

This method requires the choice of block length r and threshold level u satisfying $r_n\overline{F}(u_n) \to 0$; if $r_n \to \infty$ does not hold at the prescribed rate $\widehat{\theta}^{\rm bl}$ is biased. Estimators using clusters of extreme exceedences were also considered in [88].

A slight modification of the blocks estimator is the *disjoint blocks estimator* of [161]:

$$\widehat{\theta}^{\rm dbl} = \frac{\log(1 - K_n(u)/k_n)}{r\,\log(1 - N_n(u)/n)}.$$

Assuming some weak dependence condition on (X_t), the heuristic idea behind the estimator is the approximations

$$\big(\mathbb{P}(M_r \le u_n)\big)^{k_n} \approx \mathbb{P}(M_n \le u_n) \approx F^{\theta_X\, n}(u_n),$$

for a suitable sequence (u_n). Then, taking logarithms and replacing $\overline{F}(u_n)$ and $\mathbb{P}(M_n > u_n)$ by their empirical estimators $N_n(u)/n$ and $K_n(u)/k_n$, respectively, we obtain

$$\begin{aligned} \theta_X &\approx \frac{\log \mathbb{P}(M_n \le u_n)}{n\,\log F(u_n)} = \frac{\log(1-\mathbb{P}(M_n > u_n))}{n\,\log(1-\overline{F}(u_n))} \\ &\approx \frac{\log(1-K_n(u)/k_n)}{r_n\,\log(1-N_n(u)/n)} = \widehat{\theta}^{\mathrm{dbl}}\,. \end{aligned}$$

Assuming that both $K_n(u)/k_n$ and $N_n(u)/n$ converge to zero, a Taylor expansion of $\log(1+x) = x(1+o(1))$ as $x \to 0$ shows that $\widehat{\theta}^{\mathrm{bl}} \approx \widehat{\theta}^{\mathrm{dbl}}$. [161] showed that $\widehat{\theta}^{\mathrm{dbl}}$ has a smaller asymptotic variance than $\widehat{\theta}^{\mathrm{bl}}$. [150] proposed a sliding blocks version of $\widehat{\theta}^{\mathrm{dbl}}$ with an even smaller asymptotic variance.

$$\widehat{\theta}^{\mathrm{slbl}}(u,r) = \frac{-\log\Big(\frac{1}{n-r+1}\sum_{t=1}^{n-r+1}\mathbb{1}(M_{t,t+r} \le u)\Big)}{N_n(u)/k_n}\,. \tag{3.4.12}$$

3.4.2 Runs and intervals estimator

[161] proposed the alternative *runs estimator*. It is based on the limit relation (3.2.4): the probability $\mathbb{P}(M_{\ell_n} \le u_n \mid X_0 > u_n)$ is replaced by a sample version for some sequence $l = l_n \to \infty$:

$$\widehat{\theta}^{\mathrm{runs}}_u(l) = \frac{1}{N_n(u)}\sum_{i=1}^{n-l}\mathbb{1}(X_i > u_n\,, M_{i+1,i+l} \le u_n)\,. \tag{3.4.13}$$

Clusters are considered distinct if they are separated by at least l observations not exceeding u. In [65] a complete study of the runs estimator and the inter-exceedence times is given. The thresholds (u_n) need to satisfy $r_n\overline{F}(u_n) \to 1$, and $l_n \le r_n$.

Consider the *exceedance times*:

$$S_0(u) = 0\,, \qquad S_i(u) = \min\{t > S_{i-1}(u) : X_t > u_n\}\,, \qquad i \ge 1\,,$$

with *inter-exceedance times* $T_i(u) = S_i(u) - S_{i-1}(u)$, $i \ge 1$. The sequence $(T_i(u))_{i\ge 2}$ constitutes a stationary sequence. If $r_n\,\overline{F}(u_n) \to 1$, [65] noticed that $(n\,T_2(u))$ converges in distribution to a limiting mixture given by $(1-\theta_X)\mathbb{1}_0(x)+\theta_X\,(1-\mathrm{e}^{-\theta_X\,x})$, $x \ge 0$. Calculation yields to the coefficient of variation ν of $T_2(u)$ whose square is given by

$$\nu^2 = \mathrm{var}(T_2(u))/(\mathbb{E}[T_2(u)])^2 = \mathbb{E}[T_2^2(u)]/(\mathbb{E}[T_2(u)])^2 - 1 = 2/\theta_X - 1\,,$$

leading to overdispersion $\nu > 0$ if and only if $\theta_X < 1$. Replacing the moments on the left-hand side by sample versions and adjusting the empirical moments for bias, [65] arrived at the *intervals* *estimator*

$$\widehat{\theta}^{\rm int}(u) \quad = \quad 1 \wedge \frac{2\Big(\sum_{i=2}^{N_n(u)}(T_i(u)-1)\Big)^2}{(N_n(u)-1)\sum_{i=2}^{N_n(u)}(T_i(u)-1)(T_i(u)-2)}\,. \tag{3.4.14}$$

See also [155, 159].

3.4.3 Northrop's estimator

Assume for the moment that (X_i) is iid and F is continuous. Then $F(X)$ is uniform on $(0,1)$. Hence for $r=r_n$ and $x>0$,

$$\begin{aligned}\mathbb{P}\big(-r_n\,\log F(M_r)>x\big) &= \mathbb{P}(F(M_r)\le \mathrm{e}^{-x/r})\\ &= \mathbb{P}(\max_{i=1,\ldots,r_n} F(X_i)\le \mathrm{e}^{-x/r})\\ &= \big(\mathbb{P}(F(X)\le \mathrm{e}^{-x/r})\big)^r = \mathrm{e}^{-x}\,.\end{aligned}$$

For a weakly dependent sequence (X_i) with marginal distribution F, assume the existence of the extremal index for $(F(X_t))$ which, by monotonicity of F, coincides with θ_X:

$$\mathbb{P}\big(-r_n\,\log F(M_r)>x\big)=\mathbb{P}(\max_{i=1,\ldots,r_n}F(X_i)<\mathrm{e}^{-x/r})\to \mathrm{e}^{-\theta_X\,x}\,,\qquad x>0\,.$$

Thus the random variables $(-r_n\,\log F(M_r))$ are asymptotically $\mathrm{Exp}(\theta_X)$ distributed. For iid $\mathrm{Exp}(\theta_X)$ random variables the maximum likelihood estimator of θ_X is given by the reciprocal of the sample mean. These ideas lead to Northrop's estimators [129]. Mimicking the maximum likelihood estimator of iid $\mathrm{Exp}(\theta_X)$ data for a stationary sequence (X_t), one considers the quantities $-r_n\,\log F(M_{t,t+r})$, $t=1,\ldots,n-r_n$, and constructs sliding or disjoint blocks estimators of θ_X:

$$\widehat{\theta}^{\rm Nsl}(r) \quad = \quad \Big(\frac{1}{n-r+1}\sum_{t=1}^{n-r+1}(-r\,\log F_n(M_{t,t+r}))\Big)^{-1}\,, \tag{3.4.15}$$

$$\widehat{\theta}^{\rm Ndbl}(r) \quad = \quad \Big(\frac{1}{[n/r]}\sum_{i=1}^{[n/r]}(-r\,\log F_n(M_{r\,(i-1)+1,r\,i}))\Big)^{-1}\,. \tag{3.4.16}$$

Here F_n is the empirical distribution function of the data. This particular choice of estimator of F depends on the whole sample, hence introduces additional dependence. This fact requires an optimal choice of block length r_n for implementation.

3.4.4 An estimator based on the spectral cluster process

In this subsection we consider a stationary non-negative regularly varying process (X_t) with index $\alpha>0$, spectral tail process (Θ_t) and normalizing sequence (a_n) satisfying $n\,\mathbb{P}(X>a_n)\to 1$. Proposition 3.3.10 yields the alternative representation $\theta_X=\mathbb{E}[Q_{T^*}^{\alpha}]$ where (Q_t) is the spectral cluster process of (X_t). We will construct an estimator based on this identity.

We consider sums and maxima over disjoint blocks of size $r = r_n = o(n)$:

$$S_{i,r}^{(\alpha)} := \sum_{t=(i-1)\,r+1}^{i\,r} X_t^{\alpha}\,, \qquad M_{i,r} = \max_{t=(i-1)r+1,\ldots,i\,r} X_t\,, \qquad i = 1,\ldots,k_n\,.$$

The following limit relation is proved in [26]:

$$\lim_{n\to\infty} \mathbb{E}\big[\, M_{1,r}^{\alpha}/S_{1,r}^{(\alpha)} \mid S_r^{(\alpha)} > a_n^{\alpha}\big] = \mathbb{E}[Q_{T^*}^{\alpha}], \tag{3.4.17}$$

which is based on large deviation results for regularly varying stationary sequences; see for example [26]. Now we build an estimator of θ_X from an empirical version of the left-hand expectation. Define the corresponding estimator by

$$\widehat{\theta}_v^{\mathrm{scp}}(r) := \frac{\sum_{i=1}^{k_n} \frac{M_{i,r}^{\alpha}}{S_{i,r}^{(\alpha)}}\, \mathbb{1}\big(S_{i,r}^{(\alpha)} > v\big)}{\sum_{i=1}^{k_n} \mathbb{1}\big(S_{i,r}^{(\alpha)} > v\big)}\,. \tag{3.4.18}$$

Here we choose $v = S_{(s),r}^{(\alpha)}$, the sth largest among $(S_{i,r}^{(\alpha)})_{i=1,\ldots,k_n}$ for an integer sequence $s = s_n$ such that $s_n = o(k_n)$.

3.5 A Monte-Carlo study of the estimators

We run a short study based on 1 000 simulated processes $(X_t)_{t=1,\ldots,5000}$ for comparing the performances of some of the aforementioned estimators. First, (X_t) is an AR(1) process with parameter $\varphi = 0.2$ and iid student(1) noise, resulting in a regularly varying process with index 1 and $\theta_{|X|} = 0.8$. Second, we consider the regularly varying solution of an affine stochastic recurrence equation with iid $\log A_t \sim N(-0.5, 1)$, $B_t \equiv 1$, and $\theta_X \approx 0.2792$; see (3.3.9).

Figures 3.5.1 and 3.5.2 show boxplots of the simulation study.

- $\widehat{\theta}^{\mathrm{bl}}$ and $\widehat{\theta}^{\mathrm{runs}}$ are functions of the block and run lengths, respectively. u is the largest $[n^{0.6}]$th upper order statistic of the sample.
- $\widehat{\theta}^{\mathrm{slbl}}$ is a function of r. u is the rth upper order statistic of the sample.
- $\widehat{\theta}^{\mathrm{int}}$ is a function of x. u is the $[n/x]$th upper order statistic.
- $\widehat{\theta}^{\mathrm{Nsl}}$, $\widehat{\theta}^{\mathrm{scp}}$ are functions of r.
- For $\widehat{\theta}^{\mathrm{scp}}$ we choose $s = [n^{0.6}/r]$. The tail index α is estimated by the Hill estimator from [83] based on $[n^{0.8}]$ upper order statistics of the sample.

According to the folklore in the literature, Northrop's estimator $\widehat{\theta}^{\mathrm{Nsl}}$ outperforms the classical estimators (runs, blocks); it has smallest variance but it may be difficult to control its bias. Our experience with $\widehat{\theta}^{\mathrm{scp}}$ shows that it performs better than the other estimators as regards the bias, especially when θ_X is small. The intervals estimator $\widehat{\theta}^{\text{int}}$ is preferred by practitioners because the choice of the hyperparameter x is robust with respect to different values of θ_X. This cannot be said about the other estimators with the exception of $\widehat{\theta}^{\text{scp}}$. In our experiments with sample size $n = 5000$, the choices $x = 32$ and $r = 64$ work well for $\widehat{\theta}^{\text{int}}$ and $\widehat{\theta}^{\text{scp}}$, respectively. We did not fine-tune the hyperparameter s in $\widehat{\theta}^{\text{scp}}$ in our experiments.

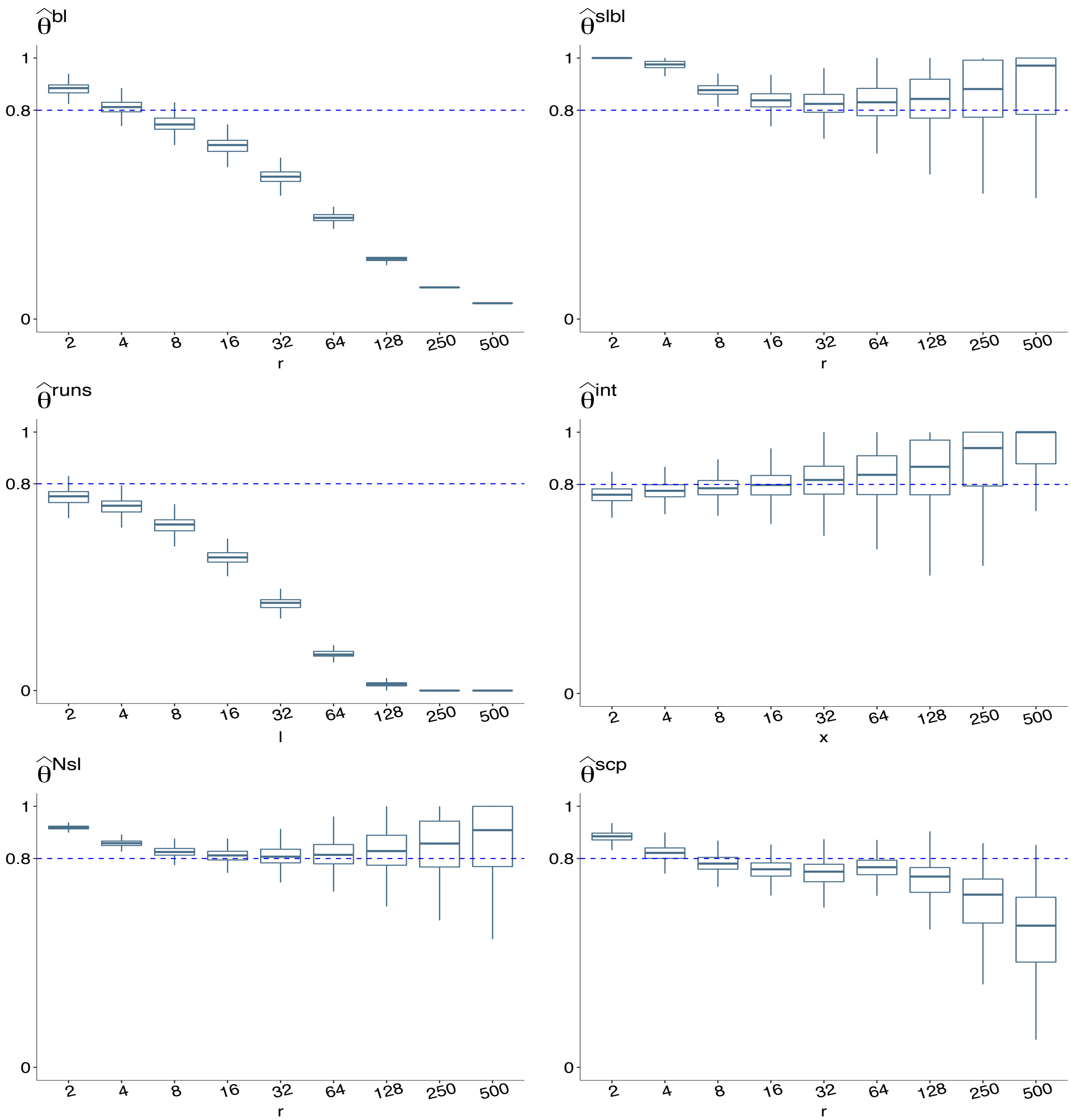


Figure 3.5.1. *Boxplots based on* 1 000 *simulations for the estimation of* $\theta_{|X|} = 0.8$ *in the* $AR(1)$ *model with* $\varphi = 0.2$ *and iid student*(1) *noise.*

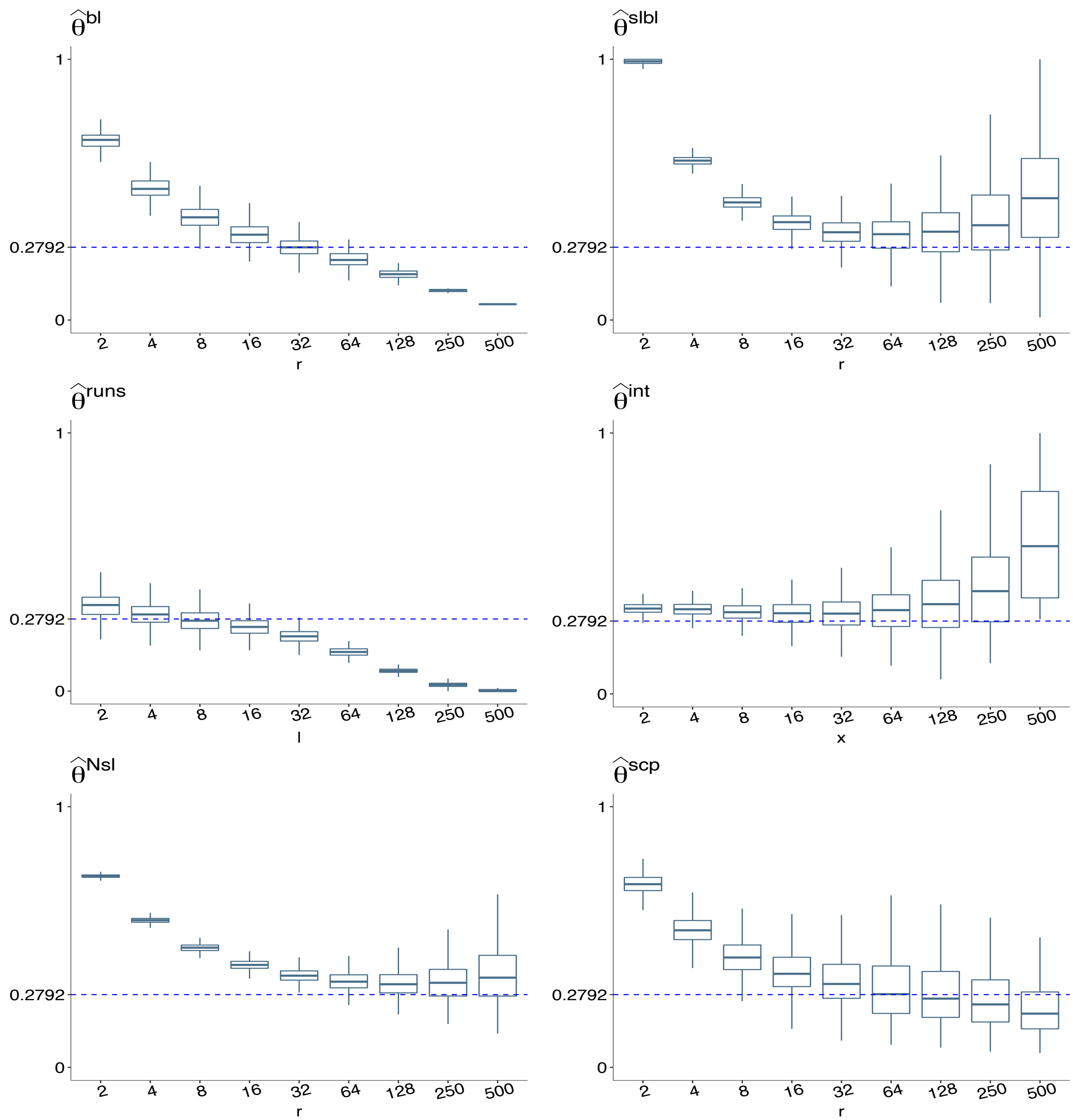


Figure 3.5.2. *Boxplots based on* 1 000 *simulations for the estimation of* $\theta_X \approx 0.2792$ *for the solution to a stochastic recurrence equation.*

Chapter 4: Stable sums to infer high return levels of multivariate rainfall time series

Abstract

We introduce the stable sums method to infer extreme return levels for stationary time series in a multivariate context. We base our method on large deviation principles for regularly varying time series, allowing the incorporation of time and space extreme dependencies in the analysis. First, we avoid classical declustering steps as our strategy is implemented at independent and dependent observations from the stationary model in the same way. Already in the univariate setting, a comparison with the main estimators from extreme value theory, where detecting clustering in time is required, shows improvement of the coverage probabilities of confidence intervals obtained from our approach against its competitors. Also, further numerical experiments point to a smaller mean squared error with the multivariate stable sums method than its component-wise implementation. We apply our method to infer high return levels of daily fall precipitation amounts from a national network of weather stations in France.

keywords: *Environmental time series; multivariate regular variation; stable distribution; stationary time series; cluster process.*

Main contributions

I address **(Quest. 4)** in this Chapter. The following are my main contributions:

- **(Quest. 4)**: This chapter presents the stable sums method for inference of multivariate heavy-tailed return levels. Numerical experiments illustrate the robustness of these estimates with respect to time dependencies.

  Consider a sample $\mathbf{X}_{[1,n]}$ from a heavy-tailed stationary time series $(\mathbf{X}_t)$ with (tail) index $\alpha > 0$. For inference purposes, consider an integer sequence (b_n) such that $n/b_n \to +\infty$, as $n \to +\infty$, and the sub-sample

$$\textstyle\sum_{t=1}^{b_n}|\mathbf{X}_t|^\alpha \ , \ \sum_{t=b_n+1}^{2b_n}|\mathbf{X}_t|^\alpha \ , \qquad , \ \sum_{t=n-b_n+1}^{n}|\mathbf{X}_t|^\alpha. \tag{4.0.1}$$

  Notice the large deviations principles in (2.2.1) yield the approximation

$$\mathbb{P}(X_0(j) > x_{b_n}) \sim m(j)(b_n)^{-1}\mathbb{P}(\textstyle\sum_{t=1}^{b_n}|\mathbf{X}_t|^\alpha \ > x_{b_n}^\alpha), \quad n \to +\infty, \tag{4.0.2}$$

  for a suitable sequence (x_n) satisfying $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$, as $n \to +\infty$, where $X_0(j)$ is the j-th coordinate of $\mathbf{X}_0$, and $m(j) \in (0,1]$ traces the spatial features of $\mathbf{X}_0$ (see (4.1.3)). Since $|\mathbf{X}_1|^\alpha$ is regularly varying of unit (tail) index, the central limit theorem of heavy-tailed increments in Theorem 4.7.1 holds. Then, replacing $m(j)$ by an estimate of this constant, we can model the right-hand side of Equation (4.0.2) using stable distributions.

  Algorithm 1 fits a stable distribution to the sub-sample in (4.0.1), replacing α with an estimate $\widehat{\alpha}^n$, and then extrapolates high quantiles using (4.0.2). To tune the sum length parameter (b_n), Algorithm 1 assesses goodness-of-fit with the stable limit with unit stable parameter.

  Section 4.4 implements this new strategy for multivariate high quantile estimation to the data set of daily rainfall records in France, already introduced in Chapter 1.

Contents

4.1 Introduction

Nowadays, extreme value theory [34] is frequently applied to meteorological time series to capture extremal climatological features in temperatures, winds, precipitation and other atmospheric variables [104, 166, 171]. For example, due to its high societal impacts in terms of flooding, heavy rainfall have been analyzed at various spatial and temporal scales [93]. In particular, storms/fronts duration and spatial coverage can produce potential temporal and spatial dependencies among recordings from nearby weather stations [92]. In this multivariate context, the analysis of consecutive extremes, even in the stationary case, can be theoretically complex [26, 9]. Although marginal behaviors of heavy rainfall is today well modeled, the temporal dynamic is rarely taken in account in applied studies, especially for multivariate time series [61, 170, 1, 38, 64]. To produce accurate high return level estimates from multivariate time series of extreme daily precipitation, we propose an approach to jointly incorporate the temporal dependence and the multidimensional structure among heavy rainfall. This joint modeling appears necessary to perform a full risk assessment, as ignored correlations may lead to erroneous confidence intervals. The latter is particularly important when the practitioner has to provide them about extreme occurrences, i.e. extrapolating beyond the largest observed value.

4.1.1 Motivation

For our case study, we analyze daily precipitation records from a national network in France from 1976 to 2015. To contrast different climate types, we choose three stations in three different regions in France: oceanic in the northwest (Brest, Lanveoc, and Quimper), mediterranean in the south (Hyeres, Bormes-les-Mimosas and Le Luc), and continental in the northeast (Metz, Nancy, and Roville). Concerning seasonality, we will focus on Fall (September, October, and November) as heavy rainfall has been the strongest in France during this season. Concerning marginal behaviors, records within the same region reach similar precipitation intensity levels. For example, the south of France registers higher precipitation amounts than the other two regions, but the south attains high levels at a similar rate; see Figure 4.1. We assume that heavy-tailed margins from the same region are asymptotically equivalent, up to a constant. This is a reasonable modeling assumption if we believe extreme episodes within a region have the same driver, let's say, a big storm.

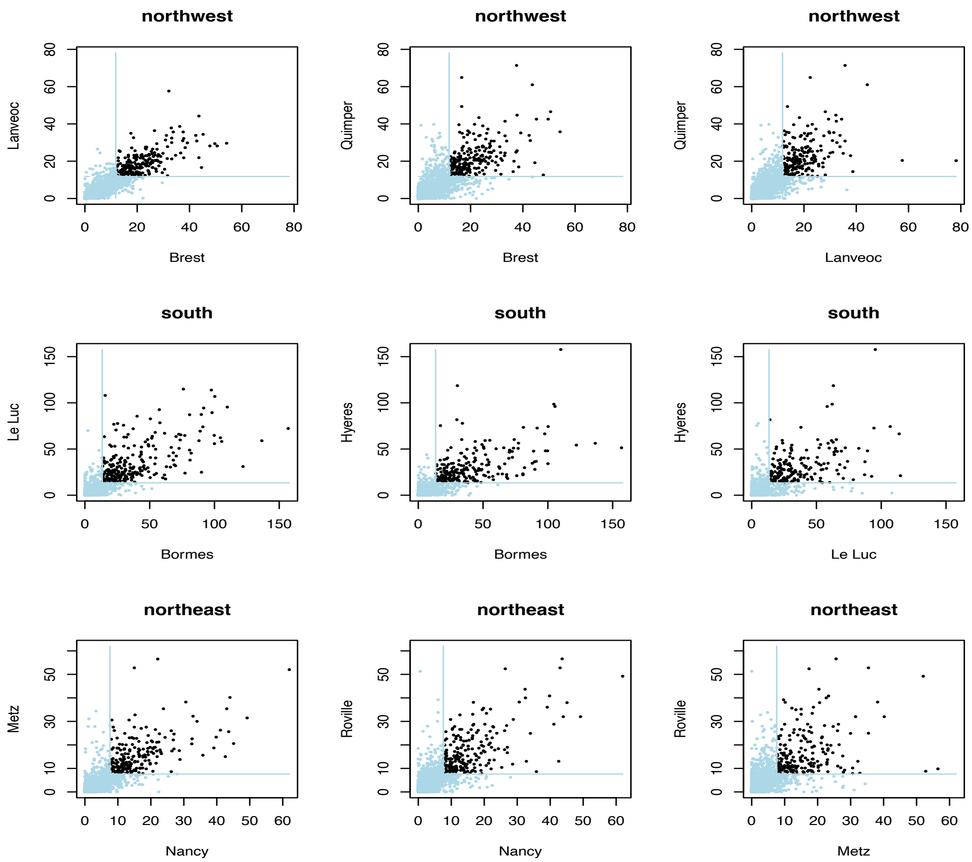


Figure 4.1: Scatter plots of fall daily rainfall in France from 1976 to 2015. The top, middle, and bottom panels refer to three climatological regions: continental (northwest), oceanic (west), and mediterranean (south), respectively. Simultaneous exceedances of the 95-th order statistic of the sample of daily maxima of a region are in black.

While it is reasonable to assume independence between regions, the stations' spatial proximity within a region imposes a tri-variate analysis by region. Figure 4.1 illustrates how high rainfall values often co-occur at two close stations pointing to a spatial dependence of large values. Concerning the temporal ties, we see that at all nine stations, recording high rainfall levels at one day is often followed by measures from rainy days later since an extreme weather condition can last numerous hours. This extremal dependence in time is well illustrated by the temporal extremogram[1] introduced in [44] as can be seen in Figure 4.2. Overall, we can explain the spatial and temporal links by the weather dynamics. As mentioned, It is reasonable to think that the main climatological event impacting an area often has the same source but is manifested at different time lags and locations. Our goal is to improve inference of its extremal features by aggregating all measurements collected of it in space and time. We do so by introducing the stable sums method.

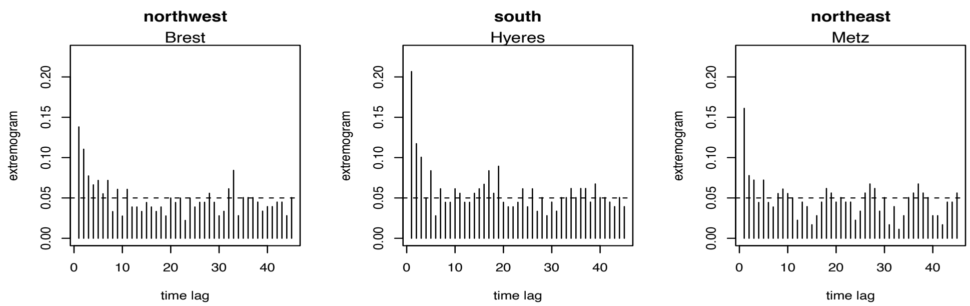


Figure 4.2: Empirical temporal extremogram of the 95-th order statistic of fall daily rainfall in France from 1976 to 2015. The first, middle and last correspond to three climatological regions as in Figure 4.1. As a baseline, the extremogram takes the value pointed by the dotted line on independent time lags.

4.1.2 Asymptotics of the stable sums method

The practical goal of this study is to infer the 50 years return levels at each weather station, while taking in account the tri-variate dependence and the temporal memories. The theoretical added value is that we address the extremal multivariate structure without assuming temporal independence. In particular, we do not need to decluster the 3-dimensional time series to make observations independent in the upper tail. Declustering is particularly challenging in a multivariate context [146]. To bypass these hurdles, we build on a stable sum method. This approach takes its roots in large deviation principles and central limit theory for weakly dependent regularly varying time series [59]. In terms of notations, $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ will always represent a stationary time series with tail index $\alpha > 0$, where $\mathbf{X}_t = (X_t(1), \cdots, X_t(d))$ takes values in $\mathbb{R}^d$, that we endow with a norm $|\cdot|$. To model heavy-tailedness, we assume all vectors $(\mathbf{X}_t)_{|t|=0,\dots,h}$ are multivariate regularly varying, $h \geq 0$; see Equation (4.2.7) for a precise definition. In this setting, under short-rage dependence conditions [10], preventing extreme records to affect indefinitely future ones, the following large deviation approximation holds: for any $p \geq \alpha$, $j = 1, \dots, d$,

$$\mathbb{P}(X_0(j) > x_n) \;\approx\; m(j)\,(n\,c(p))^{-1}\,\mathbb{P}\big(S_{1,n}(p) > x_n^p\big), \quad n \to +\infty, \tag{4.1.3}$$

where $X_0(j)$ is the jth–coordinate of $\mathbf{X}_0$, $S_{1,n}(p) = \sum_{t=1}^n |\mathbf{X}_t|^p$, (x_n) corresponds to a suitable sequence verifying $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ as $n \to +\infty$, $m(j)$ takes values in $[0,1]$, for all $j = 1, \dots, d$,

[1]The temporal extremogram is defined over time lags by $t \mapsto \lim_{x\to+\infty} \mathbb{P}(X_t > x \mid X_0 > x)$.

and $p \mapsto c(p)$ is a decreasing function. We confer to [26] the proof of (4.1.3) and its extension for $\alpha/2 < p < \alpha$.

The practical key aspect of (4.1.3) is that, whenever the constants $m(j)$ and $c(p)$ are adequately estimated, all marginal features of the multivariate vector $\mathbf{X}_0$ can be easily deduced from the single univariate sum $S_{1,n}(p)$. In practice, this means that any extreme quantile of a weather station, say j, can be directly deduced from the sum $S_{1,n}(p)$ computed over the group of three neighbouring stations, albeit the knowledge of the two constants $m(j)$ and $c(p)$ in (4.1.3). To interpret these two quantities, we write them as follows

$$\lim_{n\to+\infty} \frac{\mathbb{P}(X_0(j) > x_n)}{\mathbb{P}(|\mathbf{X}_0| > x_n)} \;=\; m(j), \qquad \lim_{n\to+\infty} \frac{\mathbb{P}((\,S_{1,n}(p)\,)^{1/p} > x_n)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \;=\; c(p). \tag{4.1.4}$$

The ratio between the norm feature $\mathbb{P}(|\mathbf{X}_0| > x_n)$ and the marginal feature $\mathbb{P}(X_0(j) > x_n)$ does not depend on t (as $t = 0$), and consequently, the constants $m(j)$ trace back the d-dimensional structure of extremes, but not the temporal dynamic. In contrast, the constant $c(p)$ captures, throughout the ℓ^p–norm, the temporal clustering among extremes when compared to independent observations. Recall for our case study, the three stations within a region are assumed to have the same tail index and margins within the same region are assumed to be asymptotically equivalent, up to a constant. This is in compliance with the left-hand of (4.1.4). Practically, this is justified by the close proximity among the three stations within each of our region. Theoretically, the multivariate Breiman's theorem [72] tells us that tail equivalences can be obtained whenever a multiplicative or linear lighter-tailed noise impacts the variables at hand.

We recall [26] also showed $c(\alpha) = 1$, regardless of the temporal ties. Thus our goal is to motivate the choice $p = \alpha$ in (4.1.3). This modelling strategy obviously implies that the index of regular variation, α, needs to be estimated, a necessary step in any Pareto based quantile estimation. The main challenge now is to infer the distribution of $S_{1,n}(p)$, for $p > 0$, from the sampled multivariate vector $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. To reach this goal, we consider the following partial sums from disjoint time periods of length b_n as

$$\underbrace{S_{1,b_n}(p)}_{:=\sum_{t=1}^{b_n} |\mathbf{X}_t|^p}\;,\;\underbrace{S_{2,b_n}(p)}_{:=\sum_{t=b_n+1}^{2\,b_n} |\mathbf{X}_t|^p}\;,\quad\cdots\quad,\;\underbrace{S_{\lfloor n/b_n\rfloor,b_n}(p)}_{:=\sum_{t=\lfloor n/b_n\rfloor-b_n+1}^{\lfloor n/b_n\rfloor} |\mathbf{X}_t|^p}\;, \tag{4.1.5}$$

with the convention $S_{b_n}(p) := S_{1,b_n}(p)$. The new sequence of variables $(S_{t,b_n}(p))_{t=1,\ldots,\lfloor n/b_n\rfloor}$ provides us a transformed dataset from which the inference of $x \mapsto \mathbb{P}(S_{1,b_n}(p) > x)$ becomes possible. The natural question is then what is the appropriate model for $S_{b_n}(p)$. As $S_{b_n}(p)$ is a sum of regularly varying increments, then, assuming $n/b_n \to +\infty$ as $n \to +\infty$, the central limit theorem for weakly dependent stationary time series holds. There exists positive and real sequences $(a_n(p)), (d_n(p))$, such that $(S_{b_n}(p) - d_{b_n}(p))/a_{b_n}(p)$ converges to a stable distribution with stable parameter α/p, as the sums length b_n goes to infinity. Two important elements can be highlighted from this convergence. First, the family of α-stable distributions (see Section 4.2.1) appears as the natural parametric family to fit the sequence $(S_{t,b_n}(p))_{t=1,\dots,\lfloor n/b_n \rfloor}$. Second, the aforementioned choice of taking $p = \alpha$ is reinforced as the stable parameter α/p equals to one for this choice. This produces a solid yardstick to select the right b_n. In other words, an appropriate selection of b_n corresponds to the case where the distribution of $(S_{t,b_n}(\alpha))_{t=1,\dots,\lfloor n/b_n \rfloor}$ follows a stable distribution with a stable unit parameter. The algorithm behind this strategy will be explained in Section 4.2.3.

Notice that in our methodology, the choice of p in $c(p)$ is up to the practitioner. The special case of $p = \infty$ has a strong connection with the so-called declustering technique [65]. Typically, the constant $c(\infty)$ equals the *extremal index* of the time series $(|\mathbf{X}_t|)_{t \in \mathbb{Z}}$, which has been understood as the reciprocate of the mean number of consecutive high levels recorded in a short period; cf. [109, 110]. For univariate time series, the conventional choice of taking $p = \infty$, describes block of maxima, which can be modeled with classical extreme value theory based on generalized extreme value distributions [34]. However, it brings the difficult problem of inferring the extremal index. The exceedances approach built on generalized Pareto distributions also needs to be corrected in this setting [30]. Typically, applied studies base inference only on the maxima of clusters [167], but if $c(\infty) < 1$, estimates of marginal features are biased [60, 62, 63]. Instead, choosing $p = \alpha$ completely bypasses the estimation of $c(p)$ or any declustering strategy.

From a theoretical point of view, our method is motivated by equation (4.1.3) proven in [26]. We then use central limit theory to justify the parametric model for the partial sums in (4.1.5). Borrowing classical telescopic sum arguments, we prove the limit with stable parameter one in Section 4.6.2, which interests us as we take $p = \alpha$. This proof uses the *cluster process* defined in [26]. Our proof simplifies the assumptions in [5] and [9] who might have overlooked the unit stable domain, usually receiving less attention. Overall, (4.1.3) justifies inference of extreme quantiles in the scope of the sequence of threshold levels (x_{b_n}). The order of magnitude of the sequence (x_{b_n}) was studied for classical examples as linear processes in [120] and for solutions to recurrence equations in [24, 106]. For further references on large deviation probabilities for weakly dependent processes with no long-range dependence of extremes we refer to [40, 98, 97, 121]. Central limit theory for stationary weakly dependent sequences was first studied in [40] using weak convergence of point processes. Further, [98, 97] show it using classical telescopic sum arguments and large deviation limits. A modern treatment is conferred to [5].

Concerning the implementation of our stable sum method, Section 4.2 details the ingredients of our algorithm and its assumptions. The important step of setting the inputs of our alogrithm is explained in Section 4.2.3. In particular, the estimation of the stable parameter, α, is treated there. It is followed by the description of our algorithm. Our simulation study is described in Section 4.3. Univariate and multivariate cases are investigated. Comparisons with other approaches are implemented and commented. In Section 4.4, the rainfall dataset introduced with Figures 4.2 and 4.1 is analyzed in depth. In section 4.5 we discuss future perspectives. The theoretical aspects of our method are detailed in Section 4.6.

4.2 Stable sums algorithm

4.2.1 Preliminaries

Let $\mathbf{X} = (X(1), \ldots, X(d))$ be a random vector taking values in $\mathbb{R}^d$. For $T > 0$, the multivariate T–return level is $\mathbf{z}_T = (z_T(1), \ldots, z_T(d))$, where $z_T(j)$ is the T–return level associated to the j-th coordinate: $z_T(j) = \inf\{z(j) : \mathbb{P}(X(j) > z(j)) \leq 1/T\}$. We recall below the definition and basic properties of stable distributions.

Definition 4.2.1. *The random variable $\xi_a := \xi_a(\sigma, \beta, \mu)$ follows a stable distribution with parameters (a, σ, β, μ) if and only if, for all $u \in \mathbb{R}$,*

$$\mathbb{E}\big[\exp\{i\, u\, \xi_a\}\big] = \begin{cases} \exp\{-\sigma^a |u|^a (1 - i\, \beta\, sign(u)\, \tan \frac{\pi a}{2}) + i\, \mu\, u\} & \text{if } a \neq 1, \\ \exp\{-\sigma |u| (1 - i\, \beta\, sign(u)\, \frac{2}{\pi} \log |u|) + i\, \mu\, u\} & \text{if } a = 1, \end{cases} \tag{4.2.6}$$

where $a \in (0, 2]$ is the stable parameter, $\sigma \in [0, +\infty)$ is a scale parameter, $\beta \in [-1, 1]$ is a skweness parameter, and $\mu \in \mathbb{R}$ is a location parameter.

Classical examples of stable distributions are the Gaussian distribution with $a = 2$ and $\beta = 0$, the Cauchy distribution with $a = 1$ and $\beta = 0$; and the Lévy distribution with $a = 1/2$ and $\beta = 1$. Stable distributions satisfy the reflection property: if $\xi_a := \xi_a(1, \beta, 0)$ is a stable random variable with parameters $(a, 1, \beta, 0)$, then $-\xi_a$ is a stable random variables with parameters $(a, 1, -\beta, 0)$. The stable distribution is symmetric when $\beta = 0$, and has support in $\mathbb{R}$ when $|\beta| \neq 1$. If $\beta = 1$ there are three cases: if $a < 1$ then the support of its density admits a finite lower bound. If $a = 1$ the density is supported in $\mathbb{R}$ but only the right tail is regularly varying. Otherwise, the stable distribution admits two heavy tails. A full summary on stable distributions can be found in [67, 130, 151].

4.2.2 Model assumptions

In the remaining of the article we assume $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ be a regularly varying time series taking values in $(\mathbb{R}^d, |\cdot|)$, with index of regular variation $\alpha > 0$; cf. [10]. This means there exists an $\mathbb{R}^d$-valued time series $(\Theta_t)_{t \in \mathbb{Z}}$ verifying $|\Theta_0| = 1$ a.s. and

$$\mathbb{P}((\mathbf{X}_t)_{|t|=0,\ldots,h} \in \cdot \mid |\mathbf{X}_0| > x) \xrightarrow{d} \mathbb{P}(Y(\Theta_t)_{|t|=0,\ldots,h} \in \cdot), \quad x \to +\infty, \tag{4.2.7}$$

where Y is (α)–Pareto distributed, $\mathbb{P}(Y > y) = y^{-\alpha}$, for all $y > 1$, independent of $(\Theta_t)_{t \in \mathbb{Z}}$. We fix $|\cdot|$ to be the supremum norm, i.e. $|\mathbf{X}_0| := \max_{j=1,\ldots,d} |X_0(j)|$, but any choice of norm is possible under minor modifications. We call $(\Theta_t)_{t \in \mathbb{Z}}$ the spectral tail process.

For now, we suppose the approximation in (4.1.3) holds and the renormalized process of partial sums $S_{b_n}(p)$ converges to a stable distribution with stable parameter $a = \alpha/p$ as $n \to +\infty$. These assumptions are satisfied for classical examples of weakly dependent regularly varying time series.

We postpone the asymptotic theory behind it to Section 4.6 (see Lemma 4.6.2 and Proposition 4.6.3). Motivated by Proposition 4.6.3, we also set the skweness parameter $\beta = 1$ to simplify computations.

4.2.3 Choice of the algorithm inputs

To construct the time series $(S_{t,b_n}(\alpha))_{t=1,\dots,\lfloor n/b_n \rfloor}$ defined in (4.1.5), we need to determine the sum length, b_n, and α. Also, the indexes of spatial clustering $m(j)$ are required to use (4.1.3).

We estimate the index of regular variation α using the unbiased Hill estimator of [83], see their Equation (4.2) of $\hat{\alpha}_n$ that varies in function of order statistics k. Fixing the choice of k one obtains a point estimate[2] $\widehat{\alpha}^n = \widehat{\alpha}^n(k)$.

To select the temporal window b_n, we recall that the renormalized partial sums, that we denoted $(S_{t,b_n}(p))_{t=1,\dots,\lfloor n/b_n \rfloor}$, should follow, for $p = \alpha$, a stable distribution with stable parameter $a = 1$. So, for a given b_n, we run a ratio likelihood test for the null hypothesis $(H_0) : a = 1$ and we only keep pairs $\widehat{\alpha}^n, b_n$ such that the null hypothesis is not rejected at the 0.05 level. This heuristic allows one to discard an unsuitable choice for the couple $\widehat{\alpha}^n, b_n$.

Concerning the inference of $m(j)$, we recall that (4.1.4) and (4.2.7) imply that $m(j) = \mathbb{P}(Y\Theta_0(j) > 1) = \mathbb{E}[(\Theta_0(j))_+^\alpha]$, where Y is (α)-Pareto distributed, independent of the d–dimensional random variable Θ_0, and $|\Theta_0| = 1$ a.s. For a review on inference of the spectral measure Θ_0, we refer to [26, 39, 50]. In this context, given $\widehat{\alpha}^n$, all $m(j)$ are simply estimated by the following empirical means

$$\widehat{m}^n(j) := \frac{1}{k} \sum_{t=1}^{n} \frac{(X_t(j))_+^{\widehat{\alpha}^n}}{|\mathbf{X}_t|^{\widehat{\alpha}^n}} \mathbb{1}(|\mathbf{X}_t| \geq |\mathbf{X}_{(k)}|), \tag{4.2.8}$$

where $j = 1, \dots, d$, $|\mathbf{X}_{(k)}|$ is the k–th largest order statistic from the norm sample that we fix to be the 95–th empirical quantile for the remaining of this article.

4.2.4 Algorithm

The multivariate T–return level is estimated applying Algorithm 1 to $(\mathbf{X}_t)_{t=1,\dots,n}$. A componentwise estimator is calculated applying Algorithm 1 to $(X_t(j))_{t=1,\dots,n}$, $j = 1, \dots, d$. If $d = 1$ then $m(1) = 1$ and both estimates coincide. Confidence intervals are obtained by sampling parametric bootstrap replicates from a stable distribution with parameters $\widehat{\theta}$, as in line 5 from Algorithm 1. For each replicate, we evaluate a stable quantile at $(1 - (\widehat{m}^n(j)\, T)^{-1})^{b_n}$, for $j = 1, \dots, d$, and return the $1/\widehat{\alpha}^n$–power of the computed quantile. We then use the percentile bootstrap method with a significance level at 0.05 to obtain the confidence intervals. We refer to [54, 55, 56, 57] for large-sample theory of the maximum likelihood estimator for stable distributed sequences. Bounds for the derivatives of the density function in terms of the parameters $(x; a, \sigma, \mu)$ have been computed therein; see also [131] for an overview on maximum likelihood methods for stable distributions.

[2] Equation (4.2) in [83] yields to an estimate $\widehat{\alpha}^n(k)$, where k is a fixed number of higher order statistics. We tune the second order parameter $\widehat{\rho} \leq 0$ to the median value of $k_\rho \mapsto \widehat{\rho}(k_\rho)$, for $2 \leq k_\rho \leq k$; see [76, 83]. We then choose point estimate from a steady portion of the trajectory plot of $k \mapsto \widehat{\alpha}^n(k)$.

Algorithm 1: Stable sums estimator of the multivariate T–return level

Input: $(\mathbf{X}_t)_{t=1,\ldots,n}, b_n, \widehat{\alpha}^n$, $\widehat{m}^n$, see Section 4.2.3;

compute $(S_{t,b_n}(\widehat{\alpha}^n))_{t=1,\ldots,\lfloor n/b_n \rfloor}$ as in (4.1.5) with $p = \widehat{\alpha}^n$,

fit maximum likelihood stable parameters: $\widehat{\theta}$, and $\widehat{\theta}^{a=1}$ fixing $a = 1$,

test the null hypothesis $(H_0) : a = 1$ using ratio likelihood test,;

if (H_0) *is not rejected:* **then**

  $\widehat{\theta} = \widehat{\theta}^{a=1}$,

  if $d > 1$ **then**

    for $j = 1, \ldots, d$ **do**

      calculate $q_T(j)$ a $\widehat{\theta}$–stable quantile at $(1 - (T\,\widehat{m}^n(j)\,)^{-1})^{b_n}$; see (4.1.3),

  else

    calculate $q_T(1)$ a $\widehat{\theta}$–stable quantile at $(1 - 1/T)^{b_n}$.

  return $\widehat{\mathbf{z}}_T^n := \big((q_T(1))^{1/\widehat{\alpha}^n}, \ldots, (q_T(d))^{1/\widehat{\alpha}^n}\big)$,

else

  choose a different pair of parameters $\widehat{\alpha}^n, b_n$.

4.3 Simulation study

4.3.1 Models

We consider the following time series in our numerical experiment.

Burr model: Let $(X_t)_{t=1,\ldots,n}$ be independent random variables distributed as F with

$$F(x; c, \kappa) = 1 - \big(1 + x^c\big)^{-\kappa}, \quad x > 0, \tag{4.3.9}$$

$c, \kappa > 0$ are shape parameters thus X_1 is univariate regularly varying with index $\alpha = \frac{1}{c\kappa} > 0$.

Fréchet model: Let $(X_t)_{t=1,\ldots,n}$ be independent random variables distributed as F with

$$F(x; \alpha) = e^{-x^{-\alpha}}, \quad x > 0, \tag{4.3.10}$$

then X_1 is univariate regularly varying with tail index $\alpha > 0$.

ARMAX model: Let $(X_t)_{t\in\mathbb{Z}}$ satisfy

$$X_t = \max\big\{\lambda\, X_{t-1}, \big(1 - \lambda^\alpha\big)^{1/\alpha}\, Z_t\big\}, \quad t \in \mathbb{Z}, \tag{4.3.11}$$

where $\lambda \in [0, 1)$, and $(Z_t)_{t\in\mathbb{Z}}$ are independent identically distributed Fréchet innovations with tail index of regular variation $\alpha > 0$. Then $(X_t)_{t\in\mathbb{Z}}$ is regularly varying with same index of regular variation but with extremal index equal to $1 - \lambda^\alpha$.

mARMAX$_\tau$ model: Let $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ satisfy

$$X_t(j) := \max\big\{\, \lambda(j) X_{t-1}(j), \big(1 - (\lambda(j))^\alpha\big)^{1/\alpha}\, Z_t(j)\,\big\}, \qquad t \in \mathbb{Z}, \tag{4.3.12}$$

for $j = 1, \ldots, d$, where λ takes values in $[0,1)^d$ and $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ are independent identically distributed vectors from a Gumbel copula with Fréchet marginals and index of regular variation $\alpha > 0$. Moreover, Z_1 is distributed as G defined by

$$G(x;\alpha,\tau) = e^{-\left((x(1))^{-(\alpha/\tau)}+(x(2))^{-(\alpha/\tau)}+\cdots+(x(d))^{-(\alpha/\tau)}\right)^{\tau}}, \tag{4.3.13}$$

for $x \in \mathbb{R}^d$, and $\tau \in [0,1]$ that we refer as the coefficient of spatial dependence. The stationary solution $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ is multivariate regularly varying with index of regular variation $\alpha > 0$; cf. [68] for more details.

Moreover, straightforward computations from (4.3.13) yield

$$m(j) = \lim_{x\to+\infty} \frac{\mathbb{P}(X_0(j) > x)}{\mathbb{P}(|\mathbf{X}_0| > x)} = \lim_{x\to+\infty} \frac{1-e^{-1/x^{\alpha}}}{1-e^{-d^{\tau}/x^{\alpha}}} = \frac{1}{d^{\tau}} < 1, \tag{4.3.14}$$

for all $j = 1, \ldots, d$. Then, from (4.3.14) we recover the symmetric properties of the Gumbel copula as $m(1) = \cdots = m(d) = 1/d^{\tau}$. We can also see from (4.3.14) that the coefficient of spatial dependence $\tau \in [0,1]$ plays a key role while measuring the spatial dependence of extremes. Indeed, similar calculations allow one to compute the spatial dependence parameter between any two marginals, say j, as

$$\lim_{x\to+\infty} \mathbb{P}(X_0(j) > x \,|\, X_0(j') > x) = 2 - 2^{\tau}, \tag{4.3.15}$$

thus $\tau = 1$ points to asymptotic independence of extremes, whereas $\tau = 0$ indicates complete dependence of extremes.

4.3.2 Numerical experiment

We perform a Monte Carlo simulation study for two main purposes. We aim to compare the stable sums method commonly used method based on declustering detailed Section 4.3.4. We also aim to evaluate the multivariate approach compared to the component-wise approach.

We estimate return levels z_T for periods $T = 20, 50, 100$ years corresponding to the 99.95-th, 99.98-th and 99.99-th quantiles. We simulate 1000 trajectories of length $n = 4000$ from the models presented in Section 4.3.1 with parameters:

1. Burr(c,κ) model with $(c,\kappa) = (2,2)$ in (4.3.9).

2. Fréchet(α) model with $\alpha = 4$ in (4.3.10).

3. ARMAX(λ) model with $\alpha = 4$, for both $\lambda = 0.7$ and $\lambda = 0.8$ in (4.3.11).

4. mARMAX$_{\tau}(\lambda)$ model taking values in $[0,+\infty)^3$ with $\alpha = 4$ and $\lambda = (0.7, 0.7, 0.7)$ in (4.3.12), and for $\tau = 0.1, 0.2, \ldots, 0.9$, in (4.3.13).

Notice $\alpha = 4$ in all the models $1-4$. This corresponds to a typical rainfall tail index.

4.3.3 Implementation of stable sums method

We fix the index of regular variation to be $\widehat{\alpha}^n = \widehat{\alpha}^n(k)$ (see Section 4.2.3 for details) with $k = n^{0.7}$ for the Burr model and $k = n^{0.9}$ for the Fréchet, ARMAX and mARMAX models. Now notice that plugging in the estimates $\widehat{\alpha}^n, \widehat{m}^n$ in Algorithm 1 we can run the stable sums method as a function of the sum lengths b_n. In this way, we implement our method for the sum lengths $bl = 2^i$, with $i = 4, 5, 6, 7$. We sample $R = 100$ parametric bootstrap replicates to compute confidence intervals for the estimated return levels. For the multivariate models, we compute both the multivariate and component-wise stable sums estimator.

4.3.4 Implementation of classical methods

For the univariate models, we also run the peaks over threshold and block maxima methods [34]. We model extremal time-dependence using the theory of clusters of exceedances; see [40, 26, 108]. A brief description of both implementation procedures is given below.

The peaks over threshold method models exceeded amounts over a high threshold with a generalized Pareto distribution; see chapters 4 and 5 in [34] for an overview. In our case, we fix the threshold level to be the 95–th empirical quantile. To adjust confidence intervals, we keep only the largest peak from each clusters of exceedances, and fit a Pareto model to the exceeded amounts from this sample. For cluster detection, we follow the ideas in [65]. We use the code in the R-package *extRemes 2.0.12*, and our implementation follows the guide in [73]. We compute delta-method confidence intervals on the declustered sample.

The block maxima method models the largest records form consecutive observations with a generalized extreme value distribution; see chapters 3 and 5 in [34] for an introduction. We implement it over disjoint blocks of length $bl_{BM} = 20$. We estimate the extremal index using the interval's estimator in Ferro et al. [65], tuned with the 95–th empirical quantile. We fit a generalized extreme value distribution, perform the extremal index estimation, and the extrapolation using the R-package *extRemes 2.0.12*; see also the guide [73] for details. We compute delta-method confidence intervals for return levels.

4.3.5 Simulation study in the univariate case

Estimation of the index of regular variation, as detailed in Section 4.3.3, yields unbiased estimates for the univariate models (plots can be available upon request).

We can see from Figure 4.3 that the median estimate of the 50 years return level with the peaks over threshold method underestimates the real value when implemented at the dependent models and this underrates the risk. This bias was already observed in [63, 63], and it avert us from inferring marginal features from the maxima of clusters; see [60]. In comparison, our block maxima implementation is unbiased for all four models. However, it has a larger spread compared to the stable sums methods. Our method gives satisfactory results and, as expected, the choice of the sum length can be seen as a trade-off between bias and variance. We conclude that Algorithm 1 works fine coupled with a good estimate of the index of regular variation as the one detailed in Section 4.2.3 for all models.

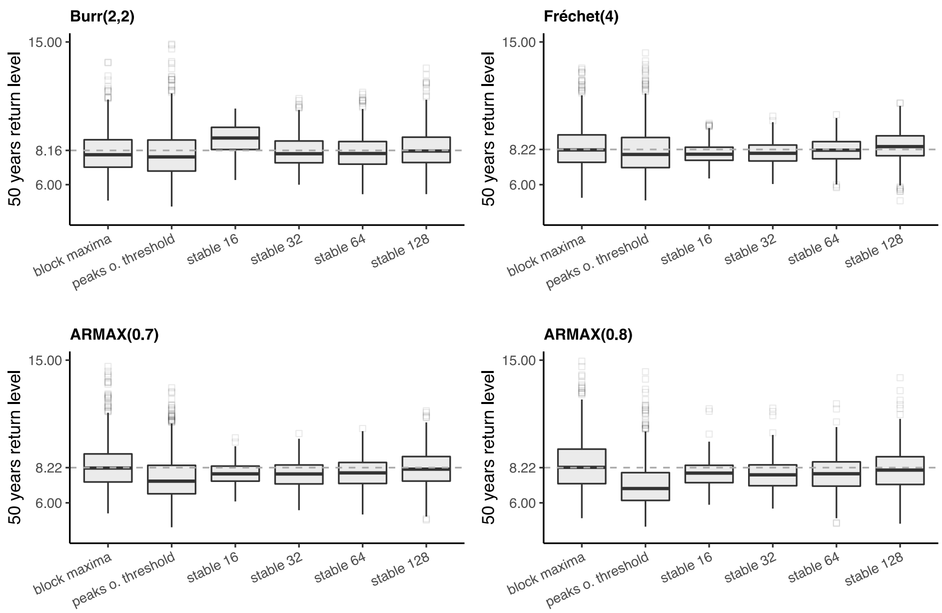


Figure 4.3: Boxplots of estimates $\widehat{z}_T^n$ with different methods such that stable 16 refers to the stable sums method with sum length $b_l = 2^4 = 16$. Dotted lines indicate the true values.

To measure the accuracy of the confidence intervals of all methods, we compute the number of times they capture the correct value. One must keep in mind that for the stable sums method Algorithm 1 only returns an estimate if the test of the stable parameter equal to one is accepted. We summarize the sample coverage probabilities and provide the proportion of acceptance of the ratio test among the 1000 simulated trajectories from each model in Table 4.1. The coverage results are not reliable when the proportion of test acceptance is small, however, it increases as the sum length increases. As a result, we notice from Table 4.1 that we automatically discard the very small sum lengths.

To sum up, we read from Table 4.1 that coverage probabilities are unsatisfactory for the peaks over threshold method, specially for the models with time dependence of extremes. The coverage for the block maxima method is not well calibrated and gives poor results for the Burr model which is the only model with a marginal distribution that does not belong to the family of generalized extreme value distributions. In this manner, we aim to point at the deficiency of the classical methods on small sample sizes. Instead, the stable sums method outperforms the block maxima and peaks over threshold methods for sum lengths between 32 and 64, where acceptance of the ratio likelihood test is significant.

Table 4.1: Table of coverage probabilities. The value in parenthesis the ratio test $(H_0) : a = 1$ acceptance proportion. In bold we highlight the optimal choice of sum length for the stable sums method. In our study, a precise coverage should be at 0.95.

years		20	50	100		20	50	100
		Burr(2,2)				Fréchet(4)		
block maxima		.91	.89	.87		.93	.93	.92
peaks o. threshold		.87	.85	.83		.89	.87	.86
stable 16	(.06)	.89	.85	.80	(.53)	.94	.95	.95
stable 32	(.51)	**.93**	**.94**	**.95**	(.83)	**.96**	**.96**	**.96**
stable 64	(.85)	**.95**	**.95**	**.97**	(.90)	.96	.99	.99
stable 128	(.94)	.87	.98	.98	(.91)	.82	.99	.99
		Armax(0.7)				Armax(0.8)		
block maxima		.93	.93	.92		.92	.91	.91
peaks o. threshold		.78	.79	.79		.66	.72	.74
stable 16	(.21)	.92	.94	.93	(.12)	.80	.82	.84
stable 32	(.66)	.90	.90	.91	(.55)	.87	.89	.90
stable 64	(.89)	**.93**	**.96**	**.96**	(.85)	**.90**	**.93**	**.93**
stable 128	(.94)	.85	.97	.98	(.92)	.83	.95	.96

4.3.6 Simulation study in the multivariate case

We inquiry now the performance of the multivariate, as opposed to the component-wise, stable sums estimator as we aim to capture the spatial features of extremes. We compute both estimates for the samples from the mARMAX$_\tau$ model with $\lambda = (0.7, 0.7, 0.7)$ as in (4.3.12); see Section 4.3.2 for details. We compare the performance of both estimators at each coordinate, $j = 1, 2, 3$, in terms of the relative percentage change of the mean squared error. More precisely, for each coordinate, we compute mean squared errors of the multivariate and univariate estimates denoted MSE_{MV} and MSE_{UV}, respectively, and relate them by

$$\text{relative percentage change of MSE} = \frac{MSE_{UV} - MSE_{MV}}{MSE_{UV}} \times 100. \tag{4.3.16}$$

We also compute the relative percentage change of the squared variance, and of the absolute bias, from equations similar to (4.3.16). Large positive values point to an improvement of the multivariate estimator, while negative values detect a deterioration of its performance.

We omit details on coverage probabilities as they both have similar coverage as the ARMAX(0.7) univariate model (as expected from (4.3.12)). We analyze in detail estimates $z_T(3)$ of the $T = 50$ years return level as similar results hold for all other coordinates. The relative percentage changes are plotted in figure 4.4 as a function of the spatial dependence coefficient τ. We notice that for the sum lengths 32 and 64 the multivariate outperforms the univariate estimator. Indeed, the choice of sum length 64 was optimal for the ARMAX(0.7) univariate model as pointed out by Table 4.1.

Moreover, for values of τ close to 0.5, the multivariate estimator has an outstanding improvement, mainly due to a diminution of bias. As τ approaches 1, and the model approaches the regime of asymptotic independence, the multivariate also outperforms the component-wise estimator though the choice of sum length becomes delicate. In contrast, the amelioration is less evident for values of τ close to 0. Recall from equation (4.3.15) that $\tau = 0$ points to asymptotic dependence and thus $m(j) = 1/d^{\tau} = 1$ can be difficult to estimate. We conclude that in general the multivariate estimator is preferable to the component-wise approach. Identifying, both theoretically and practically, which spatial features efficiently improve the multivariate inference procedures requires further investigation.

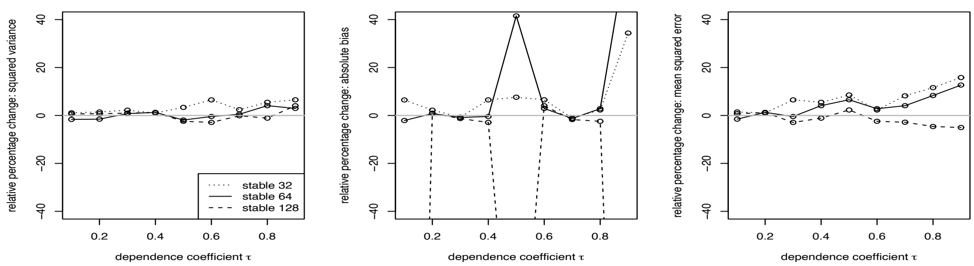


Figure 4.4: Relative percentage change of the multivariate against the component-wise estimator (for the 50 years return level $z_T(3)$ estimator) of the: squared variance, absolute bias and mean squared error as in (4.3.16), from left to right.

4.4 Case study of heavy rainfall in France

We recall the data set of fall daily rainfall introduced in Section 4.1 and our goal of computing the expected level of daily rain to be exceeded in the next 50 years at all the nine weather stations in France. We conduct our analysis separately over the three different regions: northwest, south, and northeast of France. Fall observations from the same region are modelled as a 3-dimensional sample $(\mathbf{X}_t)_{t=1,\dots,n}$ from a stationary multivariate regularly varying time series denoted $\mathbf{X}_t := (X_t(1), X_t(2), X_t(3))$, $t \in \mathbb{Z}$. We include both wet and dry days in our daily observations. In this setting, our goal of estimating the expected fall daily rainfall level to be exceeded in the next 50 years at each station traduces to estimating the 99.98-th quantile of $X_0(j)$, for $j = 1, 2, 3$.

4.4.1 Implementation

To study the samples $(\mathbf{X}_t)_{t=1,\dots,n}$ obtained from each region, we implement the stable sums method as a function of the number of order statistics k in the following way. For $k = 150, 250, 350, 450, 550$, first we compute estimates $\widehat{\alpha}^n(k)$ as described in Section 4.2.3. Then, for each estimate we search the sum length larger than 32 for which the $p-$value of the ratio likelihood test from Algorithm 1 is minimized. We look only among the sum lengths from the first 20 acceptances of the test. For comparison, we also implement classical methods as a function of the number of order statistics k as follows: we compute the peaks over threshold method for threshold levels $th(k) = X_{(k)}$ such that $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$, and we compute the block maxima method for blocks of length $bl_{BM}(k) := n/k$.

4.4.2 Analysis of the radial component

At each region, we start by studying the supremum norm observations, i.e. $(|\mathbf{X}_t|)_{t=1,\dots,n}$. We apply all three methods to estimate confidence intervals for the 50 years return level of Fall observations of the supremum norm. The obtained estimates are presented in Figure 4.5 where the rows correspond to different regions and the columns correspond to different methods. We notice that, as suggested by the simulation study in Section 4.3, the confidence intervals obtained with the peaks over threshold method might be too narrow and underestimate the expected return level. We also remark that the block maxima method varies strongly for different block length choices in Figure 4.5, thus a careful choice is required. Finally, we conclude that the stable sums method, illustrated in the third column of Figure 4.5, gives robust estimates as a function of k.

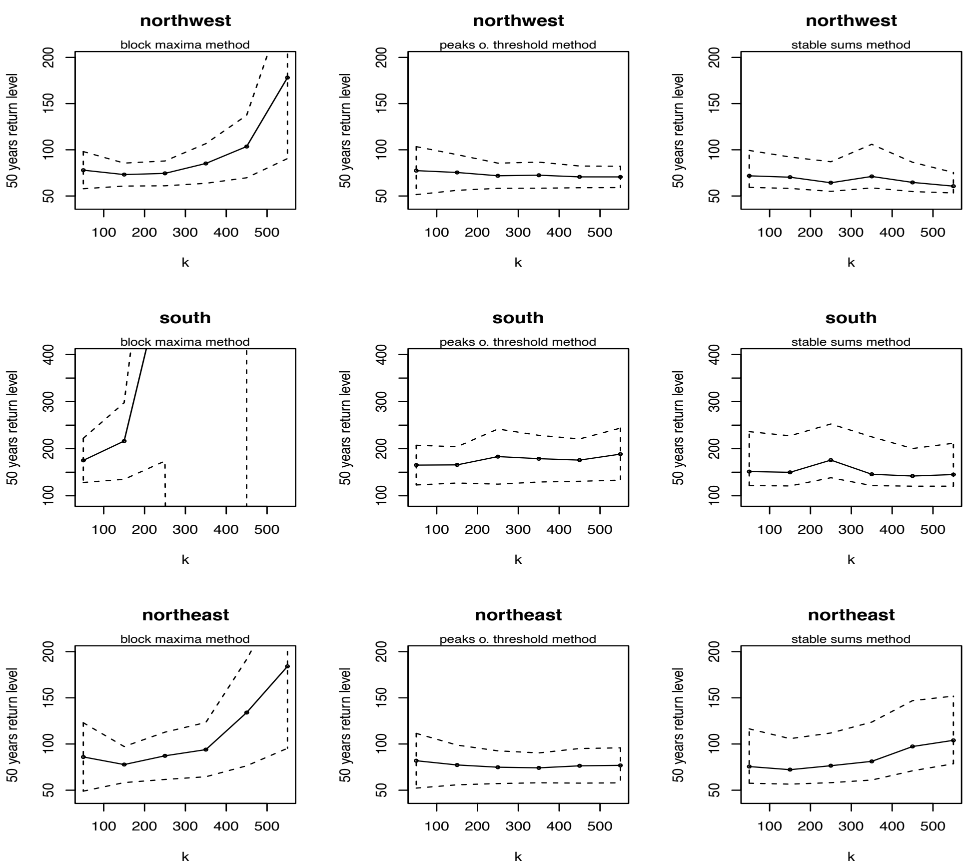


Figure 4.5: Estimates of the 50 years return level of fall supremum norm observations with confidence intervals. We write estimates as a function of k with the parametrization detailed in Section 4.4.1.

Moreover, we look at qqplots of the observed records $(S_{i,bl}(\widehat{\alpha}^n(k))^{1/\widehat{\alpha}^n(k)})_{i=1,\dots,\lfloor n/bl \rfloor}$ against the theoretical stable quantiles to the power $1/\widehat{\alpha}^n(k)$, for $k = 150, 250, 350, 450$, with sum lengths chosen as detailed in Section 4.4.1. Figures 4.6, 4.7, 4.8 contain the qqplots for the northwest, south and northeast, respectively, and allow us to assess goodness of fit for the different choices of k. We conclude from Figure 4.6 that for the northwest locations, the choice $k = 350$, $bl = 165$ captures nicely the intermediate and extreme quantiles. For the southern region, we see in Figure 4.7 that the choice $k = 250$ and $bl = 105$ gives an accurate fit. Lastly, for the northeast region, Figure 4.8 suggests the choices $k = 350$ and $bl = 70$, or $k = 450$ and $bl = 53$, for a correct alignment of intermediate and high quantiles.

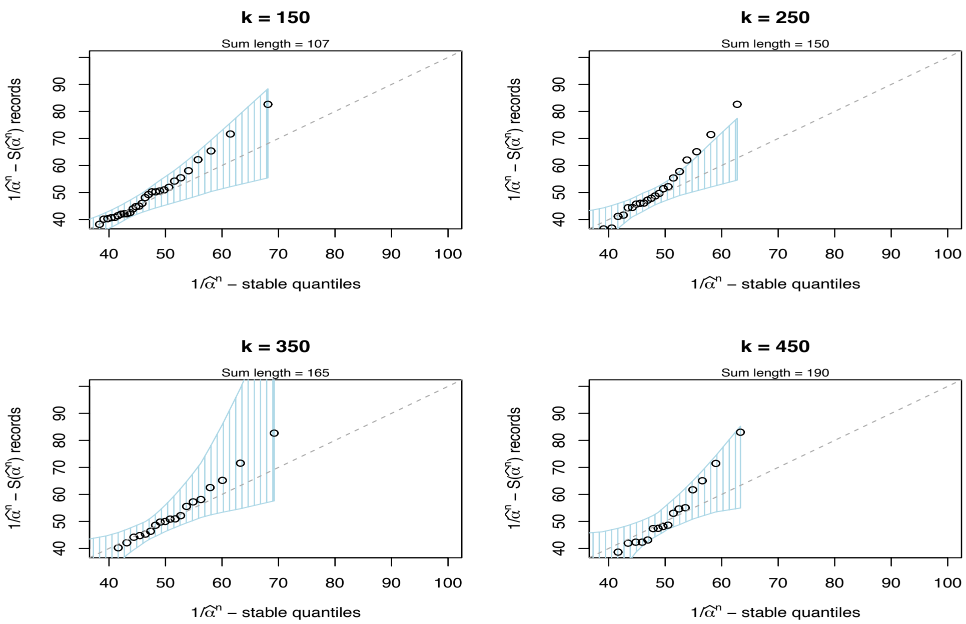


Figure 4.6: qqplots for different k values of the $1/\widehat{\alpha}^n(k)$–stable quantiles against the $1/\widehat{\alpha}^n(k)$–$(S_{t,b}(\widehat{\alpha}^n(k)))_{t=1,\dots,\lfloor n/b\rfloor}$ records with 95% confidence intervals for the northwest.

4.4.3 Analysis of the multivariate components

Finally, we turn back to the component-wise analysis. In this case, we must estimate the indexes of spatial clustering. Relying on (4.2.8) we obtain estimates: $\widehat{m}^n = (0.4966, 0.2709, 0.5744)$ corresponding to the weather stations at Brest, Lanveoc and Quimper in the northwest region; $\widehat{m}^n = (0.6064, 0.4706, 0.2866)$ for Bormes, Le Luc and Hyeres in the south; and $\widehat{m}^n = (0.3910, 0.4448, 0.5649)$ for Nancy, Metz, Roville in the northwest.

Roughly speaking, we interpret (4.1.3) to say: high daily rainfall levels at each weather station can be modelled as high quantiles of a stable distribution. In particular, letting the largest order statistics from each station play the role of the sequence of high threshold levels in (4.1.3), we deduce an empirical version of this relation which can be rewritten as

$$\mathbb{P}((S_{b_n}(\widehat{\alpha}^n))^{1/\widehat{\alpha}^n} \leq X_{(k)}(j)) \;\approx\; 1 - \frac{k}{\widehat{m}^n(j)\, n/b_n}\,, \tag{4.4.17}$$

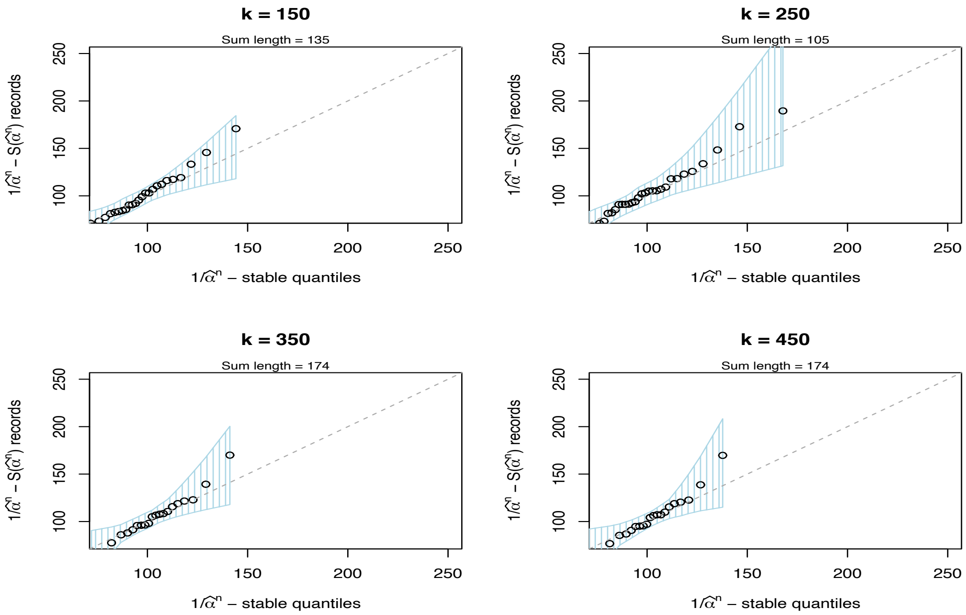


Figure 4.7: qqplots for different k values as in Figure 4.6 but for the southern region.

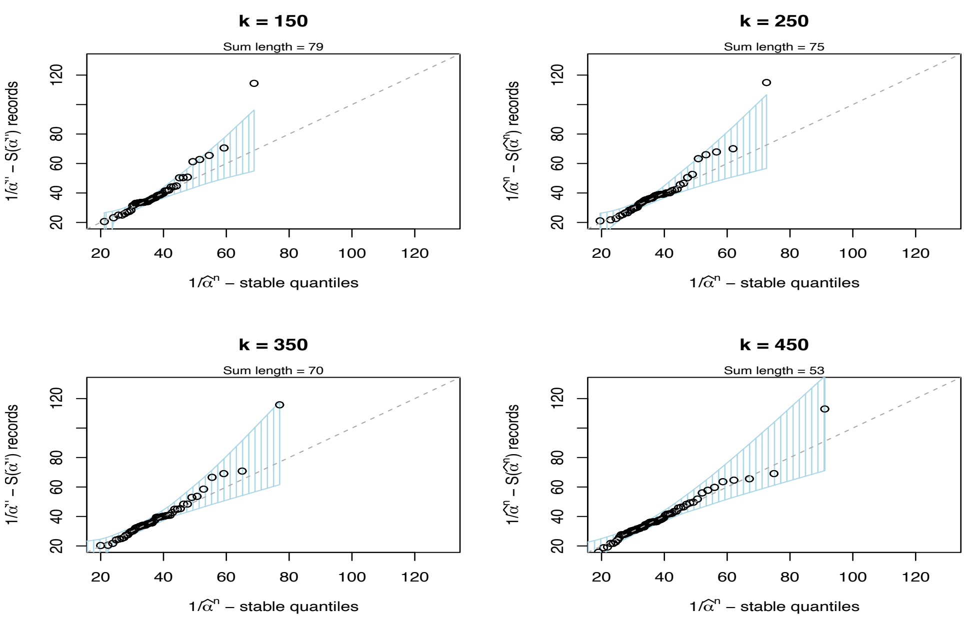


Figure 4.8: qqplots for different k values as as in Figure 4.6 but for the northeast region.

where $X_{(1)}(j) \geq X_{(2)}(j) \geq \cdots \geq X_{(\lfloor n/b_n \rfloor)}(j)$, for $j = 1, \ldots, d$. However, the approximation in (4.4.17) is only justified for the largest observations recorded. Moreover, the left-hand side in (4.4.17) can be approximated from the stable distribution fitted in line 5 of Algorithm 1.

In this way, we inspect the Equation (4.4.17) by plotting the sample largest order statistics $(X_{(t)}(j))_{t=1,\ldots,\lfloor \widehat{m}^n(j) n/b_n \rfloor}$, against the $1/\widehat{\alpha}^n$–stable quantiles from the distribution fitted in line 5 of Algorithm 1 for the multivariate stable sums method. For comparison with the univariate approach, we also plot the largest order statistics against the $1/\widehat{\alpha}^n$–stable quantiles from the univariate implementation as the following relation is also justified from (4.1.3)

$$\mathbb{P}((\textstyle\sum_{t=1}^{b_n}(X_t(j))^{\widehat{\alpha}^n(k)})^{1/\widehat{\alpha}^n} \leq X_{(k)}(j)) \;\approx\; 1 - \frac{k}{n/b_n}\,. \tag{4.4.18}$$

The plots are displayed in Figure 4.9 with points in black and grey for the multivariate and univariate approach, respectively. We use the estimates $\widehat{m}^n$, presented at the beginning of this section, and the tuning parameters $\widehat{\alpha}^n, bl$ from Section 4.4.2, pointing to a nice fit of the radial component. In particular, we set $k = 350$, $bl = 165$ for the northwest region, $k = 250$, $bl = 105$ for the south and $k = 350$, $bl = 70$ for the northeast. We interpret the largest records close to the diagonal as a nice fit. We remark from Figure 4.9 that overall the plots from the multivariate approach in black describe more accurately the most extreme observations compared to the univariate approach in gray. In particular, we can see a big improvement in the northeast region. For this region we also use the visual tool of Figure 4.9 to discard the choice of parameters $k = 450$ and $bl = 53$ as we found the fit from Figure 4.9 to be more accurate.

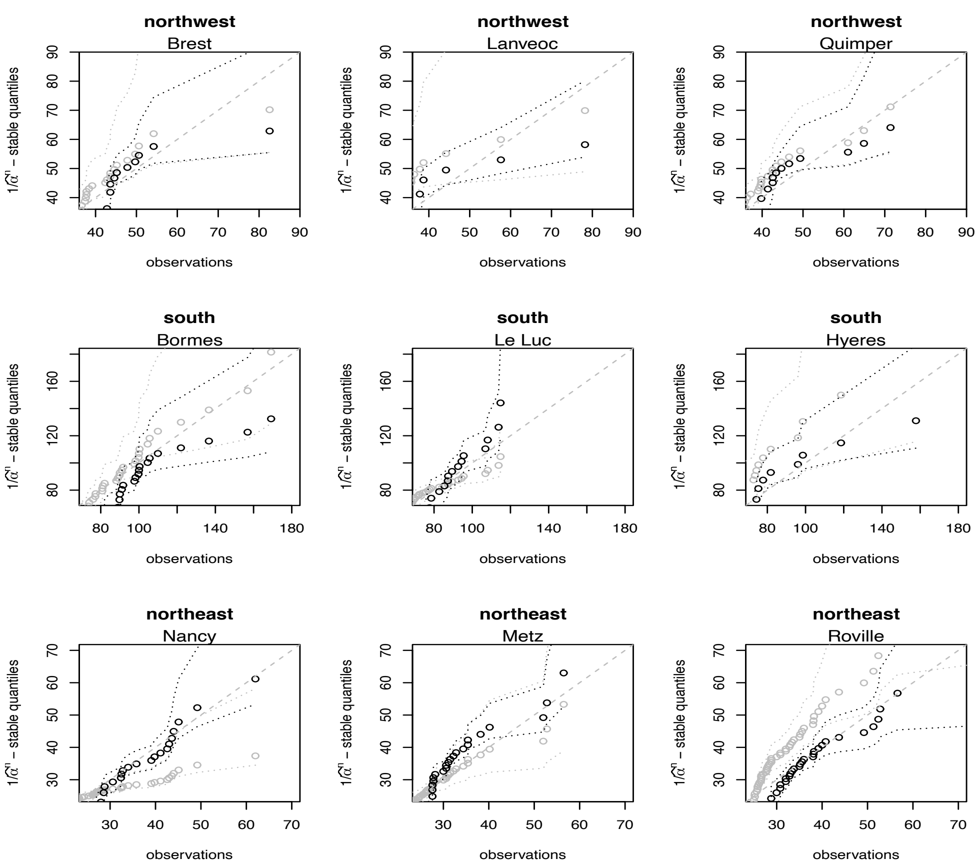


Figure 4.9: Plot of observations vs. $1/\widehat{\alpha}^n$ stable quantiles for the multivariate (in black) and component-wise estimators (in grey) with confidence intervals. The dotted line is the identity map $x \mapsto x$.

The intermediate quantiles shouldn't necessarily align, and in practice, there is not a clear procedure for knowing how many of the top quantiles should line up with the diagonal. Furthermore, we notice the observations are limited when the estimates of the spatial index: $m(j)$, are close to zero and then the graphical analysis is less reliable than for values of it close to one; see e.g. the stations of Lanveoc and Hyeres. We conclude that the multivariate method captures accurately the highest rainfall records, and supported by the numerical results from Section 4.3.6, it is justified for addressing the spatial dependencies of extremes. Using the choice of parameters k, bl, tuned for each region, and the estimates $\widehat{m}^n$, we can calculate confidence intervals for the 50-years return levels using Algorithm 1.

4.5 Conclusions

Atmospheric conditions drive the heavy-rainfall measurements. These records have a spatial and temporal coverage explained by the storm/fonts dynamics. Typically, an extreme event with a common source is recorded simultaneously at different locations and over different time lags. In this work, we have proposed the stable sums method to aggregate space and time information of dependent observations. Our ultimate goal was to extrapolate high quantiles at each weather station.

Our stable sums approach could also be used to address other environmental extremal problems. We now comment on one of them based on the idea that the asymptotics of space and time multivariate extreme events can be summarized by the univariate random variable of partial sums. In this work, we allocated weights $m(j)$ to each coordinate to compute the marginal features like the set $\{X(j) > x\}$. Conceptually, it should be also possible to study other d-dimensional extremal sets, for example, $\{X(j) > x, X(j\prime) > x\}$. Applying the theoretical results of [26] will introduce weights of the type $m(j, j')$. Still, our take-home message will remain the same: important $d-$dimensional features are accessible by fitting only the univariate sums.

4.6 Asymptotic theory

4.6.1 Invariance principle

We work under the anti-clustering condition below tailored to avoid long-range dependence of extremes. Similar conditions are also needed to justify the declustering procedures [65].

Anti-clustering condition: There exists a sequence (x_n) such that for all $\epsilon > 0$,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{P}\big(\max_{t=k,\dots,n} |\mathbf{X}_t| > \epsilon\, x_n \,\big|\, |\mathbf{X}_0| > \epsilon\, x_n \big) \;=\; 0, \tag{4.6.19}$$

and $n\,\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$ as $n \to +\infty$.

To study the long-term extremal dependencies we review below Proposition 3.2 in [26].

Lemma 4.6.1. *Let $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ be a regularly varying time series in $(\mathbb{R}^d, |\cdot|)$ with index of regular variation $\alpha > 0$, and spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$; see 4.2.7. Under (4.6.19), $\|\Theta_t\|_\alpha^\alpha < +\infty$ a.s. and $\mathbf{Q}_t = \Theta_t/\|\Theta_t\|_\alpha$, for $t \in \mathbb{Z}$, is well defined. We call $(\mathbf{Q}_t)_{t\in\mathbb{Z}}$ the cluster process.*

To study the symptotics of partial sums of p-powers, we require also an assumption to control the moderate values of sums. Similar conditions where considered in [5, 40, 121, 26].

Vanishing small values condition: There exists a sequence (x_n) verifying (4.6.19), $n/x_n^{\alpha-\kappa} \to 0$, as $n \to \infty$ for some $\kappa > 0$, and for all $\delta > 0$

$$\lim_{\epsilon\downarrow 0}\limsup_{n\to+\infty} \frac{\mathbb{P}\big(\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|\le \epsilon\, x_n\}} > \delta\, x_n^\alpha\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \;=\; 0. \tag{4.6.20}$$

Notice that since $n/x_n^{\alpha-\kappa} \to 0$ for some $\kappa > 0$ in (4.6.20), then $n\mathbb{E}[|\mathbf{X}_0/x_n|^\alpha \mathbb{1}_{\{|\mathbf{X}_0|\le x_n\}}] \to 0$, which implies $S_n(\alpha)/x_n^\alpha \xrightarrow{\mathbb{P}} 0$ as $n \to +\infty$. Hence,

$$\begin{aligned}
\frac{\mathbb{P}\big(\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|\le \epsilon\, x_n\}} > \delta\, x_n^\alpha\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} &\le \delta^{-1}\,\frac{\mathrm{Var}\big(\sum_{t=1}^n |\mathbf{X}_t/x_n|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|\le \epsilon\, x_n\}}\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \\
&\le \delta^{-1}\,\frac{\mathbb{E}\big[|\mathbf{X}/x_n|^{2\alpha} \mathbb{1}_{\{|\mathbf{X}_0|\le \epsilon\, x_n\}}\big]}{\mathbb{P}(|\mathbf{X}_0| > x_n)}\big(1 + \textstyle\sum_{t=1}^n \rho_t\big),
\end{aligned}$$

where $\rho_t \in [0,1]$ is a correlation coefficient defined as $\rho_t := corr\big(|\mathbf{X}_0|^{\alpha-\kappa}, |\mathbf{X}_t|^{\alpha-\kappa}\big)$, for some $\kappa > 0$. If $\sum_{t=1}^\infty \rho_t < +\infty$, then an application of Karamata's theorem yields an asymptotic upper bound given by $\delta^{-1}\,\epsilon^\alpha(1 + \sum_{t=1}^{+\infty}\rho_t)$ and (4.6.20) holds by letting $\epsilon \downarrow 0$.

We review Lemma 4.1 [26] showing $c(\alpha) = 1$ in (4.1.3). The proof is postponed to Section 4.7.1.

Lemma 4.6.2. *Let $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$–valued regularly varying time series with index of regular variation $\alpha > 0$. Assume it verifies conditions (4.6.19) and (4.6.20) for a sequence (x_n). Then*

$$\lim_{n\to+\infty} \frac{\mathbb{P}\big(S_n(\alpha) > x_n^\alpha\big)}{n\,\mathbb{P}(|\mathbf{X}_0| > x_n)} \;=\; 1, \quad n \to +\infty. \tag{4.6.21}$$

In this way, the choice $p = \alpha$ in (4.1.3) is a declustering method due to the invariance property in (4.6.21). Indeed, the unit limit holds regardless of the underlying time dependence dynamic of extremes for numerous examples verifying (4.6.19) and (4.6.20) [10, 24, 121].

4.6.2 Central limit theorem

We state the limit theory of partial sums of α-powers of regularly varying increments, for α the tail index. For $p/\alpha \in (0,1)\cup(1,2)$ we refer to Proposition 5.4. in [26]. We require the time series of α-powers to satisfy the mixing condition below close to condition (2.8) in [5].

Mixing condition: Let $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$–valued regularly varying time series with index of regular variation equal to one. Assume for all $\epsilon > 0$, $\mathbf{u} \in \mathbb{R}^d$, there exists integer sequences $k = k_n \to \infty$,

(a_n), satisfying $n/k_n \to \infty$, $n\,\mathbb{P}(|\mathbf{Z}_0| > a_n) \to 1$ and

$$\big|\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top \textstyle\sum_{t=1}^n \mathbf{Z}_t/a_n\big\}\big] - \mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\ \textstyle\sum_{t=1}^k \mathbf{Z}_t/a_n\big\}\big]^{\lfloor n/k \rfloor}\big| \to 0\,, \quad n \to +\infty. \tag{4.6.22}$$

The mixing condition in (4.6.22) holds if the decay of the α-mixing coefficients happens sufficiently fast; cf. Lemma 3.8. in [5] where (α_t) is defined, for all $h \in \mathbb{N}$, as

$$\alpha_h \ := \sup_{A\in\sigma(\,(X_t)_{t\le 0}\,),B\in\sigma(\,(X_t)_{t\ge h}\,)} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Proposition 4.6.3. *Consider $(\mathbf{X}_t)$ to be an $\mathbb{R}^d$–valued regularly varying time series with index of regular variation $\alpha > 0$. Let $(a_n(\alpha))$, $(d_n(\alpha))$, be sequences such that $n\,\mathbb{P}(|\mathbf{X}_0|^\alpha > a_n(\alpha)) \to 1$ and $d_n(\alpha) := \mathbb{E}[|\mathbf{X}_t|^\alpha \mathbb{1}(|\mathbf{X}_t|^\alpha \le a_n(\alpha))]$. If $(|\mathbf{X}_t|^\alpha)_{t\in\mathbb{Z}}$ verifies conditions (4.6.22), (4.6.19), (4.6.20) simultaneously; see Remark 4.6.4, then*

$$(S_n(\alpha) - d_n(\alpha))/a_n(\alpha) \ \xrightarrow{d}\ \xi_1, \quad n \to +\infty$$

where ξ_1 is stable distributed with $a = 1$ and $\beta = 1$ following the notation in (4.2.6).

Proposition 4.6.3 and Lemma 4.6.2 justify Algorithm 1 built on the partial sums of $\alpha-$powers.

Conditions (4.6.22), (4.6.19), have been verified on numerous examples under weakly mixing assumptions; cf. [5, 121] and references therein.

Remark 4.6.4. *Conditions (4.6.22), (4.6.19) and (4.6.20) hold simultaneously if there exists an integer sequence (k_n) such that (x_n), defined by $x_{k_n} = a_n(\alpha)$, satisfies (4.6.19) and (4.6.20), (k_n) satisfies (4.6.22), and $n\,\mathbb{P}(|\mathbf{X}_0|^\alpha > a_n(\alpha)) \to 1$ as $n \to \infty$. Conditions (4.6.19) and (4.6.20) are tailored to study the extreme behavior of $(|\mathbf{X}_t|^\alpha)_{t=1,\dots,n}$ as (x_n) satisfies $\mathbb{P}(\sum_{t=1}^n |\mathbf{X}_0|^\alpha > x_n) \sim n\,\mathbb{P}(|\mathbf{X}_0|^\alpha > x_n) \to 0$, and*

$$k_n\,\mathbb{P}(|\mathbf{X}_0|^\alpha > x_{k_n}) \ = \ k_n\,(n\,\mathbb{P}(|\mathbf{X}_t|^\alpha > a_n(\alpha)))/n \ \sim\ k_n/n \ \to\ 0, \quad n \to +\infty.$$

4.7 Proofs

4.7.1 Proof of Lemma 4.6.2

We rely on telescopic sum arguments from [97, 98]; see [5]. For all $\epsilon > 0$, $\delta > 0$,

$$\begin{aligned} I := \mathbb{P}(S_n(\alpha) > x_n^\alpha) = \mathbb{P}\big(S_n(\alpha) > x_n^\alpha,\, \textstyle\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|\le\epsilon\, x_n\}} < \delta\, x_n^\alpha\big) \\ + \mathbb{P}\big(S_n(\alpha) > x_n^\alpha,\, \textstyle\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|\le\epsilon\, x_n\}} > \delta\, x_n^\alpha\big). \end{aligned}$$

Referring to condition (4.6.20), the probability term above satisfies

$$\mathbb{P}\big(\textstyle\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > x_n^\alpha\big) \le I \le \mathbb{P}\big(\textstyle\sum_{t=1}^n |\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta) x_n^\alpha\big) + o(n\,\mathbb{P}(|\mathbf{X}_0| > x_n)).$$

Hence, to show (4.6.21) it suffices to prove that for all $\delta > 0$ the following relation holds

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\lim_{n\to+\infty}\frac{\mathbb{P}\big(\sum_{t=1}^{n}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\big)}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)} = 1. \tag{4.7.23}$$

Using the so-called telescopic sum argument, and by condition (4.6.19), for all $K > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\textstyle\sum_{t=1}^{n}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\big)\\
&= \textstyle\sum_{j=1}^{n-1}\Big\{\mathbb{P}\big(\sum_{t=1}^{j+1}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\,,\,|\mathbf{X}_1|>\epsilon\, x_n\big)\\
&\quad - \mathbb{P}\big(\textstyle\sum_{t=2}^{j+1}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\,,\,|\mathbf{X}_1|>\epsilon\, x_n\big)\Big\} + \mathbb{P}(|\mathbf{X}_1|^\alpha > (1-\delta)x_n^\alpha)\\
&= \textstyle\sum_{j=K}^{n-1}\Big\{\mathbb{P}\big(\sum_{t=1}^{K}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\,,\,|\mathbf{X}_1|>\epsilon\, x_n\big)\\
&\qquad\qquad - \mathbb{P}\big(\textstyle\sum_{t=2}^{K}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\,,\,|\mathbf{X}_1|>\epsilon\, x_n\big)\Big\}\\
&\qquad\qquad + \mathbb{P}(|\mathbf{X}_1|^\alpha > (1-\delta)x_n^\alpha) + o(n\,\mathbb{P}(|\mathbf{X}_0|>x_n)).
\end{aligned}$$

Then, writing $(\Theta_t)_{t\in\mathbb{Z}}$ as in (4.2.7), we obtain

$$\begin{aligned}
II := \lim_{n\to+\infty}&\frac{\mathbb{P}\big(\sum_{t=1}^{n}|\mathbf{X}_t|^\alpha \mathbb{1}_{\{|\mathbf{X}_t|>\epsilon\, x_n\}} > (1-\delta)x_n^\alpha\big)}{n\,\mathbb{P}(|\mathbf{X}_0|>x_n)}\\
&= \lim_{\epsilon\downarrow 0}\lim_{K\to+\infty}\big\{\epsilon^{-\alpha}\textstyle\int_1^\infty \mathbb{P}\big(\sum_{t=0}^{K}|\epsilon y\Theta_t|^\alpha \mathbb{1}_{\{y|\Theta_t|>1\}} > (1-\delta)\big)\\
&\qquad\qquad - \mathbb{P}\big(\textstyle\sum_{t=1}^{K}|\epsilon y\Theta_t|^\alpha \mathbb{1}_{\{y|\Theta_t|>1\}} > (1-\delta)\big)d(-y^{-\alpha})\big\}.
\end{aligned}$$

Indeed, the points of discontinuity at the limit are contained in $\cup_{t=1}^{K}\{Y|\Theta_t|=1\}$, which has zero probability. Then, using monotone convergence we take the limit as $K\to+\infty$ within the integral. Furthermore, the change of coordinates $u=\epsilon\, y$ entails

$$\begin{aligned}
II = \lim_{\epsilon\downarrow 0}&\big\{\textstyle\int_\epsilon^\infty \mathbb{P}\big(\sum_{t=0}^{\infty}|y\Theta_t|^\alpha \mathbb{1}_{\{y|\Theta_t|>\epsilon\}} > (1-\delta)\big)\\
&\qquad - \mathbb{P}\Big(\textstyle\sum_{t=1}^{\infty}|y\Theta_t|^\alpha \mathbb{1}_{\{y|\Theta_t|>\epsilon\}} > (1-\delta)\Big)d(-y^{-\alpha})\big\}.
\end{aligned}$$

We conclude by monotone convergence that we can take the limit as ϵ goes to zero at each term. As a result we obtain asymptotic equivalence with the term below

$$\begin{aligned}
&\sim \textstyle\int_0^\infty \mathbb{P}\big(\sum_{t=0}^{\infty}|y\Theta_t|^\alpha > (1-\delta)\big) - \mathbb{P}\big(\sum_{t=1}^{\infty}|y\Theta_t|^\alpha > (1-\delta)\big)d(-y^{-\alpha})\\
&= (1-\delta)^{-1}\mathbb{E}\big[\textstyle\sum_{t=0}^{\infty}|\Theta_t|^\alpha - \sum_{t=1}^{\infty}|\Theta_t|^\alpha\big] \;=\; (1-\delta)^{-1}.
\end{aligned}$$

In the last step we use that $|\Theta_0| = 1$. Finally, we conclude taking the limit as δ goes to zero that the relation 4.7.23 holds and this concludes the proof.

4.7.2 Proof of Proposition 4.6.3

To proof Proposition 4.6.3, we state in Theorem 4.7.1 a central limit theory for regularly varying increments of unitary index or regular variation. Our proof in Section 4.7.2 removes the assumption (4.10) in [9] and (**CT**) in Theorem 3.1. [5], treating the recentering term.

Theorem 4.7.1. *Let $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$–valued regularly varying time series with index of regular variation equal to one. Let (a_n), (d_n) be sequences verifying $n\,\mathbb{P}(|\mathbf{Z}_0| > a_n) \sim 1$ as $n \to +\infty$ and $d_n := \mathbb{E}[|\mathbf{Z}_t|\mathbb{1}(|\mathbf{Z}_t| \leq a_n)]$ and assume (4.6.22), (4.6.19), (4.6.20) hold simultaneously (as in Remark 4.6.4). Then, $\sum_{t=1}^{n}(\mathbf{Z}_t - d_n)/a_n$ converges in distribution to a stable distribution ξ_1 with unit stable parameter, and for all $\mathbf{u} \in \mathbb{R}^d$,*

$$\begin{aligned} &\log \mathbb{E}\big[\exp\big\{i\mathbf{u}^\top \xi_1\big\}\big] \\ &\quad = \int_0^\infty \mathbb{E}\big[\exp\big\{i\mathbf{u}^\top {\textstyle\sum_{t\in\mathbb{Z}}} y\mathbf{Q}_t^Z\big\} - 1 - i\sin\big(\mathbf{u}^\top {\textstyle\sum_{t\in\mathbb{Z}}} y\mathbf{Q}_t^Z\big)\big]d(-y^{-1}) + i\mu(\mathbf{u})\,, \end{aligned} \tag{4.7.24}$$

where the last term in (4.7.24) is a location parameter given by

$$\mu(\mathbf{u}) := \int_1^\infty \mathbb{E}\big[\sin\big(\mathbf{u}^\top {\textstyle\sum_{t\in\mathbb{Z}}} y\mathbf{Q}_t^Z\big) - \sin\big(\mathbf{u}^\top {\textstyle\sum_{t\in\mathbb{Z}}} y\mathbf{Q}_t^Z - \mathbf{u}^\top {\textstyle\sum_{t\in\mathbb{Z}}} \underline{y\mathbf{Q}_t^Z}_{\,1}\big)\big]d(-y^{-1}),$$

and $(\mathbf{Q}_t^Z)_{t\in\mathbb{Z}}$ refers to the cluster process of $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ as defined in Lemma 4.6.1 verifying $\|\mathbf{Q}_t^Z\|_1 = 1$ a.s.

Proposition 4.6.3 below follows straightforwardly from Theorem 4.7.1.

Proof of Proposition 4.6.3. Let $(\mathbf{Q}_t)_{t\in\mathbb{Z}}$ be the cluster process of $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ such that $\|\mathbf{Q}_t\|_\alpha = 1$ a.s. Define the time series $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$, $\mathbf{Z}_t = |\mathbf{X}_t|^\alpha$, with spectral process $(\mathbf{Q}_t^Z)_{t\in\mathbb{Z}}$ a.s. equal to $\mathbf{Q}_t^Z = |\mathbf{Q}_t|^\alpha$, for all $t \in \mathbb{Z}$, such that $\sum_{t=1}\mathbf{Q}_t^Z = 1$ a.s. Theorem 4.7.1 entails $\sum_{t=1}^{n}(|\mathbf{X}_t|^\alpha - d_n(\alpha))/a_n(\alpha)$ admits a stable limit ξ_1 with log-characteristic function: $u \in \mathbb{R}$

$$\log \mathbb{E}[\exp\{i\,u\,\xi_1\}] = \int_0^\infty \mathbb{E}\big[\exp\big\{i\,uy\big\} - 1 - i\sin(uy)\big]d(-y^{-1}) + i\mu(u)\,,$$

Then, following the lines of the argument in section XVII.2 of Feller [67], we deduce the skweness parameter of the stable limit ξ_1 verifies $\beta = 1$. □

Proof of Theorem 4.7.1

Proof. Let $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^d$–valued regularly varying time series with index equal to one. We introduce the truncation notation where, for $\epsilon > 0$, we denote $\mathbf{S}_n := \sum_{t=1}^{n}\mathbf{Z}_t$ and

$$\underline{\mathbf{S}_n/a_n}_{\epsilon} := {\textstyle\sum_{t=1}^{n}} \mathbf{Z}_t/a_n \mathbb{1}_{\{|\mathbf{Z}_t| > \epsilon\, a_n\}}, \quad \overline{\mathbf{S}_n/a_n}^{\epsilon} := {\textstyle\sum_{t=1}^{n}} \mathbf{Z}_t/a_n \mathbb{1}_{\{|\mathbf{Z}_t| \leq \epsilon\, a_n\}}.$$

We also consider a truncation of the centering sequence $(d_n)_{n\in\mathbb{N}}$ defined by

$$\underline{d_n/a_n}_{\epsilon} = \mathbb{E}[|\mathbf{Z}_0/a_n|\mathbb{1}_{\{\epsilon\, a_n < |\mathbf{Z}_0| \leq a_n\}}], \quad n \in \mathbb{N}.$$

To simplify we denote the cluster process $(\mathbf{Q}_t^Z)_{t\in\mathbb{Z}} := (\mathbf{Q}_t)$.

From the mixing condition in (4.6.22) we deduce there exists a sequence $k := k_n \to +\infty$ such that, for all $\mathbf{u} \in \mathbb{R}^d$,

$$\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\mathbf{S}_n/a_n - n\,d_n/a_n)\big\}\big] \sim \mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\mathbf{S}_k/a_n - k\,d_n/a_n)\big\}\big]^{\lfloor n/k\rfloor}.$$

as $n \to +\infty$. Then, taking the logarithm at both sides yields

$$\begin{aligned}\log\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\mathbf{S}_n/a_n - n\,d_n/a_n)\big\}\big] &\sim \frac{n}{k}\log\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\mathbf{S}_k/a_n - k\,d_n/a_n)\big\}\big]\\ &\sim \frac{\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\mathbf{S}_k/a_n - k\,d_n/a_n)\big\}\big]-1}{k\,\mathbb{P}(|\mathbf{Z}_0|>a_n)}\\ &\sim \frac{\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\underline{\mathbf{S}_k/a_n}_\epsilon - k\,\underline{d_n/a_n}_\epsilon)\big\}\big]-1}{k\,\mathbb{P}(|\mathbf{Z}_0|>a_n)}.\end{aligned}$$

such that the second step is granted since $k\,d_n/a_n \to 0$ as $n \to +\infty$ and also $\mathbf{S}_k/a_n \xrightarrow{\mathbb{P}} 0$ as $n \to +\infty$ and the last step follows by the vanishing-small-values condition in (4.6.20) and boundedness of the exponential function. Then, it follows from a Taylor expansion that

$$\begin{aligned}\Big|\big(\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\underline{\mathbf{S}_k/a_n}_\epsilon - k\,\underline{d_n/a_n}_\epsilon)\big\}\big]-1\big)\\ -\big(\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\,\underline{\mathbf{S}_k/a_n}_\epsilon\big\}\big]-1-i\mathbb{E}[\sin(\mathbf{u}^\top\,\overline{\underline{\mathbf{S}_k/a_n}_\epsilon}^1)]\big)\Big|\\ = O\big(k\,\underline{d_n/a_n}_\epsilon\mathbb{E}[|\underline{\mathbf{S}_k/a_n}_\epsilon|\mathbb{1}(|\underline{\mathbf{S}_k/a_n}_\epsilon|\le 1)]\big).\end{aligned}$$

Moreover, $|k\,\underline{d_n/a_n}_\epsilon\mathbb{E}[|\underline{\mathbf{S}_k/a_n}_\epsilon|\mathbb{1}(|\underline{\mathbf{S}_k/a_n}_\epsilon|\le 1)]|/k\,\mathbb{P}(|\mathbf{Z}_0|>a_n) \to 0$ as $n \to +\infty$. Thus, we obtain the asymptotic equivalence

$$\log\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top(\underline{\mathbf{S}_n(\alpha)/a_n}_\epsilon - n\,\underline{d_n/a_n}_\epsilon)\big\}\big] \sim \frac{\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\,\underline{\mathbf{S}_k/a_n}_\epsilon\big\}-1-i\sin(u\,\overline{\underline{\mathbf{S}_k/a_n}_\epsilon}^1)\big]}{k\,\mathbb{P}(|\mathbf{Z}_0|>a_n)}.$$

as $n \to +\infty$. Furthermore,

$$\begin{aligned}&\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\,\underline{\mathbf{S}_k/a_n}_\epsilon\big\}-1-i\sin(\mathbf{u}^\top\,\overline{\underline{\mathbf{S}_k/a_n}_\epsilon}^1)\big]\\ &\quad= \mathbb{E}\big[\big(\exp\big\{i\,\mathbf{u}^\top\,\underline{\mathbf{S}_k/a_n}_\epsilon\big\}-1\big)\mathbb{1}_{\{|\mathbf{S}_k|>\epsilon\,a_n\}}\big] - \mathbb{E}\big[i\sin(\mathbf{u}^\top\,\overline{\underline{\mathbf{S}_k/a_n}_\epsilon}^1)\mathbb{1}_{\{|\mathbf{S}_k|>\epsilon\,a_n\}}\big].\end{aligned}$$

For all $\mathbf{x} \in \mathbb{R}^d$, we denote $\mathbf{x}\mathbb{1}_{|\mathbf{x}|>\epsilon}$ by $\underline{\mathbf{x}}_\epsilon$, and similarly we denote $\mathbf{x}\mathbb{1}_{|\mathbf{x}|\le 1}$ by $\overline{\mathbf{x}}^1$. Then, conditioning to the event $\{|\mathbf{S}_k| > \epsilon\,a_n\}$, we use the limit relation in (4.6.21) and Proposition 4.2. in [26] and take the limit as n goes to infinity in the above expression. Hence,

$$\begin{aligned}&\frac{\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\underline{\mathbf{S}_k/a_n}_\epsilon\big\}-1-i\sin(\mathbf{u}^\top\,\overline{\underline{\mathbf{S}_k/a_n}_\epsilon}^1)\big]}{k\mathbb{P}(|\mathbf{Z}_0|>a_n)}\\ &\quad\sim \textstyle\int_0^\infty\mathbb{E}\big[\exp\big\{i\mathbf{u}^\top\sum_{t\in\mathbb{Z}}\underline{y\mathbf{Q}_t}_\epsilon\big\}-1-i\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}}(\overline{\underline{y\mathbf{Q}_t}_\epsilon}^1)\big)\big]d(-y^{-1}),\end{aligned}$$

where $(\mathbf{Q}_t)_{t\in\mathbb{Z}}$ is the cluster process of the stationary process $(\mathbf{Z}_t)$. In particular, it takes values in $\mathbb{R}^{\mathbb{Z}}$ and verifies $\sum_{t\in\mathbb{Z}}|\mathbf{Q}_t| = 1$ with probability one.

Let $\delta > 0$ and let's divide the integral above on the events $\{y > \delta\}$ and $\{y \le \delta\}$. Over the event $\{y \le \delta\}$ we have that given we choose $\delta < 1$,

$$
\begin{aligned}
&\textstyle\int_0^\infty \mathbb{E}\big[\exp\big\{i\mathbf{u}^\top \sum_{t\in\mathbb{Z}} \underline{y\mathbf{Q}_t}_\epsilon\big\} - 1 - i\sin\big(\mathbf{u}^\top \sum_{t\in\mathbb{Z}} \overline{\underline{y\mathbf{Q}_t}_\epsilon}^1\big)\big]\mathbb{1}(y\le\delta)\,d(-y^{-1}) \\
&\qquad = \textstyle\int_0^\infty \mathbb{E}\big[\exp\big\{i\mathbf{u}^\top \sum_{t\in\mathbb{Z}} \underline{y\mathbf{Q}_t}_\epsilon\big\} - 1 - i\sin\big(\mathbf{u}^\top \sum_{t\in\mathbb{Z}} \underline{y\mathbf{Q}_t}_\epsilon\big)\big]\mathbb{1}(y\le\delta)\,d(-y^{-1}).
\end{aligned}
$$

Then, using the inequality $|\exp\{iz\} - 1 - i\sin(z)| \le |z|^2$ for all $z \in \mathbb{R}$, the integral above is bounded in absolute value by

$$
\textstyle\int_0^\infty \mathbb{E}\big[\big|\mathbf{u}^\top \sum_{t\in\mathbb{Z}}|y\mathbf{Q}_t|\big|^2\big]\mathbb{1}(y\le\delta)d(-y^{-1}) \le \delta\mathbb{E}\big[\big|\mathbf{u}^\top\sum_{t\in\mathbb{Z}}|\mathbf{Q}_t|\big|^2\big] = \delta|\mathbf{u}|^2 < +\infty.
$$

Then, we conclude that

$$
\begin{aligned}
&\log\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\,(\mathbf{S}_k/a_n - n\,d_n/a_n)\big\}\big] \\
&\qquad \sim \lim_{\delta\to 0}\textstyle\int_\delta^\infty \mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big\} - 1 - i\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}}\overline{y\mathbf{Q}_t}^1\big)\big]d(-y^{-1}).
\end{aligned}
$$

Moreover, we can rewrite the term above as the sum of two integrals as shown below

$$
\begin{aligned}
&\lim_{\delta\to 0}\textstyle\int_\delta^\infty \mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big\} - 1 - i\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big)\big]d(-y^{-1}) \\
&\qquad + i\textstyle\int_1^\infty \mathbb{E}\big[\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big) - \sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t - \mathbf{u}^\top\sum_{t\in\mathbb{Z}}\underline{y\mathbf{Q}_t}_1\big)\big]d(-y^{-1}) \\
&\quad = I + II
\end{aligned}
$$

such that the for the last term we can use the trigonometric relation $\sin(p) - \sin(p-q) = 2\sin(p/2)\cos(p-(q/2))$, for $p, q \in \mathbb{R}$, to obtain that the second term II is bounded in absolute value by $\int_1^{+\infty} y^{-2} = 1$. This term can be interpreted as a location parameter.

Finally, by monotone convergence, using the bound previously derived, we can take the limit as δ goes to 0 in I yielding

$$
\begin{aligned}
&\log\mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\,(\mathbf{S}_k/a_n - n\,d_n/a_n)\big\}\big] \\
&\qquad \sim \textstyle\int_0^\infty \mathbb{E}\big[\exp\big\{i\,\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big\} - 1 - i\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big)\big]d(-y^{-1}) \\
&\qquad\quad + i\textstyle\int_1^\infty \mathbb{E}\big[\sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t\big) - \sin\big(\mathbf{u}^\top\sum_{t\in\mathbb{Z}} y\mathbf{Q}_t - \mathbf{u}^\top\sum_{t\in\mathbb{Z}}\underline{y\mathbf{Q}_t}_1\big)\big]d(-y^{-1}).
\end{aligned}
$$

as $n \to +\infty$. We have shown that the limit relation from equation (4.7.24) holds and this concludes the proof. $\square$

Chapter 5: (Ongoing work)
Asymptotic normality of ℓ^p–cluster inference

Abstract

In the stationary regularly varying time series framework, extreme behavior occurs as short periods with several heavy recordings, known as extremal blocks. To summarize this clustering of heavy observations, we let functionals act on extremal blocks. Summary statistics of this type are known as cluster statistics. For example, we can recover the extremal index in this way. To address cluster statistics inference, we review the blocks estimators in [26]. These are built on large deviation principles of extremal ℓ^p-blocks, i.e., consecutive observations with large ℓ^p-norms, for $p > 0$. We show asymptotic normality of these blocks estimators relying on the theory in [52]. The goal of this work is twofold. First, we aim to review the arguments of empirical process theory. Second, we aim to study mixing conditions to guarantee asymptotic normality of the ℓ^p-blocks estimators in the stationary setting. We conclude by verifying the main assumptions on classic models as linear models and stochastic recurrence equations.

keywords: *Cluster processes; multivariate regular variation; stable distribution; stationary time series; extremal index; empirical processes theory.*

Main contributions

I address (**Quest. 3**) in this Chapter. The following are my main contributions:

- (**Quest. 3**): Consider the p-clusters $\mathbf{Q}^{(p)} \in \ell^p$, satisfying $\|\mathbf{Q}^{(p)}\|_p = 1$, a.s., introduced in Theorem 2.2.1. Let's denote $\widehat{f_p^{\mathbf{Q}}}$ the disjoint blocks estimators from Theorem 2.4.1, tailored to infer the cluster statistic

$$f_p^{\mathbf{Q}} = \mathbb{E}[f(Y\mathbf{Q}^{(p)})], \tag{5.0.1}$$

such that $f : \ell^p \to \mathbb{R}$ is a bounded continuous function, and Y is Pareto distributed, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, independent of $\mathbf{Q}^{(p)}$. Theorem 5.2.1 states that if

$$u \mapsto f(\mathbf{x}_t/u), \quad (\mathbf{x}_t) \in \ell^p, \quad u > 0,$$

is a non-increasing function, then

$$\sqrt{k}(\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{Var}(f(Y\mathbf{Q}^{(p)}))\big), \quad n \to +\infty. \tag{5.0.2}$$

holds under suitable mixing conditions.

Section 5.6 studies the main assumptions of Theorem 5.2.1 on classical models such as linear filters and stochastic recurrence equations.

Contents

5.1 Introduction

We consider stationary heavy-tailed time series with finite-dimensional distributions being multivariate regularly varying; see [10]. It is typical in this framework for extremal observations of the time series to cluster. Roughly speaking, this means a heavy record will trigger a short period with numerous extremal gauges. This behavior is known to perturb classical inference procedures tailored for independent observations [59]. A typical statistic summarizing the clustering effect is the *extremal index*. It was introduced initially in [109] and [110], and further popularized in the field of extremes as it corrects inferential procedures of extremes applied to dependent data [59]. We can interpret it as the reciprocate of the mean number of high levels triggered by one first heavy record in a short period [25]. In this article, we focus on cluster statistics inference, i.e., we aim to recover summary statistics of the clustering effect by letting functionals act on consecutive observations with heavy behavior. For example, we can recover the extremal index from this setting and also other important indexes in extremes. The main goal of this article is to study the asymptotic properties of the so-called blocks estimators tailored for cluster inference.

We examine cluster inference for regularly varying time series based on disjoint blocks with the framework from [26]. We review their setting and state sufficient conditions for asymptotic normality of the cluster-based blocks estimators to hold. The novelty of the approach in [26] is twofold. First, they propose to infer cluster statistics using extremal ℓ^p-blocks for $p > 0$, i.e., consecutive observations with large ℓ^p-norm. So far, cluster inference was typically addressed using blocks with at least one exceedance of a large threshold [52, 51, 28, 108]. The strategy of considering blocks exceeding a threshold for the ℓ^p-norm, for $p < \infty$, proved to be advantageous compared to $p = \infty$ regarding bias. Intuitively, as p decreases, the ℓ^p-norms increase yielding to more effective clusters crossing the high threshold for the ℓ^p-norm. Second, [26] also introduce order statistics of the ℓ^p-norms sample, and propose to replace high thresholds with order statistics of the ℓ^p-norms sample.

We prove asymptotic normality of cluster-based blocks extimators relying on the asymptotics of disjoint blocks estimators studied in Theorem 2.10. in [52], and the theory of empirical processes therein. We also follow the modern overview in [108]. In our case, to handle the asymptotics of extremal ℓ^p-blocks we rely on the large deviation principles studied in [26] and appeal to the so called p-cluster processes theory. We state asymptotic normality of the cluster-based blocks extimators in Section 5.2, with random threshold chosen as ℓ^p-norm order statistics. Preliminaries on mixing coefficients, regular variation, and p-cluster processes theory are compiled in Section 5.3. In Section 5.4, we state mixing conditions for consistency and for finite-dimensional asymptotic normality of the cluster-based blocks estimators with deterministic thresholds. Recall this last is a necessary condition to show functional central limit theorems, which we then use to replace the deterministic threshold with random ℓ^p-norm order statistics thresholds. Section 5.5 studies examples of cluster index estimation such as the extremal index and the cluster index for sums. We conclude with verification of the assumptions on classical models such as linear processes and stochastic recurrence equations in Section 5.6.

5.1.1 Notation

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

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

and define the sequential space ℓ^p as

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

with the convention that, for $p = \infty$, the space ℓ^∞ refers to sequences with finite supremum norm. For any $p \in (0, +\infty]$, the p-modulus functions induce a distance d_p in ℓ^p, and for $p \in [1, +\infty)$, it defines a norm. Abusing notation, we call them p-norms for $p \in (0, +\infty]$. Let $\tilde{\ell}^p = \ell^p / \sim$ be the shift-invariant quotient space where: $(\mathbf{x}_t) \sim (\mathbf{y}_t)$ if and only if there exists $k \in \mathbb{Z}$ such that $\mathbf{x}_{t-k} = \mathbf{y}_t$, $t \in \mathbb{Z}$. We also consider the metric space $(\tilde{\ell}^p, \tilde{d}_p)$ such that for $[\mathbf{x}], [\mathbf{y}] \in \tilde{\ell}^p$,

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

and without loss of generality, we write an element $[\mathbf{x}]$ in $\tilde{\ell}^p$ also as $(\mathbf{x}_t)$. Further details on the shift-invariant spaces are deferred to [26, 9].

Furthermore, for $a, b, \in \mathbb{Z}$, and $a \le b$, we write as $\mathbf{x}_{[a,b]}$ the vector $(\mathbf{x}_t)_{t=a,\cdots,b}$ taking values in $(\mathbb{R}^d)^{b-a+1}$. We sometimes write $\mathbf{x}_{[a,b]} \in \tilde{\ell}^p$, which means we take the natural embedding of $\mathbf{x}_{[a,b]}$ in $\tilde{\ell}^p$ defined by assigning zeros to undefined coefficients.

For a stationary time series $(\mathbf{X}_t)$, we write $(\rho_h), (\beta_h)$, for the sequence of ρ- and β-mixing coefficients with respect to past and future $\sigma-$algebras defined by $\mathcal{F}_{t\le 0} := \sigma((\mathbf{X}_t)_{t\le 0})$ and $\mathcal{F}_{t\ge h} := \sigma((\mathbf{X}_t)_{t\ge h})$, respectively. A formal definition of mixing coefficients, and their main properties are reviewed in Section 5.3.

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

$$(\overline{\mathbf{x}}_t^{\varepsilon}) \;=\; (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|\le\varepsilon})\,, \qquad\qquad (\underline{\mathbf{x}_t}_{\epsilon}) \;=\; (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|>\varepsilon})\,,$$

We sometimes write $\mathbf{x}$ for the sequence $\mathbf{x} := (\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$.

5.2 Main Result

5.2.1 Disjoint cluster blocks estimators

Consider $(\mathbf{X}_t)$ to be a stationary time series taking values in $(\mathbb{R}^d, |\cdot|)$. We assume it is regularly varying with tail index $\alpha > 0$, i.e., all its finite-cut distributions are multivariate regularly varying; see Section 5.3.2. For $p \in (0, +\infty]$, we also assume it admits a p-cluster process $\mathbf{Q}^{(p)}$, i.e., for a well chosen sequence (x_n) (see Assumptions Theorem 5.3.4) satisfying

$$\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \;\sim\; n\,c(p)\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0, \quad n \to +\infty, \tag{5.2.3}$$

for $c(p) \in (0, \infty)$, then

$$\mathbb{P}(\mathbf{X}_{[0,n]}/x_n \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_p > x_n) \;\xrightarrow{w}\; \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot\,), \quad n \to +\infty, \tag{5.2.4}$$

where Y is independent of $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., and the limit in (5.2.4) holds in $(\tilde{\ell}^p, \tilde{d}_p)$; see Section 5.3.3.

It will be convenient to write $\mathcal{G}_+(\tilde{\ell}^p)$ for the continuous non-negative functions on $(\tilde{\ell}^p, \tilde{d}_p)$ which vanish in a neighborhood of the origin. We aim to infer statistics of bounded functionals $f \in \mathcal{G}_+(\tilde{\ell}^p)$, acting on $Y\mathbf{Q}^{(p)}$, of the form

$$f_p^{\mathbf{Q}} \;=\; \mathbb{E}[f(Y\mathbf{Q}^{(p)})]\,. \tag{5.2.5}$$

We consider observations $\mathbf{X}_{[1,n]}$, and an intermediate sequence of block length (b_n), such that $b_n \to +\infty$, $n/b_n \to +\infty$, as $n \to +\infty$. We then divide observations into disjoint blocks $(\mathcal{B}_j)_{j=1,\dots,m_n}$, with $m_n := \lfloor n/b_n \rfloor$, such that

$$\underbrace{\mathbf{X}_1, \dots, \mathbf{X}_{b_n}}_{\mathcal{B}_1}, \underbrace{\mathbf{X}_{b_n+1}, \dots, \mathbf{X}_{2b_n}}_{\mathcal{B}_2}, \dots, \underbrace{\mathbf{X}_{(m_n-1)b_n+1}, \dots, \mathbf{X}_{m_n b_n}}_{\mathcal{B}_{m_n}}\,.$$

Following [26], we estimate $f_p^{\mathbf{Q}}$ in (5.2.5) using disjoint blocks estimators, that we denote $\widehat{f_p^{\mathbf{Q}}}$, defined as

$$\widehat{f_p^{\mathbf{Q}}} \quad := \quad \frac{1}{k}\sum_{t=1}^{m_n} f(\mathcal{B}_t/\|\mathcal{B}_t\|_{p,(k)})\mathbb{1}(\|\mathcal{B}_t\|_p \geq \|\mathcal{B}_t\|_{p,(k)}), \tag{5.2.6}$$

where

$$\|\mathcal{B}_t\|_{p,(1)} \geq \|\mathcal{B}_t\|_{p,(2)} \geq \cdots \geq \|\mathcal{B}_t\|_{p,(m_n)},$$

denotes the sequence of order statistics of the ℓ^p-norms, and (k_n) is a sequence satisfying $k = k_n \to +\infty$,

$$k \;=\; \lfloor m_n \mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n}) \rfloor \;\sim\; n\, c(p)\mathbb{P}(|\mathbf{X}_0| > x_b), \quad n \to +\infty, \tag{5.2.7}$$

where $c(p) \in (0,\infty)$ is as in (5.2.3). Also, we write the deterministic threshold estimator $\widetilde{f_p^{\mathbf{Q}}} := \widetilde{f_p^{\mathbf{Q}}}(1)$ by

$$\widetilde{f_p^{\mathbf{Q}}}(u) \quad := \quad \frac{1}{k}\sum_{t=1}^{m_n} f(\mathcal{B}_t/u\, x_{b_n})\mathbb{1}(\|\mathcal{B}_t\|_p > u\, x_{b_n})\,, \quad u > 0. \tag{5.2.8}$$

5.2.2 Main result

Asymptotic normality of the estimators in (5.2.6) is established in Theorem 5.2.1 below. We defer to proof to Section 5.7.1.

Theorem 5.2.1. *Let $(\mathbf{X}_t)$ be an $\mathbb{R}^d$-valued regularly varying time series with tail index $\alpha > 0$ admitting a p-cluster process $\mathbf{Q}^{(p)}$. Let (x_n), (b_n), be well chosen sequences such that (5.2.4) holds and $n/b_n \to +\infty$, as $n \to +\infty$. We fix a bounded function $f \in \mathcal{G}_+(\tilde{\ell}^p)$ and assume $u \mapsto f((\mathbf{x}_t)/u)$ is non-increasing for all $(\mathbf{x}_t) \in \tilde{\ell}^p, u > 0$. Assume also $(1), (2), (3)$, below hold. Then,*

$$\sqrt{k}\big(\,\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}\,\big) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mathrm{Var}(f(Y\mathbf{Q}^{(p)}))\,)\,, \quad n \to +\infty. \tag{5.2.9}$$

where Y is independent of $\mathbf{Q}^{(p)}$ and $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$.

(1) *Assume the mixing condition:*

$$m_n\beta_{b_n} \to 0, \quad n \to +\infty. \tag{5.2.10}$$

(2) *Assume also for $\epsilon > 0$, the bias conditions (5.2.11), (5.2.12), below hold*

$$0 \quad = \quad \lim_{n\to+\infty} \sqrt{k} \sup_{u\in[1-\varepsilon,1+\varepsilon]} \Big|\, \frac{\mathbb{E}[f(\mathcal{B}_1/u\,x_{b_n})\mathbb{1}(\|\mathcal{B}_1/x_{b_n}\|_p > u)]}{\mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n})} - u^{-\alpha} f_p^{\mathbf{Q}}\,\Big|\,, \tag{5.2.11}$$

$$0 \quad = \quad \lim_{n\to+\infty} \sqrt{k} \sup_{u\in[1-\varepsilon,1+\varepsilon]} \Big|\, \frac{\mathbb{P}(\|\mathcal{B}_1/x_{b_n}\|_p > u)}{\mathbb{P}(\|\mathcal{B}_1\|_p > x_{b_n})} - u^{-\alpha}\,\Big|\,. \tag{5.2.12}$$

(3) *Let $\mathcal{F} \subseteq \mathcal{G}_+(\tilde{\ell}^p)$ be the set with two functions: $(\mathbf{x}_t) \mapsto f(\mathbf{x}_t)$ and $(\mathbf{x}_t) \mapsto 1$. Consider the family of deterministic threshold estimators defined by*

$$\mathcal{T} \quad = \quad \{\widetilde{g_p^{\mathbf{Q}}}(u)\}_{\{g(\cdot/u):\, u\in[1-\epsilon,1+\epsilon], g\in\mathcal{F}\}}. \tag{5.2.13}$$

Assume asymptotic normality of the finite-dimensional parts of $\mathcal{T}$ holds (see Proposition 5.4.4), i.e.,

$$\sqrt{k}\big(\widetilde{g_p^{\mathbf{Q}}}(u) - u^{-\alpha}\, g_p^{\mathbf{Q}}\big) \xrightarrow{fi.di.} \mathbb{G}(g(\cdot/u)), \quad \widetilde{g_p^{\mathbf{Q}}}(u) \in \mathcal{T}, \tag{5.2.14}$$

as $n \to +\infty$, such that $\mathbb{G}$ is a centered Gaussian process satisfying

$$\mathrm{Cov}\big(\mathbb{G}(g(\cdot/u)), \mathbb{G}(h(\cdot/v))\big) \quad = \quad \int_{u\vee v}^{\infty} \mathbb{E}[g(y\mathbf{Q}^{(p)}/u)h(y\mathbf{Q}^{(p)}/v)]d(-y^{-\alpha}), \tag{5.2.15}$$

for $u, v \in [1-\epsilon, 1+\epsilon]$, and $g, h \in \mathcal{F}$.

In particular, if $(1), (2), (3)$, hold, then uniform asymptotic normality of the family $\mathcal{T}$ holds.

Asymptotic normality of finite-dimensional parts of $\mathcal{T}$ as in (5.2.14) is analyzed in depth in Section 5.4 under further mixing assumptions. We highlight from Theorem 5.2.1 the simple writing of the covariance structure thanks to the theory of p-cluster processes in [26]. We give a couple of remarks below.

Remark 5.2.2. *In Theorem 5.2.1 we assume $f \in \mathcal{G}_+(\tilde{\ell}^p)$ satisfies that $u \mapsto f((\mathbf{x}_t)/u)$ is a non-increasing function, for all $(\mathbf{x}_t) \in \tilde{\ell}^p$. In this case, $(f(\cdot\,/u))_{u\in[1-\epsilon,1+\epsilon]}$ is a linearly ordered family and thus it is a VC-class; see Remark 2.11 in [52]. This monotonicity condition is always verified by the indicator functions $u \mapsto \mathbb{1}(\|\mathbf{x}_t\|_p > u)$, for $p > 0$. It is also the case for functions projecting into the $\tilde{\ell}^p$ sphere as the cluster indexes studied in [26] (see Examples 5.5.1, 5.5.2).*

Remark 5.2.3. *In Theorem 5.2.1 we assume the mixing condition $m_n\beta_{b_n} \to 0$, as $n \to +\infty$, to show equicontinuity of the family $\mathcal{T}$ in (5.2.13). However, to show (5.2.14) holds with limit Gaussian process $\mathbb{G}$ and covariance (5.2.15), we require further mixing conditions. This discussion is deferred to Section 5.4.*

Remark 5.2.4. *The bias assumptions stated in (2) in Theorem 5.2.1 tells us we could neglect the bias by choosing the block lengths b_n sufficiently large. Indeed, recall from (5.2.7), $k \sim n\, c(p)\mathbb{P}(|\mathbf{X}_0| > x_b)$, as $n \to +\infty$, thus we see from (5.2.11), (5.2.12), that the bias term can be controlled using sufficiently large block lengths b_n for inference.*

Remark 5.2.5. *We can consider unbounded functionals $f \in \mathcal{G}_+(\tilde{\ell}^p)$ in Theorem 5.2.1 by imposing a Lindeberg assumption as (C.4.3) in [108], and assuming $\mathrm{Var}(\mathbb{G}(f(\cdot/1))) < +\infty$, following the notation from Equation (5.2.15). Recall the relationship between p-clusters established in Proposition 3.1. in [26]; see (5.3.22). We argue that for inferring $f_p^{\mathbf{Q}}$ we can first look for suitable choices of pairs f', p' such that f' is bounded and the limit $f_{p'}'^{\mathbf{Q}}$ coincides with $f_p^{\mathbf{Q}}$.*

5.3 Preliminaries

5.3.1 Mixing coefficients

Properties of stationary sequences are usually studied through mixing coefficients that we review below. For a summary on mixing conditions see [14, 47, 144].

Definition 5.3.1. *Let $(\mathbf{X}_t)$ be an $\mathbb{R}^d$-valued strictly stationary time series defined over a probability space $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P})$. Denote the past and future $\sigma-$algebras by $\mathcal{F}_{t\leq 0} := \sigma((\mathbf{X}_t)_{t\leq 0})$ and $\mathcal{F}_{t\geq h} := \sigma((\mathbf{X}_t)_{t\geq h})$, respectively. We recall the definition of mixing coefficients $(\rho_h), (\beta_h)$, below*

$$\begin{aligned}
\rho_h &:= \sup_{f\in L^2(\mathcal{F}_{t\leq 0}),\, g\in L^2(\mathcal{F}_{t\geq h})} |\mathrm{Corr}(f,g)|, \\
\beta_h &:= \frac{1}{2}\sup_{\mathcal{A},\mathcal{B}} \sum_{i\in I}\sum_{j\in J} |\mathbb{P}(A_i\cap \mathcal{B}_j) - \mathbb{P}(A_i)\mathbb{P}(\mathcal{B}_j)|, \\
&= d_{TV}\big(\, \mathbb{P}_{\mathcal{F}_{t\leq 0}\otimes \mathcal{F}_{t\geq h}}\, , \, \mathbb{P}_{\mathcal{F}_{t\leq 0}} \otimes \mathbb{P}_{\mathcal{F}_{t\geq h}}\, \big),
\end{aligned}$$

where $\mathcal{A} = \{A_i, i \in I\}, \mathcal{B} = \{\mathcal{B}_j : j \in J\}$ are finite partitions of the space, measurable with respect to $\mathcal{F}_{t\leq 0}$ and $\mathcal{F}_{t\geq h}$, respectively, and d_{TV} is the total variation distance of probability measures. Also, for any two probability spaces $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P}_1)$, $((\mathbb{R}^d)^{\mathbb{Z}}, \mathcal{A}, \mathbb{P}_2)$, we write $\mathbb{P}_1\otimes\mathbb{P}_2(A\times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$, for $A, B \in \mathcal{A}$.

For the detailed interpretation of the coefficients (β_h) in terms of the total variation distance we refer to Chapter 1.2 in [47].

Remark 5.3.2. *The β-mixing coefficients $(\beta_t)_{t\geq 0}$ are well adapted while working with Markov processes. Indeed, a strictly stationary Markov Chain $(\mathbf{X}_t)$, which is Harris recurrent verifies $\beta_h \to 0$ as $h \to +\infty$; see Theorem 3.5 in [14]. Further assumptions as Doeblin's conditions also guarantee $\beta_h \to 0$ as $h \to +\infty$ at least exponentially fast; see Theorem 3.7. in [14] and references herein.*

Remark 5.3.3. *The ρ-mixing coefficients $(\rho_t)_{t\geq 0}$ were introduced in [107] and popularized due to the Ibragimov central limit theorem for dependent stationary sequences in [95]. Mainly, it states that the mixing condition: $\sum_{t=1}^{\infty} \rho_{2^t} < +\infty$, together with a moment assumption of order $2+\delta$, are sufficient conditions for the central limit of stationary sequences to hold. The aforementioned mixing condition was studied in detail in [15, 136, 156, 169], as it is shown to be a sharp condition entailing the central limit theorem; see [14] for a review. However, aside from the Gaussian case where $\rho_h \leq \pi\beta_h$, there is not a general recipe for computing ρ–mixing rates.*

5.3.2 Regular variation

We consider stationary time series $(\mathbf{X}_t)$ taking values in $(\mathbb{R}^d, |\cdot|)$. We call it regularly varying with index $\alpha > 0$ if all finite dimensional vectors of the time series are regularly varying with index $\alpha > 0$. Borrowing the approach in [10], this holds if and only if, for all $h \geq 0$, there exists a vector

$(\Theta_t)_{|t|\le h}$, taking values in $(\mathbb{R}^d)^{2h+1}$ such that

$$\mathbb{P}(x^{-1}(\mathbf{X}_t)_{|t|\le h} \in \cdot \,|\, |\mathbf{X}_0| > x) \;\xrightarrow{d}\; \mathbb{P}(Y(\Theta_t)_{|t|\le h} \in \cdot), \quad x \to +\infty, \tag{5.3.16}$$

where Y is independent of $(\Theta_t)_{|t|\le h}$ and $\mathbb{P}(Y > y) = y^{-\alpha}, y > 1$. We call the sequence (Θ_t), taking values in $(\mathbb{R}^d)^{\mathbb{Z}}$, the *spectral tail process*, and its existence is granted by Kolmogorov's consistency theorem.

The time series (Θ_t) does not inherit the initial stationarity property. Instead, the time-change formula in [10] holds: for any $s, t \in \mathbb{Z}, s \le 0 \le t$ and for any measurable bounded function $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$,

$$\mathbb{E}[f(\Theta_{s-i}, \ldots, \Theta_{t-i})\mathbb{1}(|\Theta_{-i}| \neq 0)] \;=\; \mathbb{E}[|\Theta_i|^\alpha f(\Theta_s/|\Theta_i|, \ldots, \Theta_t/|\Theta_i|)]. \tag{5.3.17}$$

5.3.3 p-cluster processes

let $(\mathbf{X}_t)$ be a stationary regularly varying time series with tail index $\alpha > 0$, taking values in $(\mathbb{R}^d, |\cdot|)$. Recall, for $p > 0$, we say $(\mathbf{X}_t)$ admits a p-cluster process $\mathbf{Q}^{(p)} \in \ell^p$ if there exists a well chosen sequence (x_n), satisfying

$$\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \sim n\, c(p)\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0, \quad n \to \infty, \tag{5.3.18}$$

with $c(p) \in (0, \infty)$ and

$$\mathbb{P}(\mathbf{X}_{[0,n]}/x_n \in \cdot \,|\, \|\mathbf{X}_{[0,n]}\|_p > x_n) \;\xrightarrow{w}\; \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot), \quad n \to +\infty, \tag{5.3.19}$$

where Y is independent of $\mathbf{Q}^{(p)} \in \tilde{\ell}^p$, $\mathbb{P}(Y > y) = y^{-\alpha}, y > 1$, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., and the limit in (5.3.19) holds in $(\tilde{\ell}^p, \tilde{d}_p)$.

The ∞-cluster process was introduced implicetly in [40]; see [9]. The p-cluster processes, for $p < \infty$, were introduced in [26]. Anti-clustering and vanishing-small values conditions **AC**, $\mathbf{CS}_p$, respectively, guarantee the existence of $p-$clusters. We state below Theorem 2.1. in [26].

Proposition 5.3.4. *Let $(\mathbf{X}_t)$ be a stationary regularly varying time series with tail index $\alpha > 0$. Assume (x_n) satisfies $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, for some $\kappa > 0$ and for any $\delta, \epsilon > 0$, and*

AC *:* $\lim_{s\to+\infty} \lim_{n\to+\infty} \sum_{t=s}^{n} \mathbb{P}(|\mathbf{X}_t| > \epsilon\, x_n \,|\, |\mathbf{X}_0| > \epsilon\, x_n) \;=\; 0.$

$\mathbf{CS}_p$*:* $\lim_{\epsilon\downarrow 0} \limsup_{n\to+\infty} \frac{\mathbb{P}\left(\|\overline{\mathbf{X}_{[1,n]}/x_n}^{\epsilon}\|_p^p > \delta\right)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} \;=\; 0.$

For $p \ge \alpha$, the function $p \mapsto c(p)$ as in (5.3.18) is non-increasing, $c(\alpha) = 1$, and $(\mathbf{X}_t)$ admits a cluster process $\mathbf{Q}^{(p)}$ in the sense of (5.3.19). For $p < \alpha$, existence of the p-cluster process holds if there exists $c(p) < \infty$ satisfying (5.3.18).

Under **AC**, $\mathbf{CS}_\alpha$, the time series $(\mathbf{X}_t)$ admits an α-cluster $\mathbf{Q}^{(\alpha)}$, for $\alpha > 0$, the (tail) index (see

Proposition 5.3.4). In this case, appealing to Proposition 3.1. in [26] we have

$$\mathbf{Q}^{(\alpha)} \quad \stackrel{d}{=} \quad \Theta/\|\Theta\|_\alpha, \quad \in \tilde{\ell}^\alpha. \tag{5.3.20}$$

where (Θ_t) is the spectral tail process; see (5.3.16). Furthermore, if we assume in addition that $\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] < +\infty$, then Equation (5.3.18) holds with $c(p) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] < +\infty$. Sufficient conditions on $(\mathbf{X}_t)$ such that $\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] < +\infty$ are given in Section 5.7.4.

Moreover, assume $(\mathbf{X}_t)$ satisfies the conditions of Proposition 5.3.4, for $\alpha, p, p' > 0$, where $\alpha > 0$ is the (tail) index. Assume also $\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] + \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_{p'}^\alpha] < +\infty$, then the p, p'-clusters exist and are related by the change of norms formula below (see Proposition 3.1. in [26])

$$\begin{aligned}
&\mathbb{P}(\mathbf{Q}^{(p)} \in \cdot) \\
&= \quad c(p)^{-1}\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha \mathbb{1}(\mathbf{Q}^{(\alpha)}/\|\mathbf{Q}^{(\alpha)}\|_p \in \cdot)] &(5.3.21)\\
&= \quad \frac{c(p')}{c(p)}\mathbb{E}[\|\mathbf{Q}^{(p')}\|_p^\alpha \mathbb{1}(\mathbf{Q}^{(p')}/\|\mathbf{Q}^{(p')}\|_p \in \cdot)] &(5.3.22)
\end{aligned}$$

where $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. for any $p > 0$ and $\mathbb{E}[\|\mathbf{Q}^{(p')}\|_p^\alpha] = c(p)/c(p')$, where $c(p), c(p')$ are as in Equation (5.3.18).

Remark 5.3.5. *Condition* **AC** *is satisfied for m_0–dependent sequences for sequences (x_n) satisfying $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$, as $n \to +\infty$.*

Remark 5.3.6. *Condition $\mathbf{CS}_p$ is always verified for $p > \alpha$ for sequences (x_n) satisfying $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$, as $n \to +\infty$ (see Remark 5.1. in [26]). Also, using the monotonicty of norms, we see straightforwardly that $\mathbf{CS}_p$ implies condition $\mathbf{CS}_{p'}$, for $p, p' > 0$ and $p < p'$. Let $\alpha > 0$ be the (tail) index. Then, for $p \geq \alpha/2$, if $\sum_{h=0}^\infty \rho_h < +\infty$, then condition $\mathbf{CS}_p$ holds for sequences (x_n) satisfying $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, as $n \to +\infty$, for any $\kappa > 0$ (see Remark 5.2. in [26]). In particular, this means that $\mathbf{CS}_p$ holds for m_0–dependent sequences, for $p \geq \alpha/2$, for sequences (x_n) as before.*

5.4 Central limit theory

We complement the discussion on Theorem 5.2.1 in this section. In what follows, we will work under the assumptions fixed in Section 5.4.1. In Section 5.4.2 we give mixing rates for consistency of the p-cluster based blocks estimators in (5.2.6). Section 5.4.3 states sufficient assumptions for asymptotic normality of finite-dimensional parts of the family of deterministic threshold estimators $\mathcal{T}$ defined in (5.2.13); see (5.2.14).

5.4.1 Assumptions

We consider $(\mathbf{X}_t)$ to be an $\mathbb{R}^d$-valued stationary regularly varying time series of index $\alpha > 0$. We assume $(\mathbf{X}_t)$ satisfies **AC**, $\mathbf{CS}_p$, for a sequence (x_n), and that it admits a $p-$cluster process $\mathbf{Q}^{(p)} \in \ell^p$, $p > 0$ such that (5.3.19) holds for this sequence of thresholds (x_n). We fix also the sequence of block lengths $b = (b_n)$ defining the p-cluster based blocks estimators in (5.2.6). We write $m = m_n = \lfloor n/b_n \rfloor$ and $k = (k_n)$ the sequence defined in (5.2.7). We recall the family of deterministic threshold estimators, that we denote $\mathcal{T}$, defined in (5.2.13).

5.4.2 Consistency of p-cluster based blocks estimators

Recall [26] show consistency of the p-cluster based blocks estimators in (5.2.6) under the mixing condition that we provide below in Equation (5.4.23). Lemma 5.4.1 gives mixing rates for this condition to hold. The proof is left for Section 5.7.2.

Lemma 5.4.1. *Consider a stationary time series $(\mathbf{X}_t)$ satisfying the assumptions in Section 5.4.1. Let $f \in \mathcal{G}_+(\tilde{\ell}^p)$ be a bounded Lipschitz continuous function. If either 1. or 2. below hold, then the sequences (x_b) and (b_n) satisfy the relation*

$$\begin{aligned}\big|\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^{m} f(x_b^{-1}\mathcal{B}_t)\big\} - \mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\sum_{t=1}^{\lfloor m/k\rfloor} f(x_b^{-1}\mathcal{B}_t)\big\}\big]^k\big]\big| &\to 0,\\ n &\to +\infty, \qquad (5.4.23)\end{aligned}$$

where $k = k_n(b) = \lfloor m_n \mathbb{P}(\|\mathcal{B}_t\|_p > x_b)\rfloor$ is as in (5.2.7).

1. *The correlation coefficients (ρ_t) satisfy*

$$\lim_{n\to+\infty} \textstyle\sum_{t=0}^{\lfloor \log_2(n)\rfloor} \rho_{b_n 2^t}/k_n \;=\; 0.$$

2. *There exists a sequence (ℓ_n), satisfying $\ell_n \to +\infty$, as $n \to +\infty$, such that the β-mixing coefficients satisfy*

$$\lim_{n\to+\infty} m_n \beta_{\ell_n}/k_n \;=\; \lim_{n\to+\infty} \ell_n/b_n \;=\; 0.$$

Corollary 5.4.2. *Consider the assumptions in Lemma 5.4.1, and assume either condition 1. or 2. therein holds. Then for any bounded function $f \in \mathcal{G}_+(\tilde{\ell}^p)$, the p-cluster based blocks estimator in (5.2.6) satisfies*

$$\widehat{f_p^{\mathbf{Q}}} \xrightarrow{\mathbb{P}} f_p^{\mathbf{Q}}, \quad n \to +\infty.$$

Proof. This follows from Lemma 5.4.1 and applying Theorem 4.1 in [26]. □

5.4.3 Covariance p-cluster based blocks estimators

We study now a covariance structure for the deterministic threshold estimators in (5.2.8). We defer the proof of the next Lemma to Section 5.7.3.

Lemma 5.4.3. *Consider a stationary time series $(\mathbf{X}_t)$ satisfying the assumptions in Section 5.4.1. Let $g, h \in \mathcal{G}_+(\tilde{\ell}^p)$ be bounded functions and recall the deterministic threshold estimators in (5.2.8).*

Assume either condition 1. or 2. below holds, then

$$k\,\mathrm{Cov}\big(\,\widetilde{g_p^{\mathbf{Q}}}(u),\widetilde{h_p^{\mathbf{Q}}}(v)\,\big)\quad\to\quad\mathrm{Cov}\big(\mathbb{G}(g(\cdot/u)),\mathbb{G}(h(\cdot/v))\big),\quad n\to+\infty,$$

and $\mathbb{G}$ is a centered Gaussian process with covariance structure as in (5.2.15).

1. *The correlation coefficients (ρ_t) satisfy $\sum_{t=1}^{m_n}\rho_{tb_n}\to 0$, as $n\to+\infty$, and for all $t>0$, $\epsilon>0$, the additional anti-clustering assumptions holds as well:*

$$\lim_{s\to+\infty}\lim_{n\to+\infty}\textstyle\sum_{i=(t-1)b_n+s}^{tb_n}\mathbb{P}(|\mathbf{X}_t|>\epsilon\,x_{b_n}\,|\,|\mathbf{X}_0|>\epsilon\,x_{b_n})\;\;=\;\;0. \tag{5.4.24}$$

2. *There exists a sequence (ℓ_n), satisfying $\ell_n\to+\infty$, as $n\to+\infty$, and the β-mixing coefficients satisfy*

$$\lim_{n\to+\infty}\tfrac{m_n}{k_n}\textstyle\sum_{t=1}^{m_n}\beta_{\ell_n+(t-1)b_n}\;\;=\;\;\lim_{n\to+\infty}\ell_n/b_n\;\;=\;\;0. \tag{5.4.25}$$

Proposition 5.4.4. *Consider the assumptions in Lemma 5.4.3, and assume either 1. or 2. therein holds. Consider a bounded function $f\in\mathcal{G}_+(\tilde{\ell}^p)$ satisfying the bias assumptions (5.2.11), (5.2.12), from Theorem 5.2.1. Recall the family $\mathcal{T}$ in (5.2.13), then asymptotic normality of the finite-dimensional parts of $\mathcal{T}$ holds; see (5.2.14), with Gaussian limit $\mathbb{G}$ and covariance structure as in (5.2.15).*

Proof. This follows straightforwardly by the Cramér-Wold device since every linear combination of deterministic threshold estimators in (5.2.8) is again a deterministic threshold estimator. □

5.5 Examples of cluster inference

In view of Theorem 5.2.1, we can derive asymptotic normality of classical cluster index statistics in extreme value theory. In this section we work in the framework of Section 5.4.1.

Example 5.5.1. *(Extremal index) Assume the conditions of Proposition 5.3.4 and the conditions of Theorem 5.2.1 hold for $p=\alpha$. Consider f to be the function $\mathbf{x}\mapsto\|\mathbf{x}\|_\infty^\alpha/\|\mathbf{x}\|_\alpha^\alpha$, let $\theta_{|\mathbf{X}|}=\mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha]$, hence one deduces an estimator*

$$\widehat{\theta}_{|\mathbf{X}|}\quad=\quad\frac{1}{k}\sum_{t=1}^{m}\frac{\|\mathcal{B}_t\|_\infty^\alpha}{\|\mathcal{B}_t\|_\alpha^\alpha}\mathbb{1}(\|\mathcal{B}_t\|_\alpha\geq\|\mathcal{B}_t\|_{\alpha,(k)}), \tag{5.5.26}$$

such that

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

where $\theta_{|\mathbf{X}|}$ is the candidate for the extremal index. Recall, in [28] the asymptotic variance for the blocks estimator of the extremal was computed and equals σ^2 such that

$$\begin{aligned}\sigma^2 &= \theta_{|\mathbf{X}|}^2 \sum_{j\in\mathbb{Z}} \mathbb{E}[|\Theta_j|^\alpha \wedge 1] - \theta_{|\mathbf{X}|} \\ &= \theta_{|\mathbf{X}|}^2 \sum_{j\in\mathbb{Z}} \sum_{t\in\mathbb{Z}} \mathbb{E}[|\mathbf{Q}_{j+t}^{(\alpha)}|^\alpha \wedge |\mathbf{Q}_t^{(\alpha)}|^\alpha] - \theta_{|\mathbf{X}|},\end{aligned} \tag{5.5.27}$$

where the last equality follows appealing to the time-change formula in (5.3.17) and Equation (5.3.20).

Furthermore, we recall the cluster index for sums from Section 4.3 in [26]. Assume the conditions of Proposition 5.3.4 hold for $p = 1$ thus $(\mathbf{X}_t)$ admits a 1-cluster process $\mathbf{Q}^{(1)}$ and $1 \le \alpha$. Recall from (2.7.7) that $c(1)$ in (5.3.18) satisfies $c(1) = 1/\mathbb{E}[\|\mathbf{Q}^{(1)}\|_\alpha^\alpha]$; see (5.2.3). We show asymptotic normality for the disjoint blocks estimator of $c(1)$ appealing to the Delta method.

Example 5.5.2. *(Cluster index for sums) Assume the conditions of Proposition 5.3.4 and the conditions of Theorem 5.2.1 hold for $p = 1$. Consider f to be the function $\mathbf{x} \mapsto \|\mathbf{x}\|_\alpha^\alpha / \|\mathbf{x}\|_1^\alpha$, for $1 < \alpha$, and let $c(1) = 1/\mathbb{E}[\|\mathbf{Q}^{(1)}\|_\alpha^\alpha] < +\infty$. Hence one deduces an estimator*

$$\widehat{c}(1) = \Big(\frac{1}{k}\sum_{t=1}^{m} \frac{\|\mathcal{B}_t\|_\alpha^\alpha}{\|\mathcal{B}_t\|_1^\alpha} \mathbb{1}(\|\mathcal{B}_t\|_1 \ge \|\mathcal{B}_t\|_{1,(k)})\Big)^{-1}, \tag{5.5.28}$$

then an application of the Delta method yields

$$\sqrt{k}\big(\,\widehat{c}(1) - c(1)\,\big) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(\|\mathbf{Q}^{(1)}\|_\alpha^\alpha))/c(1)^2\,, \quad n \to +\infty.$$

5.6 Classical models

5.6.1 m_0–dependent linear sequences

Consider $(\mathbf{X}_t)$ to be a regularly varying time series of index $\alpha > 0$ that is m_0–dependent linear sequence.

Example 5.6.1. *Let $(\mathbf{X}_t)$ be a moving average of order $m_0 \ge 1$ defined as:*

$$\mathbf{X}_t := \mathbf{Z}_t + \varphi_1 \mathbf{Z}_{t-1} + \cdots + \varphi_{m_0} \mathbf{Z}_{t-m_0}, \quad t \in \mathbb{Z}. \tag{5.6.29}$$

with iid regularly varying innovations $(\mathbf{Z}_t)$ of index $\alpha > 0$ takinf values in $\mathbb{R}^d$. Then, by the convolution closure of regularly varying vectors, we deduce regular variation for the strictly stationary sequence $(\mathbf{X}_t)$.

Proposition 5.6.2. *Let $(\mathbf{X}_t)$ be a m_0–dependent sequence as defined in (5.6.29), regularly varying of index $\alpha > 0$. Then, for any $p > 0$, for any bounded function $f \in \mathcal{G}_+(\tilde{\ell}^p)$ such that $u \mapsto f(\cdot/u)$ is non-increasing,*

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

Corollary 5.6.3. *For m_0-dependent time series as defined in (5.6.29), the $\alpha-$cluster based estimator of the extremal index and the $1-$cluster based estimator of the cluster index of sums in (5.5.26) and (5.5.28), respectively, are asymptotically normally distributed and super efficient, i.e. its asymptotic variance is null.*

Proof Proposition 5.6.2. Conditions $\mathbf{AC}, \mathbf{CS}_p$, hold by the arguments in Remark 5.3.5 and Remark 5.3.6. Also, by Proposition 3.1. in [26], $\|\Theta\|_\alpha < +\infty$ a.s. and $\mathbf{Q}^{(\alpha)} \stackrel{d}{=} \Theta/\|\Theta\|_\alpha$. Let (φ_j) be as in (5.6.29) and $\Theta_0^{\mathbf{Z}}$ be the spectral measure of $\mathbf{Z}_0$ such that $|\Theta_0^{\mathbf{Z}}| = 1$ a.s.. Then,

$$\mathbf{Q}^{(\alpha)} \quad = \quad (\varphi_j/\|\varphi_j\|_\alpha)\,\Theta_0^{\mathbf{Z}}, \quad \in \tilde{\ell}^\alpha \tag{5.6.30}$$

and $c(p) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] < +\infty$. Then, all assumptions in Section 5.4.1 are satisfied. Moreover, m_0 dependent sequence have mixing coefficients $(\beta_h), (\rho_h)$ equal to zero for $h > m_0$. Then all assumptions from Theorem 5.2.1 and Lemma 5.4.3 hold and this concludes the proof. □

Proof Corollary 5.6.3. Notice that if $p = \alpha$, the expression in (5.6.30) yields to a deterministic expression of $|\mathbf{Q}^{(\alpha)}|$ in the shift-invariant space. Then, the index estimators in (5.5.26) and (5.5.28) with $f : \mathbf{x} \mapsto \|\mathbf{x}\|_\infty^\alpha/\|\mathbf{x}\|_\alpha^\alpha$ and $f : \mathbf{x} \mapsto \|\mathbf{x}\|_\alpha^\alpha/\|\mathbf{x}\|_1^\alpha$, respectively, satisfy $\mathrm{Var}(f(Y\mathbf{Q}^{(\alpha)})) = 0$. Finally, using the change of norms formula in (5.3.21) we can also show $\mathrm{Var}(f(Y\mathbf{Q}^{(p)})) = 0$, for any $p > 0$, and this concludes the proof. □

5.6.2 Linear processes

We consider linear processes $(\mathbf{X}_t)$ satisfying the equation

$$\mathbf{X}_t \quad = \quad \sum_{t\in\mathbb{Z}} \varphi_j \mathbf{Z}_{t-j}, \quad t \in \mathbb{Z}, \tag{5.6.31}$$

for a sequence of iid innovations (Z_t), regularly varying of index $\alpha > 0$. We assume the sequence (φ_j) satisfies $\|\varphi_j\|_{\alpha-\delta} < +\infty$ for some $\delta > 0$. Then a stationary solution $(\mathbf{X}_t)$ exists [31]; see and [120] for sharper conditions guaranteeing existence. Moreover,

$$\frac{\mathbb{P}(|\mathbf{X}_0| > x_n)}{\mathbb{P}(|\mathbf{Z}_0| > x_n)} \quad \to \quad \sum_{t\in\mathbb{Z}} |\varphi_j|^\alpha < +\infty, \quad n \to +\infty \tag{5.6.32}$$

This relation is also shown in [31]. Examples of the linear model are stationary ARMA processes with iid regularly varying noise (Z_t) (cf. [18]). Moreover, if $\Theta_0^{\mathbf{Z}}$ is the spectral measure of $\mathbf{Z}$ such that $|\Theta_0^{\mathbf{Z}}| = 1$ a.s., then

$$\mathbf{Q}^{(\alpha)} \quad = \quad (\varphi_j/\|\varphi_j\|_\alpha)\,\Theta_0^{\mathbf{Z}}, \quad \in \tilde{\ell}^\alpha \tag{5.6.33}$$

thus $\mathbf{Q}^{(\alpha)}$ admits a deterministic expression when projected to the shift-invariant space $\tilde{\ell}^\alpha$. In this setting, we also have $c(p) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_p^\alpha] = \|\varphi_j\|_p^\alpha/\|\varphi_j\|_\alpha^\alpha$ which is finite if $\|\varphi_j\|_p < \infty$. In what follows the linear processes we considered satisfy the aforementioned conditions.

Proposition 5.6.4. *Let $(\mathbf{X}_t)$ be the linear process as in (5.6.31). Let $\alpha/2 < p < \alpha$ and (x_n) be a sequence satisfying $n/x_n^{p\wedge(\alpha-\kappa)} \to 0$, for any $\kappa > 0$. If $\|\varphi\|_{p\wedge(\alpha-\kappa)} < +\infty$, then for all $\delta > 0$,*

$$\lim_{s\to+\infty} \limsup_{n\to+\infty} \frac{\mathbb{P}(\|\mathbf{X}_{[0,n]}/x_n - \mathbf{X}^{(s)}_{[0,n]}/x_n\|_p^p > \delta)}{n\mathbb{P}(|\mathbf{X}_0| > x)} \quad = \quad 0. \tag{5.6.34}$$

*where $\mathbf{X}_t^{(s)} := \sum_{|j|\le s} \varphi_j \mathbf{Z}_{t-j}$. Moreover, **AC** and $\mathbf{CS}_p$ are verified, and for any $f \in \mathcal{G}_+(\tilde{\ell}^p)$, bounded continuous*

$$\widehat{f_p^{\mathbf{Q}}} \;\xrightarrow{\mathbb{P}}\; f_p^{\mathbf{Q}}, \quad n \to +\infty.$$

Moreover, assume the β-mixing conditions (5.2.10), (5.4.25), and the bias condition (5.2.11), (5.2.12), then for any $f \in \mathcal{G}_+(\tilde{\ell}^p)$, such that $u \mapsto f(\cdot/u)$ is non-increasing,

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

Remark 5.6.5. *Computation of β-mixing bound rates for linear models are available in the literature. Typically, additional conditions on the sequence (φ_t) are assumed; see for example Lemma 15.3.1. in [108].*

Corollary 5.6.6. *For linear models in (5.6.31), verifying the assumptions of Proposition 5.6.4, the estimators for the extremal index and the cluster index of sums in (5.5.26) and (5.5.28), respectively, are asymptotically normally distributed and super efficient, i.e. its asymptotic variance is null.*

Remark 5.6.7. *For the extremal index, we can compare the asymptotic variance of the $\alpha-$cluster based estimator against the asymptotic variance σ^2 for the blocks estimator in (5.5.27). For example, for autoregressive process of order one $AR(1)$, $\sigma^2 = 1 - \theta_{|\mathbf{X}|} \ge 0$, thus the α-cluster based estimator has a an optimal asymptotic variance. The main drawback of the α-cluster based estimator is that α also needs to be estimated. We could use a consistent Hill estimator of $1/\alpha$; see [87, 141], but this subject requires further investigation.*

Proof Corollary 5.6.6. Under the assumptions of Proposition 5.6.4 asymptotic normality holds. Then, as if the steps of the proof of Corollary 5.6.3 we can compute the asymptotic variance using the expression in (5.6.33). For the extremal and cluster index, both variance computation are zero and this concludes the proof. □

5.6.3 Solutions to affine stochastic recurrence equations under Kesten's conditions

We study the causal solution to the equation:

$$\mathbf{X}_t \quad = \quad \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{B}_t, \qquad t \in \mathbb{Z}, \tag{5.6.35}$$

where $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$ is an iid sequence of non-negative random $d \times d$ matrices with generic element $\mathbf{A}$, and random vectors with generic element $\mathbf{B}$, taking values in $\mathbb{R}^d$.

The celebrated paper of Kesten [103] demonstrates sufficient conditions for the existence of a stationary solution to (5.6.35), admitting the causal representation

$$\mathbf{X}_t \quad = \quad \sum_{i \geq 0} \mathbf{A}_{t-i+1} \dots \mathbf{A}_t \, \mathbf{B}_{t-i}, \qquad t \in \mathbb{Z}, \tag{5.6.36}$$

where the summand for $i = 0$ is $\mathbf{B}_t$, for $i = 1$ is $\mathbf{A}_t \mathbf{B}_{t-1}$, and so on; see [6, 24] and references therein. We quote below Theorem 2.1. in [6] to explicit sufficient conditions for the existence of a causal stationary solution as in (5.6.36). In this case, if Θ_0 is the exponent measure of $\mathbf{X}_0$ such that

$$\Theta_t \quad = \quad \mathbf{A}_t \cdots \mathbf{A}_1 \Theta_0, \quad t \geq 0$$

where $(\mathbf{A}_t)$ is an iid sequence distributed as $\mathbf{A}$.

Theorem 5.6.8. *(Existence of stationary solution) Let $\mathbf{A}$, $\mathbf{B}$ be a $d \times d$ random matrix and a $\mathbb{R}^d$-valued random vector, respectively. Assume*

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

and assume the Lyapunov exponent of the iid sequence $(\mathbf{A}_t)$, distributed as $\mathbf{A}$, is negative. Then, the solution in (5.6.36) converges a.s. and is the unique strictly stationary causal solution to the stochastic recurrence equation in (5.6.35).

Solutions $(\mathbf{X}_t)$ to the stochastic recurrence equation in (5.6.35) have attired high attention because they can be regularly varying even when innovations $((\mathbf{A}_t, \mathbf{B}_t))$, are light-tailed as first noticed in [103]. Under the so-called Kesten's assumptions, the extremes come from the products of arbitrary length in (5.6.36); see [13] for a review, [74] for the Goldie and Kesten's conditions for univariate innovations, and [6] a treatement in the multivariate setting. Moreover, under these Kesten's type assumptions, the index of regular variation of $(\mathbf{X}_t)$, denoted $\alpha > 0$, is given as the unique solution to the equation

$$\lim_{n \to +\infty} n^{-1} \log \mathbb{E}\big[|\mathbf{A}_1 \cdots \mathbf{A}_n|_{op}^{\alpha}\big] \quad = \quad 0. \tag{5.6.37}$$

The multivariate solution to (5.6.35) models typical econometric time series models such as the squared ARCH(p) and the volatility of GARCH(p, q) processes. We review Theorem 2.4. and Corollary 2.7. in [6] to state exact conditions leading to heavy-tailed solutions $(\mathbf{X}_t)$, from light-tailed innovations $((\mathbf{A}_t, \mathbf{B}_t))$.

Theorem 5.6.9. *(Heavy-tailedness) Let $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$ be an iid sequence of $d \times d$ matrices $\mathbf{A}$ with non-negative entries, and non-negative $\mathbb{R}^d$-valued random vectors $\mathbf{B}$, $\mathbf{B} \neq \mathbf{0}$ a.s., Assume also the conditions below hold*

1. *For some $\kappa > 0$, $\mathbb{E}[|\mathbf{A}|_{op}^{\kappa}] < 1$.*

2. $\mathbf{A}$ *has no zero rows a.s.*

3. *The set* Γ *in* (5.6.38) *generates a dense group on* $\mathbb{R}$.

$$\Gamma \quad = \quad \{\ln |\mathbf{a}_n \cdots \mathbf{a}_1|_{op} : n \geq 1, \mathbf{a}_n \cdots \mathbf{a}_1 > 0,\ \mathbf{a}_n, \ldots, \mathbf{a}_1 \text{ are in the support of } \mathbf{A}\text{'s law } \}. \tag{5.6.38}$$

4. *There exists* $\kappa_1 > 0$, $\mathbb{E}\big[(\min_{i=1,\ldots,d} \sum_{t=1}^{d} A_{ij})^{\kappa_1}\big] \geq d^{\kappa_1/2}$ *and* $\mathbb{E}[|\mathbf{A}|_{op}^{k_1} \ln^+ |\mathbf{A}|_{op}] < +\infty$.

Then, there exists a unique solution $\alpha > 0$ *to* (5.6.37)*, and there also exists a unique strictly stationary causal solution* $(\mathbf{X}_t)$ *to* (5.6.37)*. Moreover, if* $\mathbb{E}[|\mathbf{B}|^\alpha] < +\infty$ *and either* $d = 1$ *or* $d > 1$ *and* $\alpha > 0$ *is not an even integer, then the finite-dimensional distributions of* $(\mathbf{X}_t)$ *are regularly varying of index* $\alpha > 0$.

Proposition 5.6.10. *Assume* $((\mathbf{A}_t, \mathbf{B}_t))$ *is an iid sequence of* $d \times d$ *matrices* $\mathbf{A}$*, and* $\mathbb{R}^d$*-valued vectors* $\mathbf{B}$*, which satisfies the conditions of Theorem 5.6.9* and *Theorem 5.6.8. Let* $(\mathbf{X}_t)$ *be the causal regularly varying stationary solution to* (5.6.35)*, with index* $\alpha > 0$*. Assume also*

$$\mathbb{E}[|\mathbf{A}|_{op}^{\alpha-\delta}] < 1, \quad \text{and} \quad \mathbb{E}[|\mathbf{A}|_{op}^{1+\delta}] + \mathbb{E}[|\mathbf{B}|^{1+\delta}] < +\infty,$$

for some $\delta > 0$*. Then, for all* $p > \alpha/2$*,* **AC***,* $\mathbf{CS}_p$ *hold for all sequences* (x_n) *such that* $n/x_n^{p\wedge\alpha-\kappa} \to 0$*, for some* $\kappa > 0$*, as* $n \to +\infty$.

Proposition 5.6.11. *Assume the conditions of Proposition 5.6.10* and *let* $(\mathbf{X}_t)$ *be the causal regularly varying stationary solution to* (5.6.35)*, with index* $\alpha > 0$*. Assume also the* β*-mixing conditions* (5.2.10)*,* (5.4.25)*, and the bias condition* (5.2.11)*,* (5.2.12)*. Then for any* $f \in \mathcal{G}_+(\tilde{\ell}^p)$*, such that* $u \mapsto f(\cdot/u)$ *is non increasing*

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

Remark 5.6.12. *Under the assumptions of Theorem 5.6.9* and *Theorem 5.6.8, the stationary solution* $(\mathbf{X}_t)$ *of* (5.6.35) *is a Markov Chain. From Remark 5.3.2, further assumptions on* $\mathbf{A}$ *yield* (β_h) *has an exponential decay under the assumptions; see [23].*

Example 5.6.13. *Define the positive random variables* A*,* B *by* $\log A \stackrel{d}{=} N - 0.5$*, where* N *denotes a centered and reduced Gaussian random variable, and* B *any bounded positive random variable. Consider* (X_t) *to be the univariate stationary solution in* (5.6.35)*. Then,* (X_t) *is regularly varying with index* $\alpha = 1$ *and* $\mathbb{E}[(A)^{1-\delta}] < 1$ *for all* $\delta > 0$.

Remark 5.6.14. *Let* $d > 1$ *and assume conditions 1. - 4. in Theorem 5.6.9. In this case, regular variation as in* (5.3.16) *might not hold. In this case we can only show that linear projections* $\mathbf{x}^T\mathbf{X}_1$ *are regularly varying, for all* $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$*; see (2.11) in [6]. However, a Cramèr-Wold devise entailing multivariate regular variation as in* (5.3.16) *only holds under further assumptions. For* *example, if $\alpha > 0$ is not an even integer then the assumptions of Theorem 5.6.9 imply regular variation as in* (5.3.16)*; see [89, 6, 7].*

5.7 Proofs

5.7.1 Proof Theorem 5.2.1

Proof. Let $f \in \mathcal{G}_+(\tilde{\ell}^p)$ be a fixed bounded function. Recall the set $\mathcal{F}$ containing two functions: $(\mathbf{x}_t) \mapsto f(\mathbf{x}_t)$ and $(\mathbf{x}_t) \mapsto 1$, and the family $\mathcal{T}$ in (5.7.39) defined by

$$\mathcal{T} \quad = \quad \{\widetilde{g_p^{\mathbf{Q}}}(u)\}_{\{g(\cdot/u):\, u\in[1-\epsilon,1+\epsilon], g\in\mathcal{F}\}}. \tag{5.7.39}$$

for $\epsilon > 0$ as in (5.2.11).

We separate the proof in two steps. We start by assuming the uniform asymptotic normality of estimators indexed by $\mathcal{T}$ holds, and we show in this case (5.2.9) holds. In the second part of the proof we will show the uniformity of the limit Gaussian process indexed by $\mathcal{T}$.

For the first step of the proof, assume

$$\sqrt{k}(\widetilde{g_p^{\mathbf{Q}}}(u) - u^{-\alpha} g_p^{\mathbf{Q}}) \xrightarrow{d} \mathbb{G}(g(\cdot/u)), \quad \widetilde{g_p^{\mathbf{Q}}} \in \mathcal{T},$$

as $n \to +\infty$ holds uniformly and $\mathbb{G}$ is a Gaussian process with structure as in (5.2.15), that we recall below

$$\begin{aligned} \mathrm{Cov}(\mathbb{G}(g(\cdot/u)), \mathbb{G}(h(\cdot/v))) \quad &= \quad \int_{u\vee v}^{\infty} \mathbb{E}[g(y\mathbf{Q}^{(p)}/u)h(y\mathbf{Q}^{(p)}/v)]d(-y^{-\alpha}) \\ &= \quad c(g(\cdot/u), h(\cdot/v)) \end{aligned}$$

for $g, h \in \mathcal{F}$, and $u, v \in [1-\epsilon, 1+\epsilon]$. We also write $c(g(\cdot/u), g(\cdot/u)) = c(g(\cdot/u))$ in what follows.

Then, for $g \in \mathcal{F}$, $u \in [1-\epsilon, 1+\epsilon]$, we have

$$\sqrt{k}(\widetilde{g_p^{\mathbf{Q}}}(u)/g_p^{\mathbf{Q}} - u^{-\alpha}) \quad \xrightarrow{d} \quad \mathbb{G}(g(\cdot/u))/g_p^{\mathbf{Q}}, \quad n \to +\infty,$$

Then, taking $1 \in \mathcal{F}$ to be the constant function one: $1(\mathbf{x}_t) = 1$, yields $1_p^{\mathbf{Q}} = 1$. Moreover,

$$\begin{aligned} &\sqrt{k}(\widetilde{1_p^{\mathbf{Q}}}(u)^{\leftarrow} - (u^{-\alpha})^{\leftarrow}) \\ &\quad = \quad \sqrt{k}\big(\|\mathcal{B}_1/x_{b_n}\|_{p,(\lfloor ku \rfloor)} - u^{-1/\alpha}\big), \quad u \in [1-\epsilon, 1+\epsilon]. \end{aligned}$$

Then, by an application of Vervaat's lemma; see (5.9.5),

$$\sqrt{k}\big(\|\mathcal{B}_1/x_{b_n}\|_{p,(\lfloor k \rfloor)} - 1\big) \quad \xrightarrow{d} \quad -\alpha^{-1}\mathbb{G}\big(1(\cdot/1)\big),$$

as $n \to +\infty$. In particular $\|\mathcal{B}_1\|_{p,(\lfloor k \rfloor)}/x_{b_n} \xrightarrow{\mathbb{P}} 1$ and the joint convergence of distributions holds thus

$$\sqrt{k}\big(\,(\widetilde{f_p^{\mathbf{Q}}}(u),\, \|\mathcal{B}_1\|_{p,(k)}/x_b) - ((f_p^{\mathbf{Q}}(u),1)\,\big) \quad \xrightarrow{d} \quad \big(\mathbb{G}(f(\cdot/u)), -\alpha^{-1}\mathbb{G}(1(\cdot/1)\big),$$

uniformly for $u \in [1-\epsilon, 1+\epsilon]$, as $n \to +\infty$. Furthermore,

$$\begin{aligned}
\sqrt{k}\,\big(\,\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}\,\big) \quad &= \quad \sqrt{k}\,f_p^{\mathbf{Q}}\,\big(\,\widetilde{f_p^{\mathbf{Q}}}(\,\|\mathcal{B}_1\|_{p,(k)}/x_{b_n}\,)/f_p^{\mathbf{Q}} - 1\,\big) \\
&= \quad \sqrt{k}\,f_p^{\mathbf{Q}}\,\big(\,\widetilde{f_p^{\mathbf{Q}}}(\|\mathcal{B}_1\|_{p,(k)}/x_{b_n})/f_p^{\mathbf{Q}} - (\|\mathcal{B}_1\|_{p,(\lfloor k \rfloor)}/x_{b_n})^{-\alpha}\,\big) \\
&\quad\; + \sqrt{k}\,f_p^{\mathbf{Q}}\,\big(\,(\|\mathcal{B}_1/x_{b_n}\|_{p,(k)}\,)^{-\alpha} - 1\big) \\
&\xrightarrow{d} \quad f_p^{\mathbf{Q}}\,\mathbb{G}(f(\cdot/1)/f_p^{\mathbf{Q}} - 1)
\end{aligned}$$

To sum up, $\sqrt{k}\big(\,\widehat{f_p^{\mathbf{Q}}} - f_p^{\mathbf{Q}}\,\big) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(f(Y\mathbf{Q}^{(p)})))$, and we recognize that the variance term can be rewritten as in (5.2.9).

To conclude the proof, we must show

$$\sqrt{k}(\widetilde{g_p^{\mathbf{Q}}}(u) - g_p^{\mathbf{Q}}) \quad \xrightarrow{d} \quad \mathbb{G}, \quad \widetilde{g_p^{\mathbf{Q}}}(u) \in \mathcal{T},$$

uniformly as $n \to +\infty$. Recall, we assumed finite-dimensional convergence holds in (5.2.14). Under the bias assumptions (5.2.11),(5.2.12), we can replace the expected value of the estimator $\widetilde{g_p^{\mathbf{Q}}}(u)$ directly by the limit $g_p^{\mathbf{Q}}(u)$. Then, it remains to check asymptotic equicontinuity holds and this will yield the uniformity of the asymptotic normality. To do so, we split the remaining of the proof into two steps as follows. First, we have to show that the sequence $(\mathcal{B}_t)_{t=1,\dots,m_n}$ can be replaced by a sequence $(\mathcal{B}_t^*)_{t=1,\dots,m_n}$, containing iid blocks distributed as $\mathcal{B}_1$. Second, we have to show that the conditions $(i),(ii),(iii)$, of Theorem C.4.5 in [108] hold; see also Theorem 2.3. in [51].

Let's denote u_0, s_0 with $u_0 = 1-\epsilon < + < 1+\epsilon = s_0$. Notice in our case, it is enough to check equicontinuity separately over $(\widetilde{g_p^{\mathbf{Q}}}(u))_{u \in [tu_0, s_0]}$, for $g \in \mathcal{F}$, as we only consider a family $\mathcal{F}$ containing two functions. We fix $g \in \mathcal{F}$. Let's start by considering the semi-metric $d(\cdot,\cdot)$ defined as

$$d(g(\cdot/u), g(\cdot/s)) \quad := \quad |u^{-\alpha} - s^{-\alpha}|\,\mathbb{E}[g(Y\mathbf{Q}^{(p)})^2], \quad s, u > 0,$$

such that $(\widetilde{g_p^{\mathbf{Q}}}(u))_{u \in [u_0, s_0]}$ is totally bounded for this semi-metric. For the first part, for replacing $(\mathcal{B}_t)$ by $(\mathcal{B}_t^*)$ we argue as in Section 10.6. in [108]. For any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\substack{u,s \in [u_0,s_0] \\ d(g(u^{-1}\cdot), g(s^{-1}\cdot)) < \delta}} \sqrt{k}\,|\widetilde{g_p^{\mathbf{Q}}}(u) - \widetilde{g_p^{\mathbf{Q}}}(s)| > \delta\Big) \\
&\quad \le \quad 2\,\mathbb{P}\Big(\sup_{\substack{u,s \in [u_0,s_0] \\ d(g(u^{-1}\cdot), g(s^{-1}\cdot)) < \delta}} \sqrt{k}\,|\widetilde{g_{p,*}^{\mathbf{Q}}}(u) - \widetilde{g_{p,*}^{\mathbf{Q}}}(s)| > \delta/2\Big) + 4\,m_n \beta_{b_n}.
\end{aligned}$$

where $\widetilde{g_{p,*}^{\mathbf{Q}}}(u) = \frac{1}{k}\sum_{t=1,\dots,m_n, t \text{ odd}} \; g(\mathcal{B}_t^*/u\,x_b)\mathbb{1}(\|\mathcal{B}_t^*\|_p > u\,x_b)$, and the last bound follows by Lemma

2 in [58].

We are now in the framework of Theorem C.4.5. in [108] and it is enough to check conditions (i),(ii),(iii) therein. The Lindeberg condition (i) is verified since g is bounded. We now show (ii) holds (this is also condition (D1) in [51]). Under the bias conditions, convergence of the finite dimensional distributions of $\mathcal{T}$ yields, for $s > u$,

$$\begin{aligned}
&k\,\mathbb{E}[\,(\widetilde{g^{\mathbf{Q}}_{p,*}}(u) - \widetilde{g^{\mathbf{Q}}_{p,*}}(s))^2\,] \\
&\to \quad c(g(\cdot/u)) + c(g(\cdot/s)) - 2c(g(\cdot/u), g(\cdot/s)) \\
&= \quad \int_1^{+\infty} (u^{-\alpha} + s^{-\alpha})\, g(y\mathbf{Q}^{(p)})^2 - 2s^{-\alpha} g(y\mathbf{Q}^{(p)}) g((s/u)\, y\mathbf{Q}^{(p)}) d(-y^{-\alpha}).
\end{aligned}$$

Moreover, since $v \mapsto g(\cdot/v)$ is a non-increasing function, then

$$\begin{aligned}
\lim_{k\to+\infty} \mathbb{E}[\,(\widetilde{g^{\mathbf{Q}}_{p}}(u) - \widetilde{g^{\mathbf{Q}}_{p}}(s))^2\,] &\leq |u^{-\alpha} - s^{-\alpha}| \mathbb{E}[g(Y\mathbf{Q}^{(p)})^2] \\
&= d(g(\cdot/u), g(\cdot/s)).
\end{aligned}$$

Then, to sum up we have shown

$$\lim_{\delta\downarrow 0} \limsup_{n\to+\infty} \sup_{\substack{u,s\in[u_0,s_0] \\ d(g(\cdot/t),g(\cdot/s))<\delta}} k\,\mathbb{E}[\,(\widetilde{g^{\mathbf{Q}}_{p}}(u) - \widetilde{g^{\mathbf{Q}}_{p}}(s))^2\,] = 0.$$

Thus, from this we conclude (ii) holds. Finally, the entropy condition in (iii) holds with respect to the random metric $(d_n, (g(\cdot/u))_{u\in[u_0,s_0]})$ defined by

$$\begin{aligned}
&\big(d_n(g(\cdot/u), g(\cdot/s))\big)^2 \\
&= \quad \frac{1}{k}\sum_{t=1}^{m_n} \big(g(\mathcal{B}^*_t/(u\,x_b))\mathbb{1}(\|\mathcal{B}^*_t\|_p > ux_b) - g(\mathcal{B}^*_t/(s\,x_b))\mathbb{1}(\|\mathcal{B}^*_t\|_p > s\,x_b)\big)^2.
\end{aligned}$$

Indeed, the family of functions $(\widetilde{g^{\mathbf{Q}}_{p}}(t))_{t\in[t_0,s_0]}$ is linearly ordered thus it forms a VC(2)-class (cf. [52], Remark 2.11 and the discussion on condition (D3) in [51]). Hence, we conclude uniform asymptotic normality of the estimators indexed by $\mathcal{T}$. Finally, by the first step of the proof we have shown Theorem 5.2.9 holds. □

5.7.2 Proof Lemma 5.4.1

Condition (1) $\implies$ (5.4.23)

We start by denoting disjoint blocks as

$$\mathcal{B}_t \quad := \quad \mathbf{X}_{(t-1)b+[1,b]}, \qquad \mathcal{B}^*_t \quad := \quad \mathbf{X}^*_{(t-1)b+[1,b]}, \qquad t = 1,\ldots,m\,. \tag{5.7.40}$$

such that $(\mathcal{B}_t^*)_{1\le t\le m}$ is a sequence of iid blocks, distributed as $\mathcal{B}_1$, independent of $(\mathcal{B}_t)_{1\le t\le m}$ Then, the mean value theorem entails $|e^{-x}-e^{-y}|\le|x-y|$, thus

$$\begin{aligned}
&\big|\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t)\big\}\big]-\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t^*)\big\}\big]\big|^2\\
&\le\ \mathbb{E}\big[\big(\tfrac{1}{k}\textstyle\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t)-\tfrac{1}{k}\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t^*)\big)^2\big]\\
&=\ I.
\end{aligned}$$

The term I can be controlled using the correlation coefficients (ρ_h), defined in Definition 5.3.1. It follows from stationarity and Theorem 1 in [157] that there exists a constant $c>0$ such that

$$I\ \le\ 2\,c\,\tfrac{m}{k^2}\mathbb{E}\big[f(x_b^{-1}\mathcal{B}_t)^2\big]\exp\big\{c\textstyle\sum_{t=0}^{\lfloor\log_2(n)\rfloor}\rho_{b\,2^t}\big\}$$

Moreover, by (5.3.19), for any function $g\in\mathcal{G}_+(\tilde{\ell}_p)$,

$$|\tfrac{m}{k}\mathbb{E}[g(x_b^{-1}\mathcal{B}_1)]-\textstyle\int_0^\infty g(y\mathbf{Q}^{(p)})d(-y^{-\alpha})|\to 0,\quad n\to+\infty.$$

Hence, if $\exp\big\{\sum_{t=0}^{\lfloor\log_2(n)\rfloor}\rho_{b\,2^t}\big\}/k\to 0$ as $n\to+\infty$, (5.4.23) holds. Thus, since $k\to+\infty$ we deduce condition (1) in Lemma 5.4.1 implies (5.4.23) and this concludes the proof.

(2) $\implies$ (5.4.23)

Consider a sequence $\ell:=\ell_n\to+\infty$. We denote disjoint blocks as

$$\mathcal{B}_{t,\ell}\quad:=\quad\mathbf{X}_{(t-1)b+[1,b-\ell]},\qquad t=1,\dots,m.$$

such that for $\ell=0$ we keep the notation in (5.7.40). Notice that for all $\delta>0,\epsilon>0$, and for every bounded Lipschitz-continuous function $f\in\mathcal{G}_+(\tilde{\ell}^p)$

$$\begin{aligned}
&\big|\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t)\big\}\big]-\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f(\underline{x_b^{-1}\mathcal{B}_t}_{\epsilon})\big\}\big]\big|\\
&\le\ \mathbb{E}\big[\big|\tfrac{1}{k}\textstyle\sum_{t=1}^m f(x_b^{-1}\mathcal{B}_t)-\tfrac{1}{k}\sum_{t=1}^m f(\underline{x_b^{-1}\mathcal{B}_t}_{\epsilon})\big|\big]\\
&\le\ \mathbb{E}\big[\tfrac{1}{k}\textstyle\sum_{t=1}^m\big|f(x_b^{-1}\mathcal{B}_t)-f(\underline{x_b^{-1}\mathcal{B}_t}_{\epsilon})\big|\big]\\
&=\ o\big(m\mathbb{P}(\|\overline{\mathcal{B}_1/x_b}^{\epsilon}\|_p>\delta)/k\big)
\end{aligned}$$

This term vanishes by condition $\mathbf{CS}_p$. Moreover, we denote I the following term

$$I=\big|\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f_\epsilon(x_b^{-1}\mathcal{B}_t)\big\}\big]-\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f_\epsilon(x_b^{-1}\mathcal{B}_{t,\ell})\big\}\big]\big|,$$

where $f_\epsilon(\mathbf{x}_t) := f(\underline{\mathbf{x}_t}_\epsilon)$. Then, there exists a constant $c > 0$ such that

$$\begin{aligned} I &\leq c\,\frac{1}{k}\,\mathbb{P}\big(\max_{1\leq j\leq m}\max_{1\leq i\leq \ell}|\mathbf{X}_{(j-1)b-i+1}| > \epsilon\,x_b\big) \\ &\leq c\,\frac{m}{k}\mathbb{P}(\|\mathcal{B}_{1,\ell}\|_\infty > \epsilon\,x_b) \\ &\leq c\,\frac{m\ell}{k}\mathbb{P}(|\mathbf{X}_0| > \epsilon\,x_b) \sim \ell/b\,(c\,\epsilon^{-\alpha}/c(p)) \quad = \quad O(\ell/b). \end{aligned}$$

Thus, we conclude that $\lim_{n\to+\infty}\ell_n/b_n = 0$ is a sufficient condition for $I \to 0$ as $n \to +\infty$. Furthermore, recall the definition of the mixing coefficients (β_h) in Definition 5.3.1. Arguing similarly as before it follows that

$$\begin{aligned} &\big|\mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f_\epsilon(x_b^{-1}\mathcal{B}_{t,\ell})\big\}\big] - \mathbb{E}\big[\exp\big\{-\tfrac{1}{k}\textstyle\sum_{t=1}^m f_\epsilon(x_b^{-1}\mathcal{B}_t^*)\big\}\big]\big| \\ &\leq \quad \frac{m}{k}\|f\|_\infty 2\,d_{TV}\big(\mathcal{L}(\mathcal{B}_{t,l})\otimes\underbrace{\mathcal{L}(\mathbf{X}_1)\otimes\cdots\otimes\mathcal{L}(\mathbf{X}_1)}_{\ell\text{ times}},\,\mathcal{L}(\mathcal{B}_t)\big) \\ &\leq \quad \frac{m}{k}\|f\|_\infty\,2\beta_{\ell_n} \to 0, \quad n \to +\infty. \end{aligned}$$

we use first the definition of the total variation distance, and second a reformulation of the distance in terms of the mixing coefficients (β_h). Hence we deduce that (2) in Lemma 5.4.1 implies (5.4.23).

5.7.3 Proof Lemma 5.4.3

We start by proving the following auxiliary lemmas.

Lemma 5.7.1. *Consider a stationary time series $(\mathbf{X}_t)$ satisfying the assumptions in Section 5.4.1. Assume that for all $t = 1, 2, \dots,$, $\epsilon > 0$,*

$$\lim_{s\to+\infty}\lim_{n\to+\infty}\textstyle\sum_{i=(t-1)b+s}^{tb}\mathbb{P}(|\mathbf{X}_i| > \epsilon\,x_b\,|\,|\mathbf{X}_0| > \epsilon\,x_b) = 0. \tag{5.7.41}$$

Then, for all $t = 2, 3, \dots,$ we deduce for $g, h \in \mathcal{G}_+(\tilde{\ell}^p)$

$$\lim_{n\to+\infty}\frac{m}{k}\mathrm{Cov}(\,g(x_b^{-1}\mathcal{B}_1), h(x_b^{-1}\mathcal{B}_t)\,) = 0. \tag{5.7.42}$$

Lemma 5.7.2. *Consider a stationary time series $(\mathbf{X}_t)$ satisfying the assumptions in Section 5.4.1. Assume one of the two conditions below holds*

1. *The mixing correlation coefficients (ρ_h) satisfy $\rho_h \to 0$ as $h \to +\infty$.*

2. *There exists a sequences (ℓ_n), satisfying $\ell_n \to +\infty$, and*

$$\lim_{n\to+\infty} m_n\beta_{\ell_n}/k_n \quad = \quad \lim_{n\to+\infty}\ell_n/b_n \quad = \quad 0.$$

Then, the relation below holds

$$\lim_{n\to+\infty}\frac{m}{k}\sup_{t=2,\dots,m_n}\ \mathrm{Cov}(\,g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_t)\,) \;=\; 0. \tag{5.7.43}$$

Proof Lemma 5.7.1. Let $f, f' : \tilde{\ell}^p \to \mathbb{R}$ be Lipschitz-continuous bounded functions in $\mathcal{G}_+(\tilde{\ell}^p)$. From condition $\mathbf{CS}_p$ it follows that for $t = 0, 1, 2, \dots,$

$$\begin{aligned}\frac{m}{k}\mathrm{Cov}(\,f(x_b^{-1}\mathcal{B}_1)f'(x_b^{-1}\mathcal{B}_{1+t})\,) &\sim \frac{m}{k}\mathbb{E}[f(x_b^{-1}\mathcal{B}_1)f'(x_b^{-1}\mathcal{B}_{1+t})]\\ &\sim \frac{m}{k}\mathbb{E}[f(\underline{x_b^{-1}\mathcal{B}_1}_{\epsilon})f'(\underline{x_b^{-1}\mathcal{B}_{1+t}}_{\epsilon})], \qquad n\to+\infty.\end{aligned}$$

Moreover, using a telescopic sum decomposition we might rewrite the term above such that for any $s > 0$,

$$\begin{aligned}&\frac{m}{k}\mathbb{E}[f(\underline{x_b^{-1}\mathcal{B}_1}_{\epsilon})f'(\underline{x_b^{-1}\mathcal{B}_{1+t}}_{\epsilon})]\\ \sim&\ \frac{\mathbb{E}\big[\,\frac{1}{b}\sum_{i=s}^{b}\big(f(\underline{x_b^{-1}\mathbf{X}_{[0,i]}}_{\epsilon}) - f(\underline{x_b^{-1}\mathbf{X}_{[1,i]}}_{\epsilon})\big)\,f'(\underline{x_b^{-1}\mathbf{X}_{i+(t-1)b+[1,b]}}_{\epsilon})\,\big]}{c(p)\mathbb{P}(|\mathbf{X}_0| > x_b)}\\ =&\ I\end{aligned}$$

Indeed, the finite terms in s are negligible as $n \to +\infty$ by the Lipschitz-continuity of the function f. Then, the term I can be bounded in absolute value by decomposing the function f' as a telescopic sum again as follows. We write $f_\epsilon(\mathbf{x}_t) := f(\underline{\mathbf{x}_t}_\epsilon)$ and $f'_\epsilon(\mathbf{x}_t) := f'(\underline{\mathbf{x}_t}_\epsilon)$ to simplify notation. Hence

$$\begin{aligned}|I| \;\le&\ \|f\|_\infty\frac{\mathbb{E}\big[\mathbb{1}(|\mathbf{X}_0| > \epsilon\, x_b)\frac{1}{b}\sum_{i=s}^{b} f'_\epsilon(x_b^{-1}\mathbf{X}_{\lfloor t-1\rfloor b+i+[1,b]})\big]}{c(p)\mathbb{P}(|\mathbf{X}_0| > x_b)}\\ \le&\ \mathbb{E}\big[\mathbb{1}(|\mathbf{X}_0| > \epsilon\, x_b)\\ &\ \|f\|_\infty\frac{\times\frac{1}{b}\sum_{i=s}^{b}\sum_{j=1}^{b-1}\big(f'_\epsilon(x_b^{-1}\mathbf{X}_{(t-1)b+i+[1,j]}) - f'_\epsilon(x_b^{-1}\mathbf{X}_{(t-1)b+i+[1,j+1]})\big)\big]}{c(p)\mathbb{P}(|\mathbf{X}_0| > x_b)}\\ &+\|f\|_\infty\|f'\|_\infty\frac{\mathbb{E}[\mathbb{1}(|\mathbf{X}_0| > \epsilon\, x_b)\times\frac{1}{b}\sum_{i=s}^{b}\mathbb{1}(|\mathbf{X}_{(t-1)b+i+1}| > \epsilon\, x_b)]}{c(p)\mathbb{P}(|\mathbf{X}_0| > x_b)}\\ \le&\ \|f\|_\infty\|f'\|_\infty\frac{\mathbb{E}\big[\mathbb{1}(|\mathbf{X}_0| > \epsilon\, x_b)\frac{1}{b}\sum_{i=s}^{b}\sum_{j=1}^{b-1}\mathbb{1}(|\mathbf{X}_{(t-1)b+i+j}| > \epsilon\, x_b)\big]}{c(p)\,\mathbb{P}(|\mathbf{X}_0| > x_b)} + o(1)\\ =&\ O\big(\textstyle\sum_{i=(t-1)b+s}^{tb}\mathbb{P}(|\mathbf{X}_i| > \epsilon\, x_b\,|\,|\mathbf{X}_0| > \epsilon\, x_b)\big)\end{aligned}$$

Then, by letting first $n \to +\infty$ and then $s \to +\infty$ we obtain (5.7.42) for Lipschitz-continuous function and this concludes the proof. Moreover, since for all $g, h \in \mathcal{G}(\tilde{\ell}^p)$ it still holds

$$\frac{m}{k}\mathrm{Cov}(\,g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_{1+t})\,) \;\sim\; \frac{m}{k}\mathbb{E}[g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_{1+t})]$$

then, it follows by a Portmanteau argument that (5.7.42) holds for all $g, h \in \mathcal{G}_+(\tilde{\ell}^p)$.

□

Finally, the proof of Lemma 5.7.2 follows similar arguments as in the proof of Lemma 6.7 in [28] as we show next.

Proof Lemma 5.7.2. Considering the same notation as in Lemma 5.4.1. We start by showing that the condition (1) is sufficient to obtain (5.7.43).

(1) $\Longrightarrow$ (5.7.43)

We fix $t \in \mathbb{N}$ and consider $\xi \in \mathbb{N}$ such that $\xi > t > 1$. Consider bounded functions $g, h \in \mathcal{G}(\tilde{\ell}^p)$, and let $\kappa > 0$ be such that f, g, vanish at the interior of the ℓ^p-sphere of radius κ. Then, by the definition of the correlation coefficients,

$$\begin{aligned}\frac{m}{k}\mathrm{Cov}(g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_{1+\xi})]) &\leq \rho_{tb}\,\frac{m}{k}\big(\mathbb{E}[g(x_b^{-1}\mathcal{B}_1)^2]\,\mathbb{E}[h(x_b^{-1}\mathcal{B}_1)^2]\big)^{1/2}\\ &= O(\rho_{tb}\,),\end{aligned}$$

as $n \to +\infty$. Then, by letting $n \to +\infty$ we conclude

$$\lim_{t\to+\infty}\lim_{n\to+\infty}\sup_{t=2,\dots,m_n}\frac{m}{k}\mathrm{Cov}(\,g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_t)\,) = 0. \tag{5.7.44}$$

Finally, for $\xi \leq t$ we apply Lemma 5.7.1 which together with (5.7.44) yields to (5.7.43).

(2) $\Longrightarrow$ (5.7.43)

Arguing as in the proof of Lemma 5.7.1, it is enough to establish the relation (5.7.43) for Lipschitz continuous bounded functions $f, f' \in \mathcal{G}_+(\tilde{\ell}^p)$ vanishing at the interior of the unity sphere in ℓ^p. For $\epsilon > 0$, recall the notation $f_\epsilon(\mathbf{x}_t) = f(\underline{\mathbf{x}_t}_\epsilon)$. Recall also we assume the Lipschitz-continuity of the function $f \in \mathcal{G}(\tilde{\ell}^p)$ and condition $\mathbf{CS}_p$. Then, as in the proof of Lemma 5.7.1, it is enough to show that for all $\epsilon > 0$,

$$\lim_{n\to+\infty}\sup_{\xi>1}\frac{m}{k}\mathrm{Cov}(f_\epsilon(x_b^{-1}\mathcal{B}_1)f'_\epsilon(x_b^{-1}\mathcal{B}_{1+\lfloor\xi\rfloor})) = 0.$$

Moreover, by similar steps as in the proof of Lemma 5.4.1 we can replace $\mathcal{B}_t$ by $\mathcal{B}_{t,\ell}$ inside the covariance since we assumed $\ell_n/b_n \to 0$ as $n \to +\infty$. Hence,

$$\frac{m}{k}\mathrm{Cov}(f_\epsilon(x_b^{-1}\mathcal{B}_{1,\ell})f'_\epsilon(x_b^{-1}\mathcal{B}_{1+\lfloor\xi\rfloor})) \leq 2\|f\|_\infty\|f'\|_\infty\frac{m}{k}\beta_{\ell+(\xi-1)b}$$

Thus, taking the supremum for $\xi > 1$ we conclude the proof letting $n \to +\infty$. □

Proof Lemma 5.4.3. We write $m = m_n = \lfloor n/b_n \rfloor$, $k = \lfloor m\mathbb{P}(\|\mathcal{B}_t\|_p > x_b)\rfloor \sim c(p)\, n\mathbb{P}(|\mathbf{X}_0| > x_b)$. Consider bounded functions $g, h \in \mathcal{G}_+(\tilde{\ell}^p)$,

$$\begin{aligned}
&\mathrm{Cov}\big(\tfrac{1}{k}\textstyle\sum_{t=1}^m g(x_b^{-1}\mathcal{B}_t), \tfrac{1}{k}\sum_{t=1}^m h(x_b^{-1}\mathcal{B}_t)\big)\\
&= \frac{m}{k^2}\mathrm{Var}[g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_1)] + \frac{2}{k^2}\textstyle\sum_{1\le t<j\le m}\mathrm{Cov}\big(g(x_b^{-1}\mathcal{B}_t), h(x_b^{-1}\mathcal{B}_j)\big)\\
&= \frac{(I+II)}{k}
\end{aligned}$$

The second term II can be rewritten in the following way by stationarity:

$$\begin{aligned}
II &= \tfrac{m}{k}\tfrac{1}{m}\textstyle\sum_{1\le t<j\le m}\mathrm{Cov}(g(x_b^{-1}\mathcal{B}_t)h(x_b^{-1}\mathcal{B}_j))\\
&= \tfrac{1}{m}\textstyle\sum_{1\le t\le m}\sum_{1<j-t\le m-t}\tfrac{m}{k}\mathrm{Cov}(g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_{j-t}))\\
&= \tfrac{1}{m}\textstyle\sum_{1\le t\le m}\sum_{1<j\le m-t}\tfrac{m}{k}\mathrm{Cov}[g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_j)]\\
&= \tfrac{1}{m}\textstyle\sum_{1<j\le m-1}\sum_{1\le t\le m-j}\tfrac{m}{k}\mathrm{Cov}[g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_j)]\\
&= \textstyle\sum_{1<j\le m-1}\big(1-\tfrac{j}{m}\big)\tfrac{m}{k}\mathrm{Cov}(g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_j)).
\end{aligned}$$

Then, following the lines of Lemma 5.7.2 we conclude that assuming $\sum_{t=1}^{m_n}\rho_{tb_n} \to 0$ we deduce $II \to 0$. Also, if $m_n/k_n\sum_{t=1}^{m_n}\beta_{\ell_n+(t-1)} \to 0$ then $II \to 0$ as $n \to +\infty$.

Finally, notice that for bounded functions $g, h \in \mathcal{G}_+(\tilde{\ell}^p)$,

$$\begin{aligned}
I &= \frac{m}{k}\mathrm{Var}(g(x_b^{-1}\mathcal{B}_1)h(x_b^{-1}\mathcal{B}_1))\\
&\to c(g,h) := \int_0^\infty g(y\mathbf{Q}^{(p)})h(y\mathbf{Q}^{(p)})d(-y^{-\alpha}).
\end{aligned}$$

and this concludes the proof. $\square$

5.7.4 Condition finite limit $c(p)$

Lemma 5.7.3. *Let $\alpha/2 < p < \alpha$. Assume the anti-clustering condition:*

$$\lim_{l\to+\infty}\limsup_{n\to+\infty}\sum_{|t|=l}^{n}\frac{\mathbb{E}[(|x_n^{-1}\mathbf{X}_t|^p\wedge 1)\mathbb{1}(|\mathbf{X}_0|>x_n)]}{\mathbb{P}(|\mathbf{X}_0|>x_n)} = 0. \tag{5.7.45}$$

Then $\|\Theta\|_p < +\infty$ a.s. and also $c(p) < +\infty$.

Remark 5.7.4. *Conditions of the type (5.7.45) were considered also in [5] in the context of stable central limit theorems for regularly varying time series.*

5.7.5 Proof Lemma 5.7.3:

Let $p < \alpha$. Condition (5.7.45) implies the classical anti-clustering condition which implies $|\Theta_t| \to 0$ as $|t| \to +\infty$. Moreover, by the mean value theorem, $c(p) < +\infty$ if one proves

$\mathbb{E}[(\sum_{t=0}^{\infty}|\Theta_t|^p)^{\alpha/p-1}] < +\infty$. Moreover, for $p > \alpha/2$, by Jensen's inequality and subadditivity we deduce it suffices to check

$$\sum_{t=0}^{\infty}\mathbb{E}\Big[|\Theta_t|^{\alpha-p}\mathbb{1}(|\Theta_t| > 1)\Big] + \sum_{t=0}^{\infty}\mathbb{E}[|\Theta_t|^p \wedge 1] < \infty. \tag{5.7.46}$$

We start by showing

$$\sum_{t\in\mathbb{Z}}\mathbb{E}[|\Theta_t|^p \wedge 1] < \infty. \tag{5.7.47}$$

Notice that (5.7.47) implies $\|\Theta\|_p < +\infty$ a.s. since $|\Theta_t| \to 0$ as $|t| \to +\infty$ a.s., and (5.7.47) follows by (5.7.45) and the Cauchy criterion as for any $\epsilon > 0$ there exists a K sufficiently large such that for all $l \geq K$, $h \geq 0$,

$$\limsup_{n\to+\infty}\sum_{|t|=l}^{l+h}\frac{\mathbb{E}[(|x_n^{-1}\mathbf{X}_t|^p \wedge 1)\mathbb{1}(|\mathbf{X}_0| > x_n)]}{\mathbb{P}(|\mathbf{X}_0| > x_n)} = \sum_{|t|=l}^{l+h}\mathbb{E}[|Y\Theta_j|^p \wedge 1] \leq \epsilon,$$

where the last equality follows by regular variation. Thus we conclude (5.7.47) holds. Finally, by the time-change formula we deduce

$$\begin{aligned}\infty &> \sum_{t\in\mathbb{Z}}\mathbb{E}[|\Theta_t|^{\alpha}(|\Theta_t|^{-p} \wedge 1)] \\ &> \sum_{t\in\mathbb{Z}}\mathbb{E}[|\Theta_t|^{\alpha-p}\mathbb{1}(|\Theta_t| > 1)],\end{aligned}$$

thus (5.7.46) holds. This concludes the proof.

5.8 Proofs examples

5.8.1 Proof Proposition 5.6.4

Proof. First, notice that (5.6.34) can be written as, for all $\delta > 0$,

$$\lim_{s\to+\infty}\limsup_{n\to+\infty}\frac{\mathbb{P}(\sum_{t=1}^{n}|\sum_{|j|>s}\varphi_j\mathbf{Z}_{t-j}/x_n|^p > \delta)}{n\mathbb{P}(|\mathbf{X}_0| > x)} = 0. \tag{5.8.48}$$

Once we have verified (5.8.48), **AC** and $\mathbf{CS}_p$ will follow straightforwardly since, for all $s > 0$, $(\mathbf{X}_t^{(s)})$ given by $\mathbf{X}_t^{(s)} = \sum_{|j|\leq s}\varphi_j\mathbf{Z}_{t-j}$ is an m_0-dependent sequence with $m_0 = 2k+1$, and satisfies $\mathbf{AC}, \mathbf{CS}_p$, from Example 5.6.1. Moreover, consistency of Lipschitz continuous functionals holds.

Then, for all $s > 0$,

$$\begin{aligned}
\mathbb{P}(|\widetilde{f^{\mathbf{Q}}_{p,s}} - \widetilde{f_p}^{\mathbf{Q}}| > \epsilon) &\leq \epsilon^{-1}\frac{\mathbb{E}[|f(\mathbf{X}^{(s)}_{[0,b]}) - f(\mathbf{X}_{[0,b]})|]}{\mathbb{P}(\|\mathbf{X}_{[0,b]}\|_p > x_b)} \\
&\leq \delta + \frac{\mathbb{P}(\|\mathbf{X}_{[0,b]}/x_b - \mathbf{X}^{(s)}_{[0,b]}/x_b\|_p > \delta\epsilon^{-1})}{\mathbb{P}(\|\mathbf{X}_{[0,b]}\|_p > x_b)}
\end{aligned}$$

where $\widetilde{f^{\mathbf{Q}}_{p,s}}$ is the deterministic threshold estimator in (5.2.8) for $\mathbf{X}^{(s)}$. Then, we can conclude letting first n and then s tend to infinity and finally δ to zero. To resume, for any $f \in \mathcal{G}_+(\tilde{\ell}^p)$ bounded Lipschitz continuous $\widetilde{f^{\mathbf{Q}}_p} \xrightarrow{\mathbb{P}} f^{\mathbf{Q}}_p$ and following the proof steps in Theorem 4.1. [26] we conclude consistency of $\widehat{f^{\mathbf{Q}}_p}$

In what follows, we focus on showing (5.8.48). Let's consider a Taylor decomposition of the function $|\cdot|^p$. For $a, b, \in \mathbb{R}$,

$$|a+b|^p = |a|^p + p\,\text{sign}(a)|a|^{p-1}b + \frac{p(p-1)}{2}|a|^{p-2}b^2 + \cdots + R_{[p]}(a,b)$$

where the remaining term can be written as

$$R_{[p]} = R_{[p]}(a,b) \leq \frac{p(p-1)\cdots(p-[p])}{[p]!}|b - \xi a|^{p-[p]}b^{[p]},$$

and $\xi \in [0,1]$. Then,

$$\begin{aligned}
|\sum_{|j|>k}\varphi_j\mathbf{Z}_{t-j}/x_n|^p &\leq |\varphi_t\mathbf{Z}_0/x_n|^p \\
&\quad + p\,\text{sign}(\varphi_t\mathbf{Z}_0)|\varphi_t\mathbf{Z}_0/x_n|^{p-1}\Big(\sum_{\substack{|j|>k\\ j\neq t}}|\varphi_j\mathbf{Z}_{t-j}/x_n|\Big) \\
&\quad + \cdots + R_{[p]}.
\end{aligned}$$

Then, by subadditivity of the function $x \mapsto x^{p-[p]}$, denoting $(|\mathbf{Z}_t|)$ by (Z_t),

$$\begin{aligned}
|\sum_{|j|>k}\varphi_j\mathbf{Z}_{t-j}/x_n|^p &\leq |\varphi_t Z_0/x_n|^p + c|\varphi_t Z_0/x_n|^{p-1}\Big|\sum_{\substack{|j|>k\\ j\neq t}}|\varphi_j|Z_{t-j}/x_n\Big| \\
&\quad + \cdots + c|\varphi_t Z_0/x_n|^{p-[p]}\Big|\sum_{\substack{|j|>k\\ j\neq t}}|\varphi_j|Z_{t-j}/x_n\Big|^{[p]} \\
&\quad + c\sum_{\substack{|j|>k\\ j\neq t}}||\varphi_j|Z_{t-j}/x_n|^{p-[p]}\Big|\sum_{\substack{|j|>k\\ j\neq t}}|\varphi_j|Z_{t-j}/x_n\Big|^{[p]} \\
&= I_{0,t} + I_{1,t} + \cdots + I_{[p],t} + I_{[p]+1,t}
\end{aligned}$$

where $c > 0$ is again a constant of no interest, only depending on p, and we have decomposed the previous sum into finite terms. In fact, for $p \le 1$, we can use the subadditivity yielding

$$|\sum_{|j|>k} \varphi_j \mathbf{Z}_{t-j}|^p \;\; \le \;\; \sum_{|j|>k} |\varphi_j \mathbf{Z}_{t-j}|^p \; = \; I_{[p]+1,t}.$$

In the following, we show that for all $i = 0, \dots, [p]+1$,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \frac{\mathbb{P}(\sum_{t=1}^n I_{i,t} > \delta)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} = 0. \tag{5.8.49}$$

To show (5.8.49) we consider a typical truncation argument. It will be enough to show that for all $\epsilon, \delta > 0$,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \frac{\mathbb{P}(\sum_{t=1}^n \overline{I_{i,t}}^\epsilon > \delta)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} + \frac{\mathbb{P}(\sum_{t=1}^n \underline{I_{i,t}}_\epsilon > \delta)}{n\mathbb{P}(|\mathbf{X}_0| > x_n)} \;\; = \;\; 0. \tag{5.8.50}$$

where the truncated terms are defined as follows, for $i = 0, \dots, [p]+1$

$$\begin{aligned}
\overline{I_{i,t}}^\epsilon &:= |\overline{\varphi_t Z_0/x_n}^\epsilon|^{p-i}| \sum_{\substack{|j|>k\\ j\neq t}} \overline{|\varphi_j| Z_{t-j}/x_n}^\epsilon|^i, \\
\overline{I_{[p]+1,t}}^\epsilon &:= \sum_{\substack{|j|>k\\ j\neq t}} |\overline{\varphi_j Z_{t-j}/x_n}^\epsilon|^{p-[p]} | \sum_{\substack{|j|>k\\ j\neq t}} \overline{|\varphi_j| Z_{t-j}/x_n}^\epsilon|^{[p]}, \\
\underline{I_{i,t}}_\epsilon &:= |\underline{\varphi_t Z_0/x_n}_\epsilon|^{p-i}| \sum_{\substack{|j|>k\\ j\neq t}} \underline{|\varphi_j| Z_{t-j}/x_n}_\epsilon|^i, \\
\underline{I_{[p]+1,t}}_\epsilon &:= \sum_{\substack{|j|>k\\ j\neq t}} |\underline{\varphi_j Z_{t-j}/x_n}_\epsilon|^{p-[p]}| \sum_{\substack{|j|>k\\ j\neq t}} \underline{|\varphi_j| Z_{t-j}/x_n}_\epsilon|^{[p]},
\end{aligned}$$

We study first the truncation from below. For $i = 0, \cdots, [p]$, by the Markov's inequality,

$$\begin{aligned}
\mathbb{P}(\textstyle\sum_{t=1}^n \underline{I_{i,t}}_\epsilon > \delta) &\le \delta^{-1} n \mathbb{E}\big[|\underline{\varphi_t Z_0/x_n}_\epsilon|^{p-i}\big]\mathbb{E}\big[| \sum_{\substack{|j|>k\\ j\neq t}} \underline{|\varphi_j| Z_{t-j}/x_n}_\epsilon|^i\big] \\
&= o(n\mathbb{P}(|Z_0| > x_n))
\end{aligned}$$

where the last relation holds by Karamata's theorem. For the last term, by Markov's inequality

$$\begin{aligned}
&\mathbb{P}(\textstyle\sum_{t=1}^n \underline{I_{[p]+1,t}}_\epsilon > \delta) \\
&\le \delta^{-1} n \sum_{\substack{i_1,\cdots,i_{[p]+1} \\ |i_j|>k}} |\varphi_{i_1}\cdots\varphi_{i_{[p]}}||\varphi_{i_{[p]+1}}|^{p-[p]} \\
&\qquad\qquad \times\mathbb{E}[|\underline{Z_{i_1}/x_n}_{\epsilon\varphi_{i_1}^{-1}}\cdots\underline{Z_{i_{[p]}}/x_n}_{\epsilon\varphi_{i_{[p]}}^{-1}}||\underline{Z_{i_{[p]+1}}/x_n}_{\epsilon\varphi_{i_{[p]+1}}^{-1}}|^{p-[p]}] \\
&= \delta^{-1} n \sum_{|j|>k} |\varphi_i|^p \mathbb{E}[|\underline{Z_i/x_n}_{\epsilon\varphi_i^{-1}}|^p] + o(n\mathbb{P}(|Z_0|>x_n)) \\
&\le \delta^{-1} n\epsilon^{-\alpha+\delta} \sum_{|j|>k} |\varphi_i|^{p+\alpha-\delta} + o(n\mathbb{P}(|Z_0|>x_n))
\end{aligned}$$

where the last relation holds by an application of Karamata's theorem and Potter's bounds. Finally, we conclude

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\frac{\mathbb{P}(\sum_{t=1}^n \underline{I_{i,t}}_\epsilon > \delta)}{n\mathbb{P}(|\mathbf{X}_0|>x_n)} \quad = \quad 0,$$

letting first $n\to+\infty$ and then $k\to+\infty$.

We focus now on the terms from the truncation from above. In this case, assuming $n/x_n^p \to 0$, it will be equivalent to show

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\frac{\mathbb{P}(\sum_{t=1}^n \overline{I_{i,t}}^\epsilon - \mathbb{E}[\overline{I_{i,t}}^\epsilon] > \delta)}{n\mathbb{P}(|\mathbf{X}_0|>x_n)} = 0.$$

In this case we use Chebychev's inequality which yields

$$\mathbb{P}(\textstyle\sum_{t=1}^n \overline{I_{i,t}}^\epsilon - \mathbb{E}[\overline{I_{i,t}}^\epsilon] > \delta) \quad \le \quad 2\,\delta^{-2}\,n\displaystyle\sum_{t=0}^n \mathrm{Cov}(\overline{I_{i,0}}^\epsilon, \overline{I_{i,t}}^\epsilon)$$

As in the arguments for the truncation from above, the main difficulty lied in showing the term $i=[p]+1$ is negligeable. To handle this terms we compute the covariances

$$\begin{aligned}
&\mathrm{Cov}(\overline{I_{[p]+1,0}}^\epsilon, \overline{I_{[p]+1,t}}^\epsilon) \\
&\;= \sum_{\substack{i_1,\cdots,i_{[p]+1} \\ \ell_1,\cdots,\ell_{[p]+1} \\ |i_j|>k, |\ell_j|>k}} |\varphi_{i_1}\cdots\varphi_{i_{[p]}}||\varphi_{i_{[p]+1}}|^{p-[p]}|\varphi_{\ell_1}\cdots\varphi_{\ell_{[p]}}||\varphi_{\ell_{[p]+1}}|^{p-[p]} \\
&\qquad\times\mathrm{Cov}(|\overline{Z_{i_1}/x_n}^{\epsilon\varphi_{i_1}^{-1}}\cdots\overline{Z_{-i_{[p]}}/x_n}^{\epsilon\varphi_{i_{[p]}}^{-1}}||\overline{Z_{-i_{[p]+1}}/x_n}^{\epsilon\varphi_{i_{[p]+1}}^{-1}}|^{p-[p]}, \\
&\qquad\qquad |\overline{Z_{t-\ell_1}/x_n}^{\epsilon\varphi_{\ell_1}^{-1}}\cdots\overline{Z_{t-\ell_{[p]}}/x_n}^{\epsilon\varphi_{[p]}^{-1}}||\overline{Z_{t-\ell_{[p]+1}}/x_n}^{\epsilon\varphi_{\ell_{[p]+1}}^{-1}}|^{p-[p]})
\end{aligned}$$

such that in the above term the covariance vanishes for all except a finite number of index t. Then,

$$\begin{aligned}\sum_{t=0}^{n} Cov(\overline{I_{[p]+1,0}}^{\epsilon}, \overline{I_{[p]+1,t}}^{\epsilon}) &= \sum_{j\in\mathbb{Z}} |\varphi_j|^{2p}\mathbb{E}[|\overline{Z_{-j}/x_n}^{\varphi_j^{-1}\epsilon}|^{2p}] + o(\mathbb{P}(|\mathbf{X}_0| > x_n)) \\ &\sim \epsilon^{2p-\alpha}\sum_{j\in\mathbb{Z}} |\varphi_j|^{2p+\alpha-\delta}\mathbb{P}(|\mathbf{X}_0| > x_n) + o(\mathbb{P}(|\mathbf{X}_0| > x_n))\end{aligned}$$

The last step holds by an application of Karamata's theorem. Finally, since $p > \alpha/2$ we can let $n \to +\infty$ and then we conclude letting $\epsilon \downarrow 0$. In this case,

$$\lim_{\epsilon\downarrow 0}\limsup_{n\to+\infty} \frac{\sum_{t=0}^{n} \mathrm{Cov}(\overline{I_{i,0}}^{\epsilon}, \overline{I_{i,t}}^{\epsilon})}{\mathbb{P}(|\mathbf{X}_0| > x_n)} = 0, \tag{5.8.51}$$

for $i = [p] + 1$. Then, for $i = 0, \dots, [p]$ we can show with similar calculations that (5.8.51) holds. Finally, this shows (5.8.50), and this concludes the proof. □

5.8.2 Proof Proposition 5.6.10

Let $(\mathbf{X}_t)$ be the causal stationary solution to (5.6.35), admitting the representation in (5.6.36). We assume it is regularly varying with tail index $\alpha > 0$. We will denote the products $\Pi_{i+1,t} := \mathbf{A}_{i+1}\cdots\mathbf{A}_t$, $i, t \in \mathbb{Z}$ with $i < t$. Then, backward computations yield

$$\mathbf{X}_t = \Pi_t\mathbf{X}_0 + R_t, \quad t \geq 1, \tag{5.8.52}$$

where

$$\Pi_{i,j} := \mathbf{A}_i\cdots\mathbf{A}_j, \qquad R_t := \sum_{j=1}^{t}\Pi_{j+1,t}\mathbf{B}_j \quad t\geq 1, \quad i\leq j.$$

with the conventions: $\Pi_{1,t} = \Pi_t$ and $\Pi_{t+1,t} = Id$. Notice that the remaining R_t is measurable with respect to $\sigma\big((\mathbf{A}_i, \mathbf{B}_i)_{1\leq i\leq t}\big)$ and is independent of the sigma-field $\sigma\big((\mathbf{X}_i)_{i\leq 0}\big)$.

Condition **AC** has been shown for Theorem 4.17 in [121]. Then we focus on showing $\mathbf{CS}_p$ holds. For this purpose, we require the auxiliary results that we state below.

Lemma 5.8.1. *Let $(\mathbf{X}_t)$ be a regularly varying sequence of index $\alpha > 0$ and assume for $\alpha/2 < p \leq \alpha$, and for any sequence $\ell_n \to +\infty$, the following relation holds*

$$\sup_{x\in\Lambda_n} \frac{\mathbb{P}\big(\big|\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p - \mathbb{E}[\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p\big| > \delta\big)}{n\mathbb{P}(|\mathbf{X}_0| > x)} \tag{5.8.53}$$

$$\leq c\Big(\ell_n\mathbb{E}[|\overline{\mathbf{X}_0/x_n}^{1}|^p] + \epsilon^{2p}\sum_{t=\ell_n+1}^{n}\beta_t/\mathbb{P}(|\mathbf{X}_0| > x_n)\Big), \quad x_n > x_0. \tag{5.8.54}$$

for a constant $c > 0$. Suppose also that $(\mathbf{X}_t)$ has $\beta-$mixing coefficients (β_t) verifying $\beta_t = O(t^{-\kappa})$ for some $\kappa > 1 + \alpha/p + \delta$, then condition CS_p holds uniformly over $(x_n, +\infty)$, for any constant x_n

verifying $\lim_{n\to+\infty} n\mathbb{P}(|\mathbf{X}_0| > x_n) = 0$.

Lemma 5.8.2. *Let* $(\mathbf{X}_t)$ *be a causal stationary solution to the equation* (5.6.35)*. Assume it is regularly varying of index* $\alpha > 0$ *for a sequence of iid* $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$ *innovations and assume* $\mathbb{E}[|\mathbf{A}|_{op}^{\alpha-\delta}] < 1$ *for some* $\delta > 0$*. Then, for all* $p > \alpha/2$*,* (5.8.54) *holds for any choice of* ℓ_n*. Then, if* $n/x_n^p \to 0$ *and* $p > \alpha/2$*,* $\mathbf{CS}_p$ *holds.*

Remark 5.8.3. *In Lemma 5.8.1 and Lemma 5.8.2 he have shown a little more. Actually, we showed that* (β_h) *mixing condition are sufficient to show condition* $\mathbf{CS}_p$ *holds uniformly in the sense* $\lim_{\epsilon\to 0}\limsup_{n\to+\infty}$ (5.8.53) $= 0$ *over the uniform regions* $\Lambda_n := (x_n, +\infty)$ *where* (x_n) *satisfies* $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$*. In this case it is possible to extend the large deviation results for* $p-$*norms to uniform regions following the approach in [121]. This result extends the setting in Theorem 4.17 [121] where only bounded regions where considered.*

Finally, from Lemma 5.8.1 and Lemma 5.8.2 we conclude Proposition 5.6.10 holds. The proof is concluded as long as we have shown these auxiliary results.

Proof Lemma 5.8.1. The bound in (5.8.54) can be decomposed as $c(I_1 + I_2)$. For I_1 to go to zero one can choose $\ell_n := x_n^{p-\delta}$ for some $\delta > 0$. In this case, one can bound the second term up I_2 by

$$\begin{aligned}
\sum_{t=\ell_n+1}^{n} \beta_t/\mathbb{P}(|\mathbf{X}_0| > x_n) &= O\big(\textstyle\sum_{t=\ell_n+1}^{n} t^{-\kappa}/\mathbb{P}(|\mathbf{X}_0| > x_n)\big) \\
&= O\big(\ell_n^{-\kappa+1}/\mathbb{P}(|\mathbf{X}_0| > x_n)\big) \\
&\leq O(x_n^{-\kappa(p-\delta)+(p-\delta)+(\alpha+\delta)}).
\end{aligned}$$

Then I_2 vanishes as $n \to +\infty$ if $\kappa > 1 + (\alpha+\delta)/(p-\delta) > 1 + \alpha/p + \delta$ where the last bound holds by Karamata'a theorem and plugging in the value for ℓ_n. □

Proof Lemma 5.8.2.
Case $p > \alpha/2$ We will use that is $p < \alpha$ then $n\mathbb{E}[|\overline{\mathbf{X}_t/x_n}^{\epsilon}|^p] = o(1)$.

Let's write $\mathbf{X}'_t := \Pi_t \mathbf{X}'_0 + R_t$ with $\mathbf{X}'_0$ independent identically distributed as $\mathbf{X}_0$. Indeed, this corresponds to the solution of equation (5.6.35) with innovations $\big((\mathbf{A}'_t, \mathbf{B}'_t)\big)$ where $(\mathbf{A}'_t, \mathbf{B}'_t) = (\mathbf{A}_t, \mathbf{B}_t)$ for $t \leq 0$ and $(\mathbf{A}'_t, \mathbf{B}'_t)_{t\geq 1}$ is an iid sequence independent of $(\mathbf{A}_t, \mathbf{B}_t)_{t\leq 0}$. Then,

$$\mathbb{P}\big(\big|\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p - \mathbb{E}[\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p\big| > \delta\big) \quad \leq \quad 2\,n \sum_{t=0}^{n} I_t$$

where $I_t = \mathrm{Cov}(|\overline{\mathbf{X}_0}^\epsilon|, |\overline{\mathbf{X}_t}^\epsilon|)$. Thus,

$$
\begin{aligned}
I_t &= \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\mathbf{X}_t/x_n}^\epsilon|^p\big] - \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\mathbf{X}_t'/x_n}^\epsilon|^p\big] \\
&\leq \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p \left(|\overline{\mathbf{X}_t/x_n}^\epsilon|^p - |\overline{\mathbf{X}_t'/x_n}^\epsilon|^p\right)_+\big] \\
&\leq \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p \left(|\mathbf{X}_t/x_n|^p - |\mathbf{X}_t'/x_n|^p\right)_+ \mathbb{1}(|\mathbf{X}_t/x_n| \leq \epsilon)\mathbb{1}(|\mathbf{X}_t'/x_n| \leq \epsilon)\big] \\
&\quad + \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\, |\overline{\mathbf{X}_t/x_n}^\epsilon|^p \mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big] \\
&= I_{t,1} + I_{t,2}.
\end{aligned}
$$

Assume $p > 1$. Then, for the first term $I_{t,1}$, a Taylor decomposition yields

$$
\begin{aligned}
I_{t,1} &\leq p\,\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\mathbf{X}_t'/x_n - \mathbf{X}_t/x_n}^{2\,\epsilon}||\overline{\mathbf{X}_t'/x_n}^\epsilon + \xi\big(\overline{\mathbf{X}_t'/x_n - \mathbf{X}_t/x_n}^{2\epsilon}\big)|^{p-1}\,\big] \\
&= p\,\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\Pi_t\mathbf{X}_0'/x_n - \Pi_t\mathbf{X}_0/x_n}^{2\,\epsilon}||\overline{\mathbf{X}_t'/x_n}^\epsilon + \xi\big(\overline{\Pi_t\mathbf{X}_0'/x_n - \Pi_t\mathbf{X}_0/x_n}^{2\epsilon}\big)|^{p-1}\,\big]
\end{aligned}
$$

for some random variable $\xi \in (0,1)$ a.s. Then, by Jenssen's inequality we have that there exists a constant $c > 0$ such that

$$
\begin{aligned}
I_{t,1} &\leq p2^{(p-1)_+}\,\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\Pi_t\mathbf{X}_0'/x_n - \Pi_t\mathbf{X}_0/x_n}^{2\,\epsilon}||\overline{\mathbf{X}_t'/x_n}^\epsilon|^{p-1}\big] \\
&\quad + p2^{(p-1)_+}\,\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^p|\overline{\Pi_t\mathbf{X}_0'/x_n - \Pi_t\mathbf{X}_0/x_n}^{2\,\epsilon}|^p\big] \\
&\leq c\,\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^{2p}\big]^{1/2}\mathbb{E}\big[|\overline{\Pi_t\mathbf{X}_0'/x_n}^{4\,\epsilon}|^{2p}\big]^{1/2} \\
&\leq c\left(\mathbb{E}\big[\,|\overline{\mathbf{X}_0/x_n}^\epsilon|^{2p}\big]\mathbb{E}\big[|\Pi_t|^{\alpha-\delta}\big]\mathbb{P}(|\mathbf{X}_0| > x_n)\right)^{1/2}. \\
&= O\big(\big(\epsilon^{2p-\alpha}\mathbb{E}[|\Pi_t|^{\alpha-\delta}]\big)^{1/2}\mathbb{P}(|\mathbf{X}_0| > x_n)\,\big).
\end{aligned}
$$

The last result holds by an application of Potter's bounds and Karamata's theorem. Then, if $\mathbb{E}[|A|^{\alpha-\delta}] < 1$ we conclude $\lim_{\epsilon\downarrow 0}\limsup_{n\to+\infty}\sum_{t=1}^n I_{t,1}/\mathbb{P}(|\mathbf{X}_0| > x_n) = 0$. If $p < 1$ then we can use a subadditivity argument and we conclude by similar steps. Now we look at the term $I_{t,2}$

$$
\begin{aligned}
I_{t,2} &:= \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\, |\overline{\mathbf{X}_t/x_n}^\epsilon|^p \mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big] \\
&\leq \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\, |\overline{\Pi_t\mathbf{X}_0/x_n}^\epsilon|^p \mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big] + \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\, |\overline{R_t/x_n}^{2\epsilon}|^p \mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big] \\
&\quad + \mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\, \mathbb{1}(|\Pi_t\mathbf{X}_0/x_n| > \epsilon)\mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big]. \\
&= O(\mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\big]\mathbb{E}\big[|\overline{R_t/x_n}^{2\epsilon}|^p \mathbb{1}(|\mathbf{X}_t'/x_n| > \epsilon)\big]
\end{aligned}
$$

Then,

$$
\lim_{\epsilon\downarrow 0}\limsup_{n\to+\infty}\sum_{t=1}^n I_{t,2}/\mathbb{P}(|\mathbf{X}_0| > x_n) = O\big(n\mathbb{E}\big[|\overline{\mathbf{X}_0/x_n}^\epsilon|^p\big]\mathbb{P}(|\mathbf{X}_0/x_n| > \epsilon x_n)\big)
$$

In this case we argue as in Lemma 5.8.1 and write

$$\begin{aligned}
&\mathbb{P}\big(\big|\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p - \mathbb{E}[\|\overline{\mathbf{X}_{[1,n]}/x}^{\epsilon}\|_p^p\big| > \delta\big) \\
&\leq \; 2\,n \sum_{t=0}^{\ell_n} I_t + \sum_{t=\ell_n+1}^{n} I_t. \\
&\leq \; O(\ell_n \mathbb{E}[|\overline{\mathbf{X}_0/x_n}^{\epsilon}|^p] + \sum_{t=\ell_n+1}^{n} I_t) \\
&\leq \; O(\ell_n n \mathbb{E}[|\overline{\mathbf{X}_0/x_n}^{\epsilon}|^p] + n\epsilon^{2p} \sum_{t=\ell_n+1}^{n} \beta_t)
\end{aligned}$$

where in the last bound we use the covariance inequality for the (β_h) mixing coefficients. Finally, notice that if $n/x_n^p \to 0$ then $\mathbf{CS}_p$ holds without additional mixing conditions. □

5.9 Annexes

Theorem 5.9.1. *(Karamata's theorem) Let* $\mathbf{X}$ *be a regularly varying random variable with index* $\alpha > 0$*. Then, the following asymptotic bounds hold*

$$\begin{aligned}
\lim_{x\to+\infty} \frac{\mathbb{E}[|\mathbf{X}/x|^p \mathbb{1}(|\mathbf{X}| \leq x)}{\mathbb{P}(|\mathbf{X}| > x)]} &= \frac{p-\alpha}{\alpha}, && \text{if } p > \alpha \\
\lim_{x\to+\infty} \frac{\mathbb{E}[|\mathbf{X}/x|^p \mathbb{1}(|\mathbf{X}| > x)]}{\mathbb{P}(|\mathbf{X}| > x)} &= \frac{p-\alpha}{\alpha}, && \text{if } p < \alpha.
\end{aligned}$$

Theorem 5.9.2. *(Potter's bounds) Let* $\mathbf{X}$ *be a regularly varying random variable with index* $\alpha > 0$*. Then, there exists* $\epsilon > 0, x_0 > 0, c_0 > 0$ *such that for all* $x > y > x_0$*,*

$$c^{-1}(x/y)^{-\alpha-\epsilon} \leq \frac{\mathbb{P}(\mathbf{X} > x)}{\mathbb{P}(\mathbf{X} > y)} \leq c(x/y)^{-\alpha+\epsilon}.$$

Consider the multivariate version of Breiman's lemma proven in [6] extending Breiman's univariate result in [19].

Theorem 5.9.3. *(Multivariate Breiman's lemma) Let* $\mathbf{X}_0$ *be an* $\mathbb{R}^d$*-values regularly varying vector with index* $\alpha \geq 0$ *and let* A *be a random matrix in* $\mathbb{R}^{q\times d}$*, independent of* $\mathbf{X}_0$*. Assume* $0 < \mathbb{E}[\,|\mathbf{A}|_{op}^{\alpha+\delta}\,] < +\infty$ *for some* $\delta > 0$*. If* $n\mathbb{P}(a_n^{-1}\mathbf{X}_0 \in \cdot) \xrightarrow{v} \mu := \int_0^\infty \mathbb{P}(y\,\Theta_0 \in \cdot)d(-y^{-\alpha})$*, where* $|\Theta_0| = 1$ *a.s., then*

$$n\mathbb{P}(a_n^{-1}\mathbf{A}\mathbf{X}_0 \in \cdot) \;\; \xrightarrow{v} \;\; \tilde{\mu} := \int_0^\infty \mathbb{P}(y\,\mathbf{A}\Theta_0 \in \cdot)d(-y^{-\alpha}), \quad n \to +\infty$$

where $\xrightarrow{v}$ *denotes vague convergence on* $\mathbb{R}^d \setminus \{0\}$ *and* $\mathbb{R}^q \setminus \{0\}$ *respectively.*

Remark 5.9.4. *In particular, under the assumptions of Theorem 5.9.3, regular variation on* $\mathbb{R}^q$ *holds if the probability* $\mathbb{P}(|\mathbf{A}\Theta_0|_{op} > 0)$ *is strictly positive.*

Theorem 5.9.5. *(Vervaat's lemma) Let* $(F_n)_{n=0,\dots}$ *be a sequence of* $D[0,+\infty)$*-valued elements and* F_0 *has continuous paths. Assume* F_n *has non decreasing paths there exists* $k_n \to +\infty$ *such that*

$$k_n(F_n(t) - t) \xrightarrow{d} F_0(t), \quad n \to +\infty$$

in $D(0,+\infty)$*. Then,*

$$k_n(F_n(t)^{\leftarrow} - t) \xrightarrow{d} -F_0(t)$$

See Proposition 3.3. in [143]

Conclusions

This thesis focuses on the multivariate study of extremal time dependencies of heavy-tailed stationary time series. The heavy-tailed phenomenon aims to model random mechanisms attaining high-intensity levels regularly. For example, in hydrology, heavy rainfall follow this principle [102]. Furthermore, extreme weather conditions produced by storms/fronts typically entail large rainfall amounts, which can be recorded on numerous days and can have extensive spatial coverage. Naturally, neighboring weather stations will record, let's say, a storm, on the same day or week. Overall, the meteorological dynamics lead to temporal and spatial dependencies in the hydrology data sets. Moreover, modeling the frequency and magnitude of high-intensity precipitation levels is critical for constructing coverage policies and infrastructure plans to protect human, societal, and environmental valuables from hazards. Extreme value theory (EVT) is essential for the risk assessment of this heavy-tailed phenomenon. Actually, a significant component of environmental risk is its interconnected aspect [69]; thus, the spatiotemporal considerations are essential for accurately modeling heavy rainfall. For example, flooding often occurs due to heavy rainfall records co-occurring in the same region and period, overloading the drainage systems. For these reasons, Chapter 4 has proposed to model regional fall daily rainfall amounts in France (see Figure 1.1), from the data set consisting of three different regions introduced in Chapter 1, as a multivariate stationary regularly varying time series.

Consider a stationary regularly varying time series $(\mathbf{X}_t)$, taking values in $(\mathbb{R}^d, |\cdot|)$, with (tail) index $\alpha > 0$. Motivated by the rainfall data presented in Chapter 1, this thesis had two main objectives. First, to present a theoretical framework for modeling heavy-tailed extremes in space and time, and second, to propose new inference procedures tailored to aggregate spatiotemporal extremes thoughtfully. To achieve the first goal, I have proposed to model spatiotemporal features of extremal ℓ^p-blocks, i.e., blocks $\mathbf{X}_{[0,n]}$ with a large ℓ^p-norm, through p-clusters $\mathbf{Q}^{(p)} \in \ell^p$, for $p > 0$, (see Theorem 2.2.1). Comparing the asymptotic results for different choices of p (see Proposition 2.3.1), has pointed to the α-cluster $\mathbf{Q} = \mathbf{Q}^{(\alpha)} \in \ell^\alpha$ to be a good candidate for trailing the spatiotemporal characteristics of extreme periods, where $\alpha > 0$ is the (tail) index. Actually, extremal ℓ^∞-blocks had been previously studied and lead to the cluster (of exceedances) theory. The asymptotics of extremal ℓ^∞-blocks already demonstrated to be helpful to obtain theoretical results for heavy-tailed time series in [40, 8, 9]. My work has extended the classical approach undertaken in [40] for modeling extremal ℓ^∞-blocks; see also [40, 10, 9, 138, 28, 108]. In my case, I have studied extremal ℓ^p-blocks, for $p \in (0, +\infty]$. For $p < \infty$, my approach was built on the theory of large deviations of sums of heavy-tailed time series; see [121, 123].

Concerning the second goal of this thesis, I have proposed new inference methodologies which aggregate spatiotemporal extremes using the ℓ^α-norm. The new procedures I presented are justified by the asymptotics derived for extremal ℓ^α-blocks in Theorem 2.2.1. I have addressed cluster inference in Chapters 2, 3, and marginal inference in Chapter 4. For cluster inference, I have proposed disjoint blocks estimators to infer cluster statistics from p-clusters using extremal ℓ^p-blocks, for $p > 0$. I have stated consistency of the disjoint blocks estimators in Theorem 2.4.1, and asymptotic normality in Theorem 5.2.1. To compare inference procedures, recall a relationship between p-clusters has been established (see Proposition 2.3.1). It follows that the same constant can be estimated from different pairs f_p, p, by letting a suitable functional $f_p : \ell^p \to \mathbb{R}$ act on extremal ℓ^p blocks. Inference based on extremal ℓ^α-blocks has proven to be advantageous compared to ℓ^∞-based inference in terms of bias. For example, the extremal index can be estimated from extremal ℓ^α-blocks (see Corollary 2.4.2). Concerning marginal inference, I have introduced the stable sums method to improve high return levels inference of $\mathbf{X}_0$. This new method is explained in Algorithm 1, and it is justified by the asymptotics of extremal ℓ^α-blocks. Roughly speaking, the method starts by computing the ℓ^α-norm of disjoint short periods to create a new sample. Then, it relies on central limit theory of heavy-tailed increments to extrapolate high quantiles from the new sample of ℓ^α-norms. Actually, using the (tail) index $\alpha > 0$ for improving inference procedures had been already studied in [10, 39, 50]. The α-cluster theory, from the new large deviation results in Theorem 2.2.1, has proven to be helpful to pursue this approach, since the new inference methodologies I presented are built on the asymptotics of extremal ℓ^α-blocks.

To conclude, the α-cluster, defined to model extremal ℓ^α-blocks, has demonstrated to be a nice candidate to capture heavy-tailed extremes with spatial and temporal coverage. Inference methodologies based on extremal ℓ^α-blocks have appeared to be robust to handle time dependencies. Finally, I want to highlight how aggregating all the recordings from the underlying extreme event (which were recorded over space and time) can be advantageous for enhancing inference. In terms of robustness, using the ℓ^α-norm proved to be more fruitful than, for example, picking only the largest value in a short period. Nevertheless, even in the stationary setting, many theoretical and practical challenges remain to improve the risk estimation of spatiotemporal data sets. For example, from a theoretical perspective, computations of the α-clusters are unavailable for many classical models. Concerning the practical difficulties, notice I have not fully accounted for the uncertainties from estimating the index of regular variation α in the methodologies presented in this thesis. Moreover, statistical methods for extremes come in several steps, where the first step is typically choosing a high threshold or selecting a block length. In this scenario, comparing and scoring statistical procedures for extremes is complex due to multiple-steps estimation and hyperparameter tuning. Overall, a list of open questions can be very long. I conclude by detailing three future perspectives of my work.

Future perspectives

- We have studied extremal ℓ^α-blocks in Chapter 2. We have mentioned the α-cluster is a good candidate to characterize extremal spatiotemporal features. They were defined as the spectral component in ℓ^α of the spectral tail measure, i.e., $\mathbf{Q}^{(\alpha)} = \Theta/\|\Theta\|_\alpha$. Computations of the forward spectral tail: $(\Theta_t)_{t\geq 0}$, are available for classical models like linear models and stochastic recurrence equations (SRE). However, less is known about the backward spectral tail $(\Theta_t)_{t\leq 0}$ for the (SRE) under the Kesten's conditions. A methodology for inferring the full spectral tail of stochastic recurrence equations should improve α-cluster inference. Indeed, we have discussed cluster inference in Chapters 2, 3, only through non-parametric approaches based on disjoint blocks estimators. Imposing a Markov structure could improve estimates of α-cluster statistics.

- We have presented in Chapter 4 the stable sums method for multivariate high quantile inference. We have proposed to aggregate spatiotemporal data through the ℓ^α-norm. In this case, the univariate ℓ^α-norm drives the extreme behavior of the d-variate time series. For high return levels inference, in the first step we model the sums of α-powers, and then we allocate weights $m(j)$, for $j = 1, \ldots, d$, to each coordinate to retrieve marginal features. Our approach focuses on extremal sets: $\{X_j > x\}$, $j = 1, \ldots, d$, but we can retrieve other $d-$variate extremal features similarly. For example, inference on sets of the form $\{\max\{X_i, X_j\} > x\}$ could be possible while allocating new weights $m(i, j)$. In practice, for a full risk assessment, it is important to model compound multivariate extremes like sites i, j, recording large amounts simultaneously. For example, in hydrology, a frequent cause of flooding is the saturation of drainage systems due to heavy rainfall recorded over the same region at the same time. This application motivates the study of compound spatiotemporal extremes. The theoretical tools in this thesis could help tackling this problems.

- Implementation of the stable blocks method from Algorithm 1 assumes that the vector coordinates reach high levels with similar intensities. Our motivation behind this assumption is that, for climatological time series, extremes can have a similar source explained by shared weather mechanisms. Extreme weather conditions then propagate in space and time following physical laws. In this setting, to apply our methodology to a larger spatial grid, we must first detect spatial patches with akin extreme climates. A spatial cluster detection for similar extreme events is studied in [112]. Coupling these techniques with our stable sums heuristic opens the road to new research perspectives.

Bibliography

[1] Asadi, P., Engelke, S. and Davison, A.C. (2018) Optimal regionalization of extreme value distributions for flood estimation. *Journal of Hydrology.* **556**, 182–193.

[2] Asmussen, S. (2003) *Applied Probability and Queues.* 2nd edition. Springer, New York.

[3] Araujo, A., and Giné, E. (1980) *The Central Limit Theorem for Real and Banach Random Variables.* Wiley, New York.

[4] Andrews, B., Calder, M. and Davis, R.A. (2009) Maximum likelihood estimation for α-stable autoregressive processes. *The Annals of Statistics.* **37**, 1946–1982.

[5] Bartkiewicz, K., Jakubowski, A., Mikosch, T. and Wintenberger, O. (2011) Stable limits for sums of dependent infinite variance random variables. *Probab. Th. Relat. Fields* **150**, 337–372.

[6] Basrak, B., Davis, R.A. and Mikosch, T. (2002) Regular variation of GARCH processes. *Stoch. Prob. Appl.* **99**, 95–115.

[7] Basrak, B., Davis, R.A. and Mikosch, T. (2002) A characterization of multivariate regular variation. *Ann. Appl. Prob.* **12**, 908–920.

[8] Basrak, B., Krizmanić, D. and Segers, J. (2012) A functional limit theorem for dependent sequences with infinite variance stable limits. *Ann. Probab.* **40**, 2008–2033.

[9] Basrak, B., Planinić, H. and Soulier, P. (2018) An invariance principle for sums and record times of regularly varying stationary sequences. *Probab. Th. Rel. Fields* **172**, 869–914.

[10] Basrak, B. and Segers, J. (2009) Regularly varying multivariate time series. *Stoch. Proc. Appl.* **119**, 1055–1080.

[11] Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002) On exponential representations of log-spacings of extreme order statistics. *Extremes.* **5**, 157–180.

[12] Berlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2006) *Statistics of extremes: theory and applications.* John Wiley & Sons, Chichester.

[13] Bingham, N., Goldie, C.M. and Teugels, J. (1987) *Regular Variation.* Cambridge University Press, Cambridge (UK).

[14] Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2**, 107-144.

[15] Bradley, R.C. (1988) A central limit theorem for stationary ρ-mixing sequences with infinite variance. *Ann. Probab. 16*, 313–332.

[16] Berghaus, B. and Bücher, A. (2018) Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Stat.* **46** , 2307–2335.

[17] Breiman, L. (1968) *Probability* Addison-Wesley, Reading, Mass.

[18] Brockwell, P.J. and Davis. R.A. (2002) *Introduction to time series and forecasting.* New York, Springer New York.

[19] Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.* **10**, 323–331.

[20] Brown, B.M. and Resnick, S.I. (1977) Extreme values of independent stochastic processes. *J. Appl. Probab.* **14**, 732–739.

[21] Bücher, A. and Jennessen, T. (2020). Method of moments estimators for the extremal index of a stationary time series. *Electr. J. Stat.* **14** , 3103–3156.

[22] Buraczewski, D., Damek, E. and Mikosch, T. (2016) *Stochastic Models with Power-Laws. The Equation $X = AX + B$.* Springer, New York.

[23] Buraczewski, D., Damek, E. and Mikosch, T. (2016). *Stochastic models with power-law tails. The Equation $X = AX + B$.* Springer Ser. Oper. Res. Financ. Eng., Springer, Chap, 10, 978-3.

[24] Buraczewski, D., Damek, E., Mikosch, T. and Zienkiewicz, J. (2013). Large deviations for solutions to stochastic recurrence equations under Kesten's condition. *The annals of Probability.* **41**, 2755–2790.

[25] Buriticá, G., Mikosch, T., Meyer, N. and Wintenberger, O. (2021) Some variations on the extremal index. *Zapiski Nauchn. Seminarov POMI*, to appear.

[26] Buriticá, G., Mikosch, T. and Wintenberger, O. (2022). Large deviations of ℓ^p-blocks of regularly varying time series. Technical report.

[27] Buriticá, G. and Naveau, P. (2022) Stable sums to infer high return levels of multivariate rainfall time series. https://arxiv.org/abs/2112.02878

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

[29] Chang, J.T. and Peres, Y. (1997) Ladder heights, Gaussian random walks and the Riemann zero function. *Ann. Probab.* **25**, 787–802.

[30] Chavez-Damoulin, V. and Davison, A.C. (2012) Modelling time series extremes. *REVSTAT-Statistical Journal.* **10**, 109–133.

[31] Cline, D. B. (1983) *Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data.* PhD Diss. Colorado State University.

[32] Cline, D.B.H. (1986) Convolution tails, product tails and domains of attraction. *Probab. Theory Related Fields.* **72**, 529–557.

[33] Cline, D.B.H. and Hsing T. (1998) *Large deviation probabilities for sums of random variables with heavy or subexponential tails, Technical Report.* Texas A&M University.

[34] Coles, B., Bawa, J., Trenner, L. and Dorazio, P. (2001) *An introduction to statistical modeling of extreme values.* Springer, London.

[35] Coles, S.G. and Tawn J.A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological).* **53**, 377–392.

[36] Coles, S.G. and Tawn J.A. (1991) Statistical methods for multivariate extremes: an application to structural design. *Journal of the Royal Statistical Society: Series B (Applied Statistics).* **43**, 1–31.

[37] Cooley, D., Nichka, D. and Naveau, P. (2007) Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association.* **102**. 824–840.

[38] Cooley, D., Davis, R.A. and Naveau, P. (2012) Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. **6**, 1406–1429.

[39] Davis, R.A., Drees, H., Segers, J. and Warchol, M. (2018) Inference on the tail process with applications to financial time series modeling. *J. Econometrics* **205**, 508–525.

[40] Davis, R.A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917.

[41] Davis, R.A. and Mikosch, T. (1998) The sample autocorrelation of heavy-tailed processes with applications to ARCH. *Annals of Statistics.* **26**, 2049–2080.

[42] Davis, R.A. and Resnick, S.I. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179–195.

[43] Davison, A.C. and Smith, R.L. (1990) Models for exceedances over high thresholds (with discussion). *J. Royal Statist. Soc., Series B.* **52**, 393–442.

[44] Davis, R.A., and Mikosch, T. (2009) The extremogram: A correlogram for extreme events. *Bernoulli* **15**, 977–1009.

[45] Davis, R.A. and Mikosch, T (2001) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Prob.* **23**, 879–917.

[46] Davis, R.A., and Resnick, S. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **3**, 179–195.

[47] Dedecker, J., Doukhan, P., Lang, G., Rafael, L. R. J., Louhichi, S. and Prieur, C. (2007) *Weak dependence: With examples and applications.* Springer, New York.

[48] Dombry, C., Hashorva, E. and Soulier, P. (2018) Tail measure and spectral tail process of regularly varying time series. *Ann. Appl. Probab.* **28**, 3884–3921.

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

[50] Drees, H., Janssen, A. and Neblung, S. (2021) Cluster based inference for extremes of time series. *Stochastic Processes and their Applications.* **142**, 1–33

[51] Drees, H. and Neblung, S. (2021) Asymptotics for sliding blocks estimators of rare events. *Bernoulli* **27**, 1239–1269.

[52] Drees, H. and Rootzén, H. (2010) Limit theorems for empirical processes of cluster functionals. *Ann. Stat.* **38**, 2145–2186.

[53] Drees, H., Segers, J. and Warchol, M. (2015) Statistics for tail processes of Markov chains. *Extremes* **18**, 369–402.

[54] DuMouchel, W.H. (1971) *Stable distributions in Statistical inference.* Dissertation Yale University.

[55] DuMouchel, W.H. (1973) On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distrbibution *The annals of Statistics.* **1**, 948–957

[56] DuMouchel, W.H. (1973) Stable distributions in statistical inference 1: Symmetric stable distributions compared to other symmetric long-tailed distributions. *JASA.* **68**, 469–477.

[57] DuMouchel, W.H. (1975) Stable distributions in statistical inference 2: Information from stable distributed sampled *JASA.* **40**, 386–393.

[58] Eberlein, Ernst. (1984) Weak convergence of partial sums of absolutely regular sequences. *Statistics and Probability Letters. 2*, 291–293.

[59] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling extremal events for insurance and finance.* Springer-Verlab, Berlin Heidelberg New York.

[60] Eastoe, E.F. and Tawn, J.A. (2012) Modelling the distribution of the cluster maxima of exceedances of subasymptotic threshold. *Biometrika.* **99**, 43–55.

[61] Evin, G., Favre, A.C. and Hingray, G. (2018) Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences.* **22**, 655–672.

[62] Fawcett, L. and Walshaw, D. (2007) Improved estimation for temporally clustered extremes. *Environmetrics.* **18**, 173–188.

[63] Fawcett, L. and Walshaw, D. (2012) Estimating return levels from serially dependent extremes *Environmetrics.* **23**, 272–283.

[64] Fawcett, L. and Walshaw, D. (2014) Estimating the probability of simultaneous rainfall extremes within a region: a spatial approach *Journal of Applied Statistics.* **41**, 959–976.

[65] Ferro, C. A. and Segers, J. (2003) Inference for clusters of extreme values. *J. Royal Statist. Soc., Series B.* **65**, 545–556.

[66] Feller, W. (1971) *An Introduction to Probability Thoery and Its Aplications I.* 3rd edition. Wiley, New York.

[67] Feller, W. (2008) *An introduction to probability theory and its applications.* John Wiley % Sons, New York.

[68] Ferreira, M. and Ferreira, H. (2013) Extremes of multivariate ARMAX processes. *TEST: An. oficial Journal of Spanish Society of Statistics and Operations Research* **22**, 606–627.

[69] Fields, C.S. and Barros, V.R. (2014) *Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects.* (IPCC) Cambridge University Press.

[70] Finkelshteyn, B.V. (1953) Limiting distribution of extreme values of the variational series of two dimensional random variables. Dokl. Ak. Nauk. SSSR 91. **2**, 209–211.

[71] Fougéres, A.L. (2004) Multivariate extremes. *Extreme values in finance, telecommunications, and the environment.* **1**, 373–388.

[72] Fougéres, A.L. and Mercadier, C. (2004). Risk measures and multivariate extension of Breiman's theorem *Journal of Applied Probability.* **49**, 364–384.

[73] Gilleland, E. and Katz, R.W. (2016) extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software.* **27**. 1–39.

[74] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

[75] Gomes, I., de Haan, L. and Rodrigues, L.H. (2008) Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 31–52.

[76] Gomes, M.I., de Haan, L. and Peng, L. (2002) Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes.* **5**, 387–414.

[77] Gumbel, E.J. (1960) Distributions de valeurs extremes en plusieurs dimensions. *Publ. Inst. Statist. Paris.* **9**, 171–173.

[78] Haan, L. de (1984) Slow variation and the characterization of domains of attraction. In: Tiago de Oliveira, J. *Statistical Extremes and Applications, p. 31-48.* Reidel, Dordrecht.

[79] Haan, L. de (1984) A spectral representation for max-stable processes. *Ann. Probab.* **12**, 1194–1204.

[80] Haan, L. de. and Ferreira, A. (2006) *Extreme Value Theory: and introduction.* Springer, New York.

[81] Haan, L. de, Resnick, S.I. (1977) Limit theory for multivariate sample extremes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. **40**, 317-.337.

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

[83] Haan, L. de, Mercadier, C. and Zhou, C. (2016) Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics* **20**, 321–354.

[84] Haan, L. de and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete.* **40**, 317–337.

[85] Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *The annals of statistics.* **1**, 1163–1174.

[86] Hult, H. and Lindskog, F. (2005) Extremal behavior of regularly varying stochastic processes. Stochastic Processes and their applications. **115**, 249–274.

[87] Hsing, T. (1991) On tail index estimation using dependent data *The Annals of Statistics.* **1**, 1547–1569.

[88] Hsing, T. (1993) Extremal index estimation for a weakly dependent stationary sequence. *Ann. Stat.* **21**, 2043–2071.

[89] Hult, H. and Lindskog, F. (2006) Regular variation for measures on metric spaces. *Publ. de l'Institut Math.* (Belgrade) **80** (94), 121–140.

[90] Hult, H. and Samorodnitsky, G. (2008). Tail probabilities for infinite series of regularly varying random vectors. *Bernoulli.* **14**, 838–864.

[91] Hult, H. and Samorodnitsky, G. (2010). Large deviations for point processes based on stationary sequences with heavy tails. *Journal of Applied Probability.* **47**, 1–40.

[92] Huser, R. and Wadsworth J.L. (2020) Advances in statistical modeling of spatial extremes *Wiley Interdisciplicary Reviews: Computational Statistics.* **14**, e1537.

[93] IPCC (2021) *Climate Change 2021: The physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergouvernmental Panel on Climate Change.* Cambridge University Press (In Press).

[94] Ibragimov, I.A. (1970) On the spectrum of stationary Gaussian sequences which satisfy the strongly mixing conditions II. Sufficient conditions. The rate of mixing. *Th. Probab. Appl.* **15**, 24–37.

[95] Ibragimov, I.A. (1975) A note on the central limit theorems for dependent random variables. *Th. Probab. Appl.* **20**, 135–141.

[96] Janssen, A. (2019) Spectral tail processes and max-stable approximations of multivariate regularly varying time series. *Stoch. Proc. Appl.* **129**, 1993–2009.

[97] Jakubowski, A. (1993) Minimal conditions in p-stable limit theorems. *Stoch. Proc. Appl.* **44**. 291–327.

[98] Jakubowski, A. (1997) Minimal conditions in p-stable limit theorems II. *Stoch. Proc. Appl.* **68**. 1–20.

[99] Kabluchko, Z. (2009) Spectral representations of sum- and max-stable processes. *Extremes.* **12**, 401–424.

[100] Karamata, J. (1933) Sur un mode de croissance régulière.Théorèmes fondamentaux. *Bulletin de la Société Mathématique de France.* **61**, 55-62.

[101] Kallenberg, O. (2017) *Random Measures, Theory and Applications.* Springer, Cham.

[102] Katz, R.W., Parlange, M. and Naveau, P. (2002) Extremes in hydrology. *Adv. Water. Resour.* **25**, 1287–1304.

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

[104] Kharin, V.V., Zwiers, F.W., Zhang, X. and Wehner, M. (2013) Changes in temperature and precipitation extremes in the CMIP5 ensemble *Climatic change.* **119**, 345–357.

[105] Kiriloük, A., Rootzén, H., Segers, J. and Wadsworth, J.L. (2019) Peaks over thresholds modeling with multivariate generalized Pareto distributions. *Technometrics.* **61**, 123–135.

[106] Konstantinides, D. G. and Mikosch, T. (2005). Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations. *The Annals of Probability.* **33**, 1992–2035.

[107] Kolmogorov, A.N. and Rozanov, Yu.A. (1960) On the strong mixing conditions for stationary Gaussian sequences. *Th. Probab. Appl.* **5**, 204–207.

[108] Kulik, R. and Soulier, P. (2020) *Heavy-Tailed Time Series.* Springer, New York.

[109] Leadbetter, M.R. (1983) Extremes and local dependence in stationary sequences. *Probab. Th. Relat. Fields* **65**, 291–306.

[110] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and related properties of random sequences and processes.* Springer, Berlin.

[111] Ledford, A.W. and Tawn, J.A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**. 169–187.

[112] Le Gall, P., Favre, A.C., Naveau, P. and Prieur, C. Improved regional frequency analysis of rainfall data (2021) https://hal.archives-ouvertes.fr/hal-03114324/.

[113] Lindskog, F., Resnick, S.I. and Roy, J. (2014) Regularly varying measures on metric spaces: hidden regular variation and hidden jumps. *Probab. Surveys* **11**, 270–314.

[114] Lin, T. and de Haan, L. (2001). On convergence toward an extreme value distribution in $C[0,1]$. The Annals of Probability. **29**, 467–483.

[115] Loynes, R.M. (1965) Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36**, 993–999.

[116] McCulloch, J.H. (2003) 13 Financial application of stable distributions. *Handbook of statistics.* **14**, 193–425.

[117] Mikosch, T. (1999) *Regular Variation, Subexponentiality and Their Applications in Probability Theory.* (Vol. 99) Eindhoven, The Netherlands:Eindhoven University of Technology.

[118] Mikosch, T. and Rodionov, I. (2021). Precise large deviations for dependent subexponential variables. *Bernoulli.* **27**, 1319–1347.

[119] Mikosch, T. and Nagaev, A. V. (1998) Large deviations of heavy-tailed sums with applications in insurance. *Extremes.* **1**, 81–110.

[120] Mikosch, T. and Samorodnitsky, G. (2000). The supremum of a negative drift random walk with dependent heavy-tailed steps. *The Annals of Applied Probability.* **10**, 1025-.1064.

[121] Mikosch, T. and Wintenberger, O. (2013) Precise large deviations for dependent regularly varying sequences. *Probab. Th. Rel. Fields* **156**, 851–887.

[122] MIKOSCH, T. AND WINTENBERGER, O. (2014) The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains. *Probab. Th. Rel. Fields* **159**, 157–196.

[123] MIKOSCH, T. AND WINTENBERGER, O. (2016) A large deviations approach to limit theory for heavy-tailed time series. *Probab. Th. Rel. Fields* **166**, 233–269.

[124] NAGAEV, S.V. (1979) Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745–789.

[125] NAGAEV, A.V. (1969) Integral limit theorems for large deviations when Cramer's condition is not fulfilled I,II. *Theory Probab. Appl* **14**, 51–64.

[126] NAVEAU, P., HUSER, R., RIBEREAU, P. AND HANNART, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without threshold selection. *Water Resources Research.* **52**, 2753–2769.

[127] NEWELL, G.F. (1964) Asymptotic extremes for m-dependent random variables. *Ann. Math. Statist.* **35**, 1322–1325.

[128] NIKIAS, C.L. AND SHAO, M. (1995) *Signal processing with α-stable distributions and applications.* Wiley, New York.

[129] NORTHROP, P.J. (2015) An efficient semiparametric maxima estimator of the extremal index. *Extremes.* **18**, 585–603.

[130] NOLAN, J.P. (2020) *Univariate stable distributions.* Berlin/Heidelberg, Germany, Springer.

[131] NOLAN, J.P. (2001) *Lévy Processes: Theory and Applications*, chapter Maximum likelihood estimation and diagnostics for stable distributions. Springer, Boston.

[132] O'BRIEN, G.L. (1974) Limit theorems for the maximum term of a stationary process. *Ann. Probab.* **2**, 540–545.

[133] O'BRIEN, G.L. (1987) Extreme values for stationary and Markov sequences. *Ann. Probab.* **15**, 281–291.

[134] OLIVEIRA, J.T. DE (1958) Extremal distributions. REV. FAC. CIENC. UNIV. LISBOA. **2**, 219–227.

[135] PICKANDS, J. (1969) Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.* **145**, 75–86.

[136] PELIGRAD, M. (1985) Invariance principles for mixing sequences of random variables. *Ann. Prob.* **10**, 968–981.

[137] Pescaroli, G. and Alexander, D. (2018) Understanding compound, interconnected, interacting, and cascading risks: a holistic framework. *Risk analysis.* **38**, 2245–2257.

[138] Planinić, H. and Soulier, P. (2018) The tail process revisited. *Extremes* **21**, 551–579.

[139] Planinic, H. (2021) Palm theory for extremes of stationary regularly varying time series and random fields. arXiv preprint arXiv:2104.03810.

[140] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Reprint 2008. Springer, New York.

[141] Resnick, S.I. and Stărică, C. (1995) Consistency of Hill's estimator for dependent data. *Journal of Applied probability.* **32**, 129–167.

[142] Resnick, S.I. (1997) Heavy tail modeling and teletraffic data. *Ann. Statist.* **25**, 1805–1869.

[143] Resnick, S.I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[144] Rio, E. (2017) *Asymptotic theory of weakly dependent random processes.* Springer, New York.

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

[146] Robert, C.Y. (2008) Estimating the multivariate extremal index function. *Bernoulli.* **14**, 1027–1064.

[147] Rootzén, H., Segers, J. and Wadsworth, J. L. (2018) Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis.* **165**, 117–131.

[148] Rootzén, H., Segers, J. and Wadsworth, J. L. (2018) Multivariate peaks over thresholds models. *Extremes.* **21**, 115–145.

[149] Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal.* **21**, 70–94.

[150] Ferro, C. A., Segers, J. and Robert, C. Y. (2009) A sliding blocks estimator for the extremal index. *Electr. J. Stat.* **3**, 993–1020.

[151] Samorodnitsky, G., Taqqu, M.S. and Linde R.W. (1996) Non-Gaussian random processes: stochastic models with infinite variance. *Bulletin of the London Mathematical Society.*

[152] Segers, J., Zhao, Y. and Meinguet, T. (2017) Polar decomposition of regularly varying time series in star-shaped metric spaces. *Extremes* **20**, 539–566.

[153] Segers, J. (2003) Functionals of clusters of extremes. *Advances in Applied Probability.* **35**, 1028–1045.

[154] Segers, J. (2005) Approximate distributions of clusters of extremes. *Stat. Probab. Letters* **74**, 330–336.

[155] Süveges, M. (2007) Likelihood estimation of the extremal index. *Extremes.* **10**, 41–55.

[156] Shao, Q.M. (1993) On the invariance principle for stationary ρ-mixing sequences. *Chinese Ann. Math.* **14B**, 27–42

[157] Shao, **Q**.M. (1995) Maximal inequalities for partial sums of ρ-mixing sequences. *The Annals of Probability.* 948–965.

[158] Sibuya, M. (1960) Bivariate extreme statistics, I. Annals of the Institute of Statistical Mathematics. **11**, 195–210.

[159] Süveges, M. and Davison, A.C. (2010) Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.* **4**, 203–221.

[160] Smith, R.L. (1989) Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statist. Sci.* **4**, 367–377

[161] Smith, R.L. and Weissman, I. (1994) Estimating the extremal index. *J. Royal Statist. Soc., Series B* **56**, 515–528.

[162] Smith, R.L., Tawn, J.A. and Coles, S.G. (1997) Markov chain models for threshold exceedances. *Biometrika.* **84**, 249–268.

[163] Starica, C. (1999) Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance.* **6**, 515–553.

[164] Tawn, J. A. (1988) Bivariate extreme value theory: Models and estimation. *Biometrika.* **75**, 397–415.

[165] Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika.* **77**, 245–253.

[166] Toulemonde, G., Guillou, A. and Naveau, P. (2013) Particle filtering for Gumbel distributed daily maxima of methane and nitrous oxide. *Environmetrics.* **24,** 51–62.

[167] Tendijck, S. Eastoe, E., Tawn, J., Randell, D. and Jonathan P. (2021) Modeling the Extremes of Bivariate Mixture Distributions with Application to Oceanographic Data. *Journal of the American Statistical Association.* **0** 1–12.

[168] Tyran-Kamińska, M. (2010) Convergence to Lévy stable processes under some weak dependence conditions. *Stochastic process. Appl.* **120**, 1629–1650.

[169] Utev, S.A. (1990) Central limit theorem for dependent random variables. *Probability Theory and Mathematical Statistics, Proceeding of the Fifth Vilnius Conference.* **2**, 519–528.

[170] Tencaliec, P., Favre, A.C., Naveau, P., Prieur, C. and Nicolet., G. (2019) Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics.* **31**, e2582.

[171] Zscheischler, J., Naveau, P., Martius, O. Engelke, S. and Raible, C.C. (2021) Evaluating the dependence structure of compound precipitation and wind speed extremes. *Earth System Dynamics Discussions.* **12**, 1–16.

Appendix A: Introduction to extremes

A.1 Outline of Chapter A

In this chapter we review classical theory and statistical methodology for extremes. We introduce univariate extreme value theory in Section A.2. We present the classical multivariate extension in Section A.3. The most common statistical aspects of both univariate and multivariate, are illustrated with the case study of fall daily heavy-rainfall in France from Chapter 1. This chapter aims to point at the drawbacks of classical approaches while applied to our data set. An expert reader can skip the introductory theory and go directly to Section A.4, where we discuss these limitations.

A.2 Univariate extremes

A.2.1 Regular variation and extreme values

In their work, Fisher and Tippett (1928) and Gnedenko (1943) derived the founding theorem of extreme value theory. All introductory books of the field include its statement, cf. Theorem 3.2.3 in [59], Theorem 2.1 in [12], Theorem 3.1. in [34], Theorem 1.1.3 in [80] or Proposition 0.3 in [140]. Mainly, it characterizes the family of distributions in the maximum domain of attraction of a distribution. Three types of domains are possible: Weibull, Gumbel, or Fréchet. Our focus is on the last one detailing heavy-tailed distributions. Typically, we aim to study distributions with a survival function decaying faster than at an exponential rate. We also want to consider distributions with an infinite variance such that extremes drive the large deviations from the median. We describe heavy-taildness in terms of regularly varying distributions following Karamata's theory in [100].

Definition A.2.1. *A function $f : \mathbb{R} \to \mathbb{R}$ is regularly varying of index $\alpha > 0$ if for all $t > 0$, $\lim_{x\to+\infty} f(tx)/f(x) = t^{-\alpha}$.*

Definition A.2.2. *A function $\ell : \mathbb{R} \to \mathbb{R}$ is slowly varying if for all $t > 0$, $\lim_{x\to+\infty} \ell(tx)/\ell(x) = 1$.*

Definition A.2.3. *A distribution function $F : \mathbb{R} \to [0, 1]$ is regularly varying with (tail)-index $\alpha > 0$ if and only if its survival function: $x \mapsto 1 - F(x)$ is regularly varying of index α.*

Naturally, from the definitions above, every regularly varying function of index $\alpha > 0$ can be written as $f(x) = x^{-\alpha}\ell(x)$ for a slowly varying function ℓ. The Table A.1 below summarizes classical examples of regularly varying distributions and their slowly varying term, e.g. Pareto distribution, the Fréchet distribution, the Burr distribution, the log-gamma; see Table 2.1. in [12]. Further examples of regularly varying distributions are the student distribution, the Lévy distribution and the Cauchy distribution.

Remark A.2.4. *The index of regular variation plays a key role in determining the heaviness of the tail. For example, if X a non-negative random variable with regularly varying distribution of index $\alpha > 0$, then for $p > \alpha$, $\mathbb{E}[X^p] = +\infty$, and for $p < \alpha$, $\mathbb{E}[X^p] < +\infty$; cf. [13].*

Remark A.2.5. *Concerning closure properties, let X_1, X_2, be non-negative independent regularly varying of index $\alpha > 0$, and let Y be non-negative such that $\mathbb{E}[Y^{\alpha+\epsilon}] < +\infty$,for $\epsilon > 0$, independent of X_1, X_2. Then, $X_1 + X_2$, $X_1 X_2$, and $Y X_2$ are regularly varying of index $\alpha > 0$. The last result is known as Breiman's lemma. We refer to [117, 32, 19].*

Remark A.2.6. *It is common to refer to heavy-tailed distributions only in the case of infinite variance, i.e., $\alpha < 2$. We study the whole family of laws verifying Definition A.2.3 with $\alpha > 0$.*

$F(x)$	$1 - F(x)$	$\ell(x)$	(tail)-index
Pareto $\alpha > 0$	$x^{-\alpha}$, $x > 1$	1	$\alpha > 0$
Fréchet $\alpha > 0$	$1 - \exp\{-x^{-\alpha}\}$, $x > 0$	$1 - \frac{x^{-\alpha}}{2} + o(x^{-\alpha})$	$\alpha > 0$
Burr $\eta, \tau, \lambda > 0$	$\left(\frac{\eta}{\eta + x^{\tau}}\right)^{\lambda}$, $x > 0$	$\left(\frac{\eta}{1+(\eta/x^{\tau})}\right)^{\lambda}$	$\lambda\tau > 0$
$\log\Gamma$ $\alpha, k > 0$	$\int_x^{\infty} \frac{\alpha^k}{\Gamma(k)} y^{-\alpha-1} (\log y)^{k-1} dy$, $x > 1$	$\frac{\alpha^{k-1}}{\Gamma(k)} (\log x)^{k-1} (1 + \frac{k-1}{\alpha \log x} + o(\frac{1}{\log x}))$	$\alpha > 0$

Table A.1: Regularly varying distribution F with (tail)-index $\alpha > 0$ satisfying $1 - F(x) = x^{-\alpha}\ell(x)$.

A.2.2 The maxima and exceedances modeling theorems

We state below the Fisher-Tippett-Gnedenko theorem for maxima modeling.

Theorem A.2.7. *(Fisher–Tippett–Gnedenko Theorem) Let (X_t) be an iid sequence of random variables. Assume there exists renormalizing constants $a_n > 0, b_n \in \mathbb{R}$ and a non-degenerate distribution G such that*

$$\lim_{n \to +\infty} \mathbb{P}(\max_{t=1,\ldots,n} X_t \leq x\, a_n + b_n) \;=\; G(x), \quad x \in \mathbb{R}, \tag{A.2.1}$$

then G belongs to the family of generalized extreme value distributions and

$$G(x) \;=\; G_\gamma(x; \mu, \sigma) \;=\; \exp\{-(1 + \gamma\tfrac{(x-\mu)}{\sigma})_+^{-1/\gamma}\}, \quad x \in \mathbb{R}. \tag{A.2.2}$$

for $\mu \in \mathbb{R}$, $\sigma > 0$ and $\gamma \in \mathbb{R}$ such as for $\gamma = 0$ one shall read the above expression as the limit as γ approaches zero.

As we mentioned, regularly varying distributions determine the maximum Fréchet-type domain of attraction, i.e., the case $\gamma > 0$ in (A.2.2) in the Fishet-Tippett-Gnedenko theorem. We rephrase Proposition 1.11 in [140] containing the original result from Gnedenko (1943).

Theorem A.2.8. *The condition in* (A.2.1) *holds for $G = G_\gamma$, where $\gamma > 0$ is the tail parameter, if and only if the survival function $x \mapsto \mathbb{P}(X_1 > x)$ is regularly varying of index $1/\gamma$ and one can choose $b_n = 0$ and (a_n) verifying $n\mathbb{P}(X_1 > a_n) \to 1$ as $n \to +\infty$.*

Theorem A.2.7 is the core for many statistical approaches to model the tail of the distribution from observed values. Another approach is built on exceedances type methods justified by the next theorem below. It relates the maximum domain of attraction of the limit in (A.2.1) with the family of generalized Pareto distributions. They key idea is to notice that if X is a random variable with distribution F and

$$F^n(a_n x + b_n) \to G(x), \quad x \in \mathbb{R}, \tag{A.2.3}$$

at the continuity points x of a non-degenerate distribution G, for constants $a_n > 0$, $b_n \in \mathbb{R}$, then the relation (A.2.3) holds if and only if

$$n(1 - F(a_n x + b_n)) \to -\log G(x), \quad x \in \mathbb{R}, \tag{A.2.4}$$

at each continuity point x of G for which $0 < G(x) < 1$. This is Theorem 1.1.2 in [80]. We rephrase below Theorem 3.4.5 in [59] or Theorem 1.1.6 in [80]; see also [78].

We state below the Pickands-Balkema-de Haan Theorem for exceedances modeling.

Theorem A.2.9. *(Pickands-Balkema-de Haan Theorem) The condition in* (A.2.1) *holds with $G = G_\gamma$, for $\gamma \in \mathbb{R}$, if and only if there exists a measurable function $a : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that*

$$\lim_{u \uparrow u^*} \mathbb{P}(X_1 - u \le x\, a(u) \mid X_1 > u) \;=\; H(x) \;=\; 1 - (1 + \gamma x/\sigma)_+^{-1/\gamma} \quad x \in \mathbb{R}, \tag{A.2.5}$$

where $u^* := \sup\{u : \mathbb{P}(X_1 \le u) < 1\}$ and H belongs to the family of generalized Pareto distributions

$$H(x) \;=\; H_\gamma(x; \mu, \sigma) \;=\; 1 - (1 + \gamma \tfrac{(x-\mu)_+}{\sigma})_+^{-1/\gamma}, \tag{A.2.6}$$

the case $\gamma = 0$ is the limit as γ approaches zero in (A.2.6).

A.2.3 Inference beyond observed values

The Fisher-Tippett-Gnedenko and Pickands-Balkema-de Haan theorems gather the foundation of most statistical univariate practices in extreme value analysis to extrapolate the tail distribution from observations. They lead to the block maxima and peaks over threshold approaches, respectively. Classical interpretations of these methods for stationary univariate time-dependent scenario lean on the extremal index modeling approach introduced in the paper [109] and the book [110].

Statistical implementation is discussed in Chapter 5 [34]. We are often interested in fixed quantities from the tail model like an extreme quantile outside the range of observations. In the hydrology community, we want to answer in 50-years, which rainfall record will we reach? This value is called the 50-years return level. Let us investigate how to implement classical approaches to compute return levels. We choose the Hyeres weather station in the South of France; see Section 1.2 for illustrating high quantiles statistical inference and uncertainties assessment.

Block maxima method

The block maxima approach follows the 1-5 steps:

1. Identify disjoint blocks,
2. keep only the largest record from each block as in Figure A.1 (in black),
3. fit a generalized extreme value distribution G_γ as in A.2.2,
4. compute the extremal index θ satisfying

$$(P(X_1 \leq xa_n))^{n\,\theta} \to G_\gamma(x), \quad n\mathbb{P}(X_1 > a_n) \to 1, \quad n \to \infty. \tag{A.2.7}$$

5. Extrapolate high quantiles and confidence intervals from (A.2.7).

We can evidence in Figure A.1 the impact from time dependence of extremes in the implementation of the block maxima method. Heavy storms that propagate for several days lead to high levels recorded on consecutive days, i.e., clusters of numerous high records in short periods. Then, the block maxima method tends to discard large values, as can be seen in Figure A.1, where several exceedances of a large order statistic are omitted and instead various lower levels are kept. In this manner, we interpret equation (A.2.7) as a bias correction. We can also read the extremal index in (A.2.7) as a measure of the clustering in time phenomenon.

Peaks over threshold method

The peaks over threshold approach follows the 1-5 steps:

1. Identify the exceedances amounts as in Figure A.2 (in black)
2. Identify clusters of numerous exceedances,
3. keep only one record from each cluster and compute its exceeded amount,
4. fit generalized Pareto distribution: H_γ as in (A.2.6),
5. extrapolate high quantiles and confidence intervals.

For the exceedances approach, Figure A.2 illustrates again the clustering phenomena we evoked. Exceedances happen as short periods with various heavy rainfall levels. In this setting, keeping all exceedances as in the iid case, lead to erroneous confidence intervals since observations due to

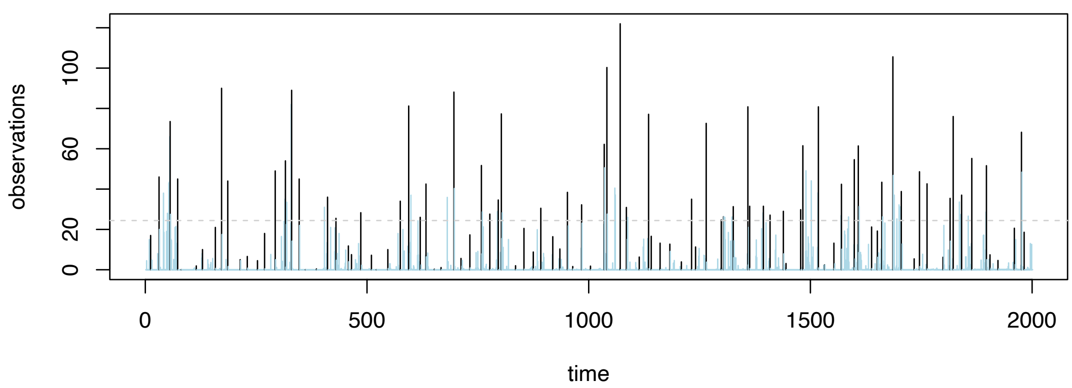


Figure A.1: Time series of fall daily rainfall measures from one weather station in the South of France. In black observations kept from the block maxima method from blocks of length 20. The dotted line the 95th order statistic of the sample.

intra-clustering dependence. For this reason, a common practice is to discard the information of all except one record from each cluster in step 3. Detecting the clusters is already a difficult task, for example, the work in [65] built the bridge between clustering detection and the extremal index as defined in (A.2.7). Furthermore, once we achieve accurate cluster identification, there is still the complicated question of which record should we keep from each cluster? Many practical studies keep the peak cluster observation; cf. [34]. However, this strategy is only justifies when the extremal index equals one and yields biased estimates otherwise; cf. [60, 63].

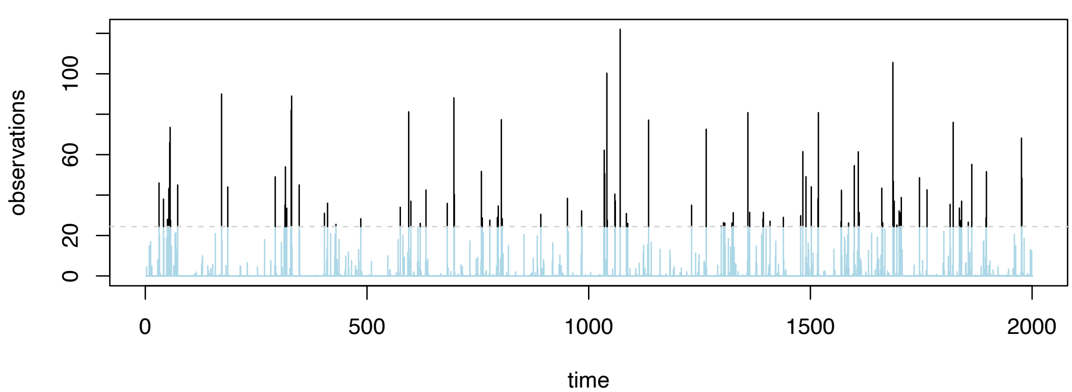


Figure A.2: Time series of fall daily rainfall measures from one weather station in the South of France. In black exceeded amounts kept from the peaks over threshold method. The dotted line the 95th order statistic of the sample.

A.2.4 Moderate levels and tail inference

Extrapolation beyond the observed values is challenging and requires a rigorous theoretical modeling. We can also be interested in intermediate values of the tail distribution. To set an example, notice that implementation of the exceedances method starts with the choice of a high empirical quantile. This random choice brings new uncertainties that we can assess by modeling moderate quantiles. Another quantity we can infer from moderate values is the (tail)-index $\alpha > 0$. It plays a crucial role in weighting the tail's heaviness and determines the return levels' magnitude. We can estimate it using the block maxima, or exceedances maximum-likelihood fit since it links to the tail parameter inhere; see Theorem A.2.8. We call this maximum-likelihood estimator. Though, this needs simultaneous estimation of the tail and shape parameter which can convey practical issues. Another approach is Hill-based estimation from the original work in [85]. This section will formulate the classical results on moderate values inference and Hill estimation of the (tail)-parameter.

We start with a brief heuristic for the iid setting. We consider observations $(X_t)_{t=1,\dots,n}$ identically distributed as X_1, such that X_1 has regularly varying distribution F of index $\alpha > 0$. We consider the quantile function $a(\cdot) : \mathbb{R} \to \mathbb{R}$ defined by

$$a(x) \;=\; F^{\leftarrow}(1 - x^{-1}), \tag{A.2.8}$$

and we consider the order statistics of the sample to be defined as

$$X_{(1)} \;\geq\; X_{(2)} \;\geq\; \cdots \;\geq\; X_{(n)}. \tag{A.2.9}$$

Typically, for $k \in \mathbb{N}$, $X_{(k)}/a(n/k) \;\xrightarrow{\mathbb{P}}\; 0$; as $n \to +\infty$. Letting $k = k_n$ such that $n/k_n \to +\infty$, $k_n \to +\infty$, as $n \to +\infty$ allows to estimate moderate quantiles. Asymptotic normality is shown in Theorem 2.2.1 below. In this setting, refering to Theorem A.2.8, we define the Hill estimator of the tail parameter $\gamma = 1/\alpha$ as

$$\widehat{\gamma}(k) \;=\; \frac{1}{k}\sum_{t=1}^{k} \log X_{(t)} - \log X_{(k-1)}. \tag{A.2.10}$$

We detail below an heuristic argument justifying the Hill estimator. Recall for regularly varying F,

$$\frac{1 - F(tx)}{1 - F(x)} \;\to\; t^{-\alpha}, \quad x \to +\infty.$$

Moreover, the tail quantile function $a(\cdot)$ in (A.2.8) satisfies $a(x) \to +\infty$ and $x(1 - F(a(x)) \to 1$ as $x \to +\infty$. Then, for $t > 0$,

$$\Big(1 - \frac{1}{t}\Big) \;\approx\; \frac{1 - F(\,a(n/t)\,)}{1 - F(\,a(n/(t-1))\,)} \;\approx\; \Big(\frac{a(n/t)}{a(n/(t-1))}\Big)^{-\alpha}, \quad n \to +\infty.$$

Then, applying logarithm at both sides of the equation and replacing $a(n/t)$ by its estimate $X_{(t)}$ yields

$$-\frac{\log(1-t^{-1})}{\alpha} \;\approx\; \log X_{(t)} - \log X_{(t-1)}, \quad n \to +\infty.$$

Moreover, $-t\log(1-t^{-1}) \to 1$ as $t \to +\infty$. Then, consider a sequence (k_n) such that $k_n \to +\infty$, $n/k_n \to +\infty$ as $n \to +\infty$. In this case, taking the average over the k largest order statistics leads

$$\begin{aligned}\gamma \;=\; \frac{1}{\alpha} \;\approx\; & \frac{1}{k}\sum_{t=1}^{k} \; t\log X_{(t)} - t\log X_{(t-1)} \\ =\; & \frac{1}{k}\sum_{t=1}^{k} \; \log X_{(t)} - \log X_{(k-1)}, \quad n \to +\infty.\end{aligned}$$

with the convention that for $t=1$ the summand is simply $\log X_{(1)}$. where the last equality holds by developing the telescopic sums.

To sum up, sufficient conditions for granting the approximations above in the iid case can be found in Theorem 2.4.1. and Theorem 3.2.5 [80] that we reformulate below.

Theorem A.2.10. *Let $(X_t)_{t=1,\dots,n}$ be iid observations with regularly varying distribution F. Let $a(\cdot)$ be as in (A.2.8). Assume F satisfies the second-order condition for $\rho \le 0$: there exists a nonnegative function $A(\cdot)$ such that $\lim_{x\to+\infty} A(x) = 0$ and for $t > 0$,*

$$\lim_{x\to+\infty} \frac{\frac{a(tx)}{a(t)} - t^{1/\alpha}}{A(x)} \;=\; t^{1/\alpha}\,\frac{t^{\rho}-1}{\rho}\,.$$

Consider (k_n), such that for $k = k_n$, $k \to +\infty$, $n/k \to +\infty$, $\sqrt{k}A(n/k) \to \lambda$. Following the notation in (A.2.9) and (A.2.10), then

$$\sqrt{k}\,\Big(\frac{X_{(k)}}{a(n/k)} - 1\Big) \xrightarrow{d} \qquad \mathcal{N}\big(0,1\big), \tag{A.2.11}$$

$$\sqrt{k}\,\big(\,\widehat{\gamma}(k) - \gamma\,\big) \xrightarrow{d} \mathcal{N}\big(\lambda(1-\rho)^{-1}, \gamma^2\big), \quad n \to +\infty. \tag{A.2.12}$$

for $\lambda \in (0, +\infty)$.

Consistency properties of the quantile and tail index estimators above for stationary, possibly dependent, observation were studied in [87, 141]. An optimal asymptotic normality result for stationary observation is shown in [49], under mild mixing assumptions. Bias correction of the Hill estimator is discussed in [11, 75, 83], and references therein.

A.3 Multivariate extremes

The precedent statistical techniques allow us to infer high quantiles from univariate time series. However, our case study, and further examples from spatial data, come in a multivariate format. We aim to study the extremes from the network of rainfall measures from weather stations in France. In the context of natural/climate events, we define risk as a compound, interconnected, interacting, and cascading event [69, 137]. For example, in hydrology, the cause of river flooding is often due to high rainfall amounts recorded simultaneously at multiple close-by stations from the national network [37]. Similarly extremal sea levels, temperatures and wind speeds have been studied from a spatial perspective; see [35]. Also from an economics point of view, the price returns of a basket of multiple assets need to avoid recording multiple extremes in periods of high volatility [163]. Moreover, embedding univariate time series in a multivariate framework might also help assess the time dependencies

Multivariate extreme value theory undertakes two main challenges. First, multivariate extreme value distributions no longer belong to a parametric family Second, but closely related, depending on the application, we can learn the extremal multivariate aspect from different perspectives. For example, we can study the vector of component-wise maxima or aggregate all the d-coordinates and study its univariate tail distribution. These both lead to possible definitions of spatial extreme episodes. Indeed, the family of d-dimensional extreme events has to be broader than the interval-type sets, typical of univariate problems. It should be flexible if we are interested in extremes where at least one coordinate is large or extremes where all coordinates are large or extremes issue from aggregating with sum-type functionals the coordinates, or further $\mathbb{R}^d$-functionals.

From the univariate setting, we saw regular variation is well-suited or studying heavy-tails in Theorem A.2.8. Therefore, a good program for a multivariate extension is to define regular variation of vectors. This approach was pursued in [140, 143] who built upon the generalization of regular variation using the notion of vague converges of measures. Instead, we state an equivalent version in terms weak convergence of probability measures using polar decomposition arguments [84, 152]. It highlights the random aspects of spatial extremes, establishes the main features of regular variation, and allows to extends the Fisher-Tippett-Gnedenko. We state it below and then explain the advantages of characterizing heavy-tailed extremes through this definition in detail below.

In terms of notation, we write variables taking values in $\mathbb{R}^d$ in bold, and we endow the space $\mathbb{R}^d$ with a norm $|\cdot|$. We focus on $\ell^p(\mathbb{R}^d)$ norms: $|\cdot| : \mathbf{x} \mapsto (|x_1|^p + |x_2|^p + \cdots + |x_d|^p)^{1/p}$, for $p > 1$ where $p = \infty$ refers to the supremum norm. Operations in $\mathbb{R}^d$ like $\mathbf{x}+\mathbf{y}$, $\mathbf{xy}$, $\mathbf{x} \leq \mathbf{y}$ or $\mathbf{x} \geq \mathbf{y}$, $(\mathbf{x})_+$, are taken component-wise unless stated otherwise. In particular, we write

$$\mathbb{P}(\max_{t=1,\ldots,n} \mathbf{X}_t \leq \mathbf{x}) \;=\; \mathbb{P}\big(\max_{t=1,\ldots,n} X_{t,1} \leq x_1, \ldots, \max_{t=1,\ldots,n} X_{t,d} \leq x^d\big).$$

If $\mathbf{X}$ is a random d-dimensional vector with distribution F, then the coordinate X_j is distributed as F_j, $j = 1, \ldots, d$. We also write $g^{\leftarrow}(\cdot)$ for the left inverse of a non decreasing function $g : \mathbb{R} \to \mathbb{R}$.

A.3.1 Multivariate regular variation

Definition A.3.1. *An $\mathbb{R}^d$-valued random variable $\mathbf{X}$ is regularly varying of (tail)-index $\alpha > 0$ if there exists a random variable Θ satisfying $|\Theta| = 1$ a.s. such that*

$$\mathbb{P}(|\mathbf{X}| > tx, \mathbf{X}/|\mathbf{X}| \in \cdot \mid |\mathbf{X}| > x) \xrightarrow{w} t^{-\alpha}\, \mathbb{P}(\Theta \in \cdot), \quad x \to +\infty. \tag{A.3.13}$$

The (tail)-index α, and the random variable Θ are informative of the magnitude and spatial distribution of extremes, respectively. Decomposing the extreme event into its magnitude and the spatial features yields to asymptotically independent component. The measure Θ is known as the spectral tail measure and takes values in the d-dimensional sphere $\mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$.

Remark A.3.2. *The closure of sums also holds for multivariate regular variation as in the univariate case; see Remark A.2.5. For the products closure, we must be careful. A Breiman's lemma extension is given in Proposition 5.1. in [6].*

Remark A.3.3. *Concerning a Cramèr-Wold type devise for the weak limit in (A.3.13), regular variation with index $\alpha > 0$ of $\mathbf{u}^\top \mathbf{X}$, for all $\mathbf{u} \in \mathbb{R}^d$ is equivalent to (A.3.13) only under certain conditions. For example, if α is a positive non-integer or if $\mathbf{X} \in [0, +\infty)^d$ and α is an odd integer; cf. Theorem 1.1. in [7].*

Any random variable Θ in the d-dimensional sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| = 1\}$ is a spectral tail measure, but notice that its support may not be the whole sphere. Two opposite behaviors capture much attention due to their interpretation. We present them in the following examples.

Example A.3.4. *(Full asymptotic independence) Let Θ be a random variable in $\mathbb{S}^{d-1}$ defined by $\mathbb{P}(\Theta \in \cdot) = \sum_{i=1}^d p_i \delta_{e_i}$, where $(p_i) \in (0,1)^d$ satisfy $\sum_{i=1}^d p_i = 1$, $\delta_\cdot$ is the Dirac measure, and e_i is the vector in $\mathbb{R}^d$ taking the value 1 at the i-th coordinate and zero elsewhere.*

Example A.3.5. *(Full asymptotic dependence) Let Θ be a random variable in $\mathbb{S}^{d-1}$ satisfying $\mathbb{P}(\Theta_j > 0, j = 1, \ldots, d) = 1$.*

Briefly speaking, asymptotic independence models vectors where two coordinates can not be large simultaneously, whereas all coordinates must be large for full asymptotic dependence. Definition A.3.1 characterizes the extremal properties of a regularly varying vector $\mathbf{X}$ after the polar decomposition $\mathbf{X} \mapsto (|\mathbf{X}|, \mathbf{X}/|\mathbf{X}|)$. Further, we recover the extremal behavior of $\mathbf{X}$ over a large family of sets following Proposition 3.1. in [152].

Theorem A.3.6. *An $\mathbb{R}^d-$valued random variable $\mathbf{X}$ is regularly varying with (tail)-index $\alpha > 0$ if and only if there exists a Borel measure μ in $\mathbb{R}^d$ and a regularly varying function V with (tail)-index $\alpha > 0$ such that*

$$\tfrac{1}{V(x)} \mathbb{P}(x^{-1}\mathbf{X} \in A) \to \mu(A), \quad x \to +\infty. \tag{A.3.14}$$

where $A \subset \mathbb{R}^d \setminus \{\mathbf{0}\}$ *and* $\mathbf{0} \notin \overline{A}$ *for continuity sets* A*. Moreover, we can take* $V(x) := \mathbb{P}(|\mathbf{X}| > x)$ *yielding*

$$\mu(\cdot) := \int_0^\infty \mathbb{P}(y\Theta \in \cdot)d(-y^{-\alpha}), \tag{A.3.15}$$

such that Θ *is the spectral tail measure satisfying* $|\Theta| = 1$ *a.s.*

Remark A.3.7. *The equivalent definition of multivariate regular variation in* (A.3.14) *appeals to the notion of vague convergence in [143], and* M_0 *convergence in [89]. Both definitions coincide in* $(\mathbb{R}^d, |\cdot|)$*, but* M_0 *convergence has the advantage that it extends to further metric spaces. This generalization allows to study extremes in the space of continuous functions in* $[0,1]$ *with the supremum norm, or càdlàg function in* $[0,1]$ *with Skorokhod* J_1*-topology, already considered in [114].*

The limit Borel measure appearing in the limit form of equation (A.3.14) details the extremal properties of a regularly varying vector $\mathbf{X}$. It is known as the exponent measure and satisfies the homogeneity relation $\mu(t\cdot) = t^{-\alpha}\mu(\cdot)$. Regular variation again characterizes the maximum domain of attraction of multivariate heavy-tailed distributions. We review this result in the theorem below; see Proposition 5.11 in Resnick [140].

Theorem A.3.8. *An* $\mathbb{R}^d$*-valued random variable* $\mathbf{X}$ *is regularly varying if and only if* $\mathbf{X}$ *has margins in the Fréchet maximum domain of attraction and there exists a distribution* G *satisfying*

$$\mathbb{P}(\max_{t=1,\dots,n} \mathbf{X}_t \leq a_n\mathbf{x} + b_n) \to G(\mathbf{x}), \quad n \to +\infty, \tag{A.3.16}$$

for a positive and real-valued sequences (a_n)*,* (b_n)*. respectively, such that* G *has non-degenerate margins. In this case,* $G(\cdot) = \exp\{-\mu(\cdot)\}$*, where the measure* μ *is the exponent measure in equation* (A.3.15)*.*

From the discussion above, regular variation is well suited for studying extremes from heavy-tailed distributions with the same tail index $\alpha > 9$. The expression in (A.3.13) highlights the important role of α and Θ for modeling rare events. In particular, any random variable Θ taking values in the d-dimensional sphere yields to an extreme value distribution G as in (A.3.16). Parametric models were studied for bivariate and then multivariate cases in [70, 134, 158], and then [77, 81, 165, 35]. The following are classical examples.

Example A.3.9. *(Logistic model)*

$$G(\mathbf{x}) := \exp\big\{-(\textstyle\sum_{i=1,\cdots,d} x_i^{-1/\tau})^\tau\big\}, \quad \tau \in [0,1].$$

Example A.3.10. *(Asymmetric logistic model)*

$$G(\mathbf{x}) := \exp\big\{-\textstyle\sum_{I\subseteq\{1,\cdots,d\}}(\sum_{i\in I}(\frac{\theta_{i,I}}{x_i})^{1/\tau_I})^{\tau_I}\big\},$$

where $\tau_I \in [0,1], \theta_{i,I} \geq 0$ *and* $\sum_{i\in I}\theta_{i,I} = 1$ *for all* $I \subseteq \{1,\cdots,d\}$*.*

Consider now a random vector $\mathbf{X}$ in $\mathbb{R}^d$ with marginals in any domain of attraction. We define multivariate extreme value distributions as the limits G with non-degenarate margins satisfying

$$\mathbb{P}(\max_{t=1,\dots,n} \mathbf{X}_t \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}), \quad n \to +\infty, \quad \mathbf{x} \in \mathbb{R}^d, \tag{A.3.17}$$

for $(\mathbf{a}_n) \in \mathbb{R}^d_{>0}$, $(\mathbf{b}_n) \in \mathbb{R}^d$. Notice this is a broader family than the limits from equation A.3.16. Also, G has multivariate extreme value margins which are thus continuous. Then, as in the univariate setting, for i.i.i. observations $(\mathbf{X}_t)$, the relation (A.3.17) holds if and only if

$$n\{1 - F(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n)\} \to -\log G(\mathbf{x}), \quad n \to +\infty, \quad \mathbf{x} \in \mathbb{R}^d, \tag{A.3.18}$$

for any $\mathbf{x}$ such that $0 < G(\mathbf{x}) < 1$. We refer the details to ([80], chapter 6). Then a natural way of defining multivariate generalized Pareto distributions is by conditioning on at least one coordinate being large. This limit of component-wise exceeded amounts was studied in [149, 12]. It generalizes the exceedances approach from the Pickands-Bakema-De Haan theorem in A.2.9, as stated in the next theorem.

Theorem A.3.11. *Let $\mathbf{X}$ be a $\mathbb{R}^d$-valued random vector with distribution F. The condition (A.3.17) holds if and only if there exists a d-dimensional random vector $\mathbf{W}$ such that*

$$\mathbb{P}(\mathbf{a}_n^{-1}(\mathbf{X} - \mathbf{b}_n)_+ \in \cdot \,|\, \max_{j=1,\cdots,d}|X_j| > b_n) \xrightarrow{d} \mathbb{P}((\mathbf{W})_+ \in \cdot\,), \quad n \to +\infty \tag{A.3.19}$$

where $(\mathbf{a}_n)$ and $(\mathbf{b}_n)$ are chosen to ensure that $0 < G_j(0) < 1$, $j = 1, \dots, d$. and H belongs to the family of multivariate generalized Pareto distributions. Moreover,

$$\mathbb{P}((\mathbf{W})_+ \leq \mathbf{x}) = \tfrac{1}{-\log G(\mathbf{0})} \log \tfrac{G(\mathbf{x})}{G(\mathbf{0})}, \quad \mathbf{x} \geq 0.$$

We confer further details of the exceedances approach to [147, 148, 105], where properties of this family of distribution and inference procedures are discussed.

The multivariate extreme value and the generalized Pareto distributions are popular since they completely characterize the three maximum domains of attraction: Weibull, Gumbel, and Fréchet. In comparison, regular variation focuses on modeling the heavy-tailed phenomena. However, for lighter tails, or coordinates with different (tail)-index, it is common to transform margins first. For example, the rank transform: $X_j \mapsto 1/(1 - F_j(X_j))$ yields to a vector with unitary Fréchet margins which allows to recover the regularly varying framework; see [12]. This practice is justified by the proposition below.

Proposition A.3.12. *Any multivariate extreme value distribution G defined from equation (A.3.17) can be written as*

$$G := G^*(\varphi_1(x_1), \cdots, \varphi_d(x_d))$$

where G^* is a multivariate extreme value distribution with unitary Fréchet margins and $\varphi_j(\cdot) := (1/(1 - G^*_j))^{\leftarrow}(\cdot)$.

For a proof see Proposition 5.10 in [140]. To conclude, in the multivariate setting, learning the marginal features and the spatial structure suffices to describe the maximum domain of attraction. The spatial aspect of extremes is carried by the spectral tail measure of the transformed vector to unitary Fréchet. We recover from Proposition A.3.12 the important role played by the spectral tail measure defined in (A.3.13). Further on, we consider uniquely heavy-tailed models.

Remark A.3.13. *Another way of characterizing extreme value distributions is through the notion of max-stability. We recall a distribution G is max-stable if for all $n \in \mathbb{N}$, there exists constants in $\mathbb{R}^d$ $\mathbf{a}_n > 0, \mathbf{b}_n \in \mathbb{R}$ satisfying*

$$\mathbb{P}(\max_{t=1,\dots,n} \mathbf{Y}_t \leq \mathbf{x}) = \mathbb{P}(\mathbf{Y}_1 \leq \mathbf{a}_n \mathbf{x} + \mathbf{b}_n),$$

where $(\mathbf{Y}_t)$ is an i.i.d. sequence distributed as G. The limit distributions from (A.3.16) are max-stable. In the univariate case, max-stable distribution coincide with extreme value distributions; see [140]. In the multivariate case, extreme value distributions coincide with max-stable distributions with non-degenerate marginals. This result was reviewed in Proposition 5.9 in Resnick [140].

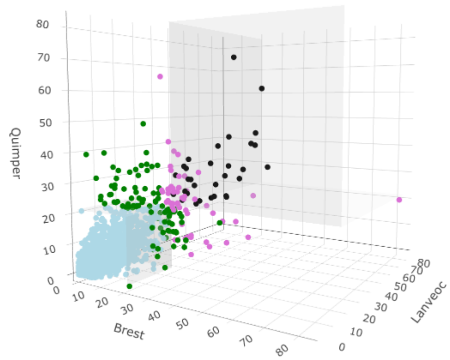
A.3.2 Statistical methods for multivariate extremes

Let $(\mathbf{X}_t)_{t=1,\dots,n}$ be observations in $\mathbb{R}^d$ identically distributed. Two challenges arise for learning the extremal properties of $\mathbf{X}_1$. First, inferring its marginal aspects, and second, inferring its extremal spatial features. The first task is typically assessed using the methods from the univariate chapter. For the second one, we can assume $\mathbf{X}_1$ is regularly varying with (tail)-index $\alpha > 0$. Indeed, this assumption holds if we have transformed our margins first, for example to unitary Fréchet with the rank transformation. It is also a reasonable assumption if all coordinates reach high levels similarly, at the same magnitude, and we can assume the heavy phenomenon has the same driver. Then we can model tail margins to be equivalent up to a constant.

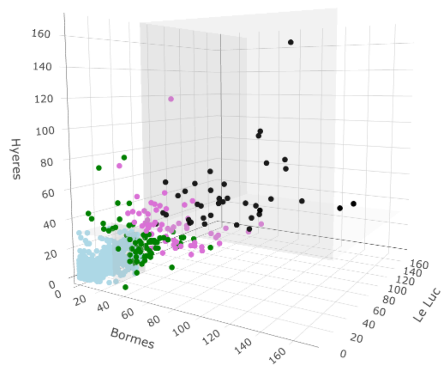
For our case study, high rainfall measures have the same magnitude and reach high levels at a similar rate whenever the weather stations are close. This is well illustrated in Figure A.3. Also, climate scientists agree that precipitations behave like a heavy-tailed phenomenon [102]. Indeed, we can think the meteorological event impacting close locations has the same source but is manifested at different time lags and locations as it travels its course. This justifies modeling daily rainfall amounts from each region as a multivariate regularly varying random vector given by $\mathbf{X}_1 = (X_{1,1}, X_{1,2}, X_{1,3})$. The main challenges of multivariate extremes then resumes to inferring features of the spectral tail measure Θ, and the tail index $\alpha > 0$. In terms of notation, we write the d-dimensional observation at time t as $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,d})$.

As mentioned previously, the domain of the spectral tail measure Θ might not be the whole sphere. Inspecting its domain can help reduce the dimension of the problem at an early stage. One way to do so is partitioning the unit sphere $\mathbb{S}^{d-1}$ into its faces

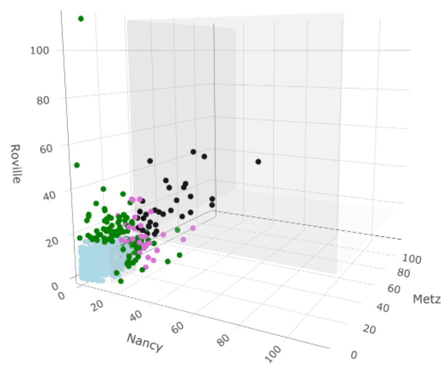
$$\mathbb{S}^{d-1}_{\{I\}} = \{\mathbf{x} \in \mathbb{S}^{d-1} : x_i > 0 \text{ if } i \in I, x_i = 0, \text{ if } i \notin I\} \tag{A.3.20}$$



(a) Northwest



(b) South



(c) Northeast

Figure A.3: Scatterplot of fall daily rainfall records from regions in France with Oceanic, Mediterranean and Continental weather in (a), (b), (c), respectively. We fix the 95th-order statistic of the norm sample at each region and then color points to distinguish records. Measures whose supremum norm does not exceed this threshold value are in blue. Records exceeding the threshold at one, two, and three coordinates are colored green, pink, and black, respectively.

.

with $I \subset \{1, \cdots, d\}$. As in the examples A.3.4 and A.3.5, this partitioning of the sphere aims to describe the coordinates of a regularly varying vector $\mathbf{X}$ that are large simultaneously. Typically, this restricts the problem to moderate dimensions. We want to recover the sparsity structure intrinsic to the measure Θ, which only puts mass in specific faces. We argue that this is crucial for both parametric and non-parametric modeling. It helps reduce the burden of dimensionality for non-parametric models and, for classical parametric models, it justifies assuming the spectral tail measure is fully asymptotically dependent. It suffices to restrict the analysis to the joint distribution over faces with positive probability. Both Example A.3.9 and Example A.3.10 model full asymptotic dependence.

For a regularly varying random vector $\mathbf{X}$, the projection $\mathbf{X}/|\mathbf{X}|$ behaves asymptotically as Θ as $|\mathbf{X}|$ reaches larger values; see (A.3.1). We use this heuristic to obtain an empirical approximation from a sample of Θ. For our precipitation data-set, the empirical sample of the spectral measure for exceedances over the 95-th quantile are plotted in Figure A.4. We can see a good proportion of points, in particular black points, lying close to $\mathbb{S}^2_{1,2,3} := \{(1,1,1)\}$; and also a significant amount of points lying around $\mathbb{S}^d_{1,2} \cup \mathbb{S}^d_{1,3} \cup \mathbb{S}^d_{2,3} := \{(x,y,z) \in \mathbb{S}^2 : (x,y) = 1 \text{ or } (x,z) = 1 \text{ or } (y,z) = 1\}$, for the majority pink. The previous sets are the faces defined in (A.3.20). We conclude that in our tri-variate analysis there is no evidence of extremal asymptotic independencies.

Measuring extremal spatial dependence

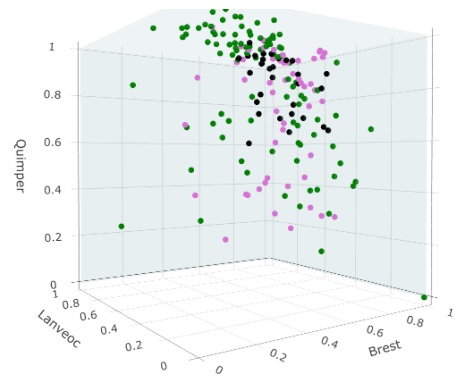
The takeaway message is that we can look at statistics of the spectral measure Θ to study the behavior of spatial extremes. Non-parametric estimation based on the subsample consisting of $\mathbf{X}_{(1)}/|\mathbf{X}_{(1)}|, \cdots, \mathbf{X}_{(k)}/|\mathbf{X}_{(k)}|$, is discussed in [143], where (k_n) satisfies, for $k = k_n$, $k \to +\infty, n/k \to +\infty$. Non-parametric estimation is discussed in [35, 162, 36, 111]. Further, summary statistics of extremal spatial dependence include the Pickands dependence function, the index of extremal dependence, the upper tail dependent coefficients are examples, which we can all write in terms of Θ. We refer to [71] for further reference on measures of spatial dependence. We only detail on the upper tail dependent coefficient.

Definition A.3.14. *(Tail dependent coefficient) Let $(\mathbf{X}_t)$ be a regularly varying time series in $(\mathbb{R}^d, |\cdot|)$ with (tail)-index $\alpha > 0$, and spectral measure Θ; see* (A.3.13). *We define the upper tail dependent coefficient by*

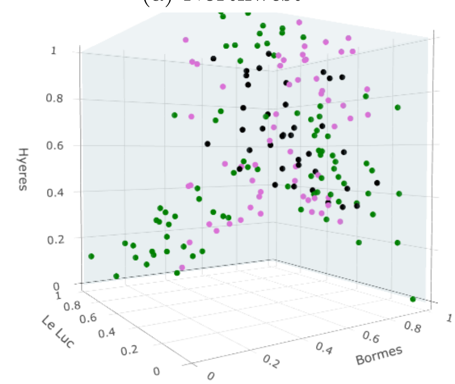
$$\chi_{i,j} \;:=\; \lim_{x \to +\infty} \mathbb{P}(X_i > x \mid X_j > x) \;=\; \frac{\mathbb{E}[(\Theta_i)_+^\alpha \wedge (\Theta_j)_+^\alpha]}{\mathbb{E}[(\Theta_j)_+^\alpha]}, \quad i,j = 1, \ldots, d.$$

In the bivariate case, the case $\chi_{1,2} = 0$ is equivalent to full asymptotic independence of extremes and $\chi_{1,2} = 1$ is asymptotic dependence of extremes.

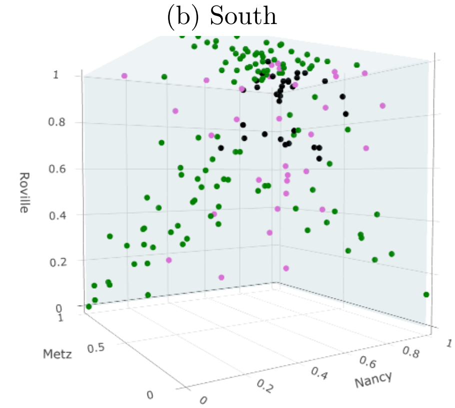
Example A.3.15. *For the logistic model with parameter $\tau \in [0,1]$, as in Example A.3.9, $\chi_{i,j} = 2 - 2^\tau$.*



(a) Northwest



(b) South



(c) Northeast

Figure A.4: Projection of fall daily rainfall exceedances of the 95th-order statistic to the supremum sphere: $\mathbf{X} \mapsto \mathbf{X}/|\mathbf{X}|$. Panels correspond to Oceanic, Mediterranean and Continental weather in France in (a), (b), (c), respectively. Records exceeding the threshold at one, two, and three coordinates are colored green, pink, and black, respectively, as in Figure A.3.

	Brest	Lanveoc	Quimper	Bormes	Le Luc	Hyeres	Nancy	Metz	Roville
Brest	-	**0.73**	**0.29**	0.01	0.01	0.01	0.11	0.09	0.10
Lanveoc	**0.39**	-	**0.12**	0.02	0.02	0.01	0.05	0.02	0.07
Quimper	**0.40**	**0.33**	-	0.02	0.01	0.01	0.12	0.13	0.07
Bormes	0.04	0.12	0.08	-	**0.33**	**0.38**	0.00	0.00	0.00
Le Luc	0.03	0.11	0.03	**0.29**	-	**0.55**	0.06	0.06	0.04
Hyeres	0.02	0.03	0.02	**0.21**	**0.36**	-	0.04	0.00	0.03
Nancy	0.04	0.03	0.04	0.00	0.01	0.01	-	**0.38**	**0.32**
Metz	0.04	0.01	0.05	0.00	0.01	0.00	**0.46**	-	**0.20**
Roville	0.07	0.07	0.04	0.00	0.01	0.01	**0.62**	**0.34**	-

Table A.2: Table of estimates of the tail dependence coefficients $\chi_{i,j}$ for fall daily rainfall between stations in France.

A.4 Limitations of classical approaches

Bivariate and multivariate extensions of extreme value theory in the i.i.d. case was undertaken initially in the work of [70, 134, 158], and then [81]. A context with both multivariate and time-dependent extreme assessment is less common. Most solutions so far presented address either time or space dependence. Also, they seldom consider a point of view outside the main heuristic of maxima of blocks and exceedances techniques which are the most common between practitioner for high quantile inference which refers to **(Quest. 4)**.

To sum up, the extremal dependence in time has been analyzed previously mainly from the extremal index point of view. Since the original work in [109, 110], the current leading branches of the domain concern estimation of the extremal index, from a practical point of view, and formalizing the notion of cluster of exceedances, from a theoretical point of view. We refer to the recent work of [65, 155, 159, 150, 129, 16, 21, 25] as for inference of the extremal index. For the mathematical theory of cluster of exceedances a list of recent publications includes [40, 10, 9, 138, 28]. In this manner, the cluster of exceedances captures short periods with consecutive large values and is one classical approach addressing **(Quest 1.)**. In the same line, inference of the extremal index tackles **(Quest 3.)** since it has interpretation as a summary statistic of the clustering of heavy records phenomenon. Nevertheless, estimating the extremal index still demonstrates to be a difficult task. To overcome these challenges we further discuss large deviations of heavy-tailed time series in Chapter B. Concerning **(Quest 2.)**, typically, to select extreme observations from the sample, we estimate moderate thresholds using order statistics. However, for dependent observations, this strategy brings new uncertainties into play, which are already difficult to track in the i.i.d. setting.

Appendix B: Large deviations of heavy-tailed time series

B.1 Outline of Chapter B

In this chapter we drop the independence assumption in time and space. We consider stationary time series $(\mathbf{X}_t)$, taking values in $(\mathbb{R}^d, |\cdot|)$, with multivariate regularly varying finite-dimensional cuts. We study the extremal behavior of $(\mathbf{X}_t)$ focusing on sum-type functionals. In Section B.2 we motivate the study of sums large fluctuations and present the classical theory in the i.i.d. setting of central limit theory and large deviation principles for heavy-tailed observations. We fix notation in Section B.3. In Section B.4 we provide assumptions for extending classical results to stationary heavy-tailed time series, following [40]. Regularly varying time series are defined in Section B.5. Then we review central limit theory for stationary regularly varying time series observations in Section B.6. To state the main results we introduce the clusters of (exceedances) with values in ℓ^∞, implicitly introduced in [40]. In Section B.7 we discuss large deviation principles for stationary regularly varying time series and motivate our definition of cluster processes in ℓ^p, for $p < +\infty$. Cluster inference is discussed in Section B.8. In Section B.9 we review two stationary regularly varying time series models.

B.2 Motivation: fluctuations of sums

B.2.1 Central limit theory for i.i.d. observations

So far, Chapter A on classical extreme value theory studies maxima and exceedances approaches. In this setting, regular variation proved to be a valuable and sufficient notion to survey these operations. Regular variation is also key for studying sum-type functionals. For instance, it appears naturally while looking at the sums fluctuations of i.i.d. observations. We recall the central limit theorem for i.i.d. random variables (X_t) with finite variance $\sigma^2 < +\infty$ and mean $\mu < +\infty$. It states that $(\sum_{t=1}^n X_t - \mu n)/\sqrt{\sigma^2\, n} \xrightarrow{d} \xi$, and ξ is standard normal distributed. This theorem extends to i.i.d. random variable with non-finite second moment, and regular variation also appears naturally as we show next. We start by stating Theorem 2.2.5 in [59].

Theorem B.2.1. *(Central Limit Theorem) Let (X_t) be an i.i.d. sequence of univariate random variables. Assume there exists renormalizing constants $a_n > 0$, $b_n \in \mathbb{R}$, and a non-degenerate distribution L such that*

$$\lim_{n\to+\infty} \mathbb{P}(\textstyle\sum_{t=1}^n X_t \le x a_n + b_n) \;=\; L(x), \quad x \in \mathbb{R}, \tag{B.2.1}$$

then $L = L_\alpha$ belongs to the parametric family of stable distributions, determined by its log-characteristic function, such that, if ξ_α is distributed as L_α, then for $u \in \mathbb{R}$

$$\log\, \mathbb{E}[e^{iu\,\xi_\alpha}] \;=\; \begin{cases} -\sigma^\alpha |u|^\alpha (1 - i\beta\,\mathrm{sign}(u) \tan\frac{\pi\alpha}{2}) + i\mu\, u & \text{if } \alpha \neq 1, \\ -\sigma|u|(1 - i\beta\,\mathrm{sign}(u)\frac{2}{\pi}\,\log|u|) + i\mu\, u & \text{if } \alpha = 1, \end{cases} \tag{B.2.2}$$

where $\alpha \in (0,2]$ is the stable tail parameter, $\sigma \in [0,+\infty)$ is a scale parameter, $\beta \in [-1,1]$ is a skweness parameter, and $\mu \in \mathbb{R}$ is a location parameter.

Theorem B.2.1 can be seen as a version of the Fisher-Tippett-Gnedenko Theorem in A.2.7 tailored for sums. Characterizing the limit distribution of partial sums is a fundamental result of probability theory, and is discussed in classical textbooks as [3, 13, 59, 66], and references therein. Moreover, as in the Fisher-Tippett-Gnedenko theorem, the limit (B.2.1) is in straight correspondence with the regular variation property of the tails. To see this we state below Corollary 2.2.17 in [59] below.

Proposition B.2.2. *Let (X_t) be an i.i.d. sequence of univariate random variables. The relation (B.2.1) holds for L_α, the α-stable distribution with $\alpha \in (0,2]$, if and only if*

- $\alpha = 2$ *and* $\mathbb{E}[(X_1)^2] < \infty$,
- $\alpha \in (0,2]$, $\mathbb{E}[(X_1)^2] = +\infty$, *and there exists non-negative constants p_-, p_+, satisfying $p_- + p_+ = 1$, such that the following tail-balance condition holds*

$$\lim_{x\to+\infty} \mathbb{P}(X_1 < -x) \sim p_-\, x^{-\alpha}\, \ell(x), \qquad \lim_{x\to+\infty} \mathbb{P}(X_1 > x) \sim p_+\, x^{-\alpha}\, \ell(x), \tag{B.2.3}$$

where $x \mapsto \ell(x)$ is a slowly varying function. We call $\alpha > 0$ the (tail)-index of (X_t).

Actually, for regularly varying random variables, there exists a tight link between the maxima and sum fluctuations given by the subexponential property. More precisely, we reformulate below a direct corollary from Lemma 1.3.1. in [59].

Proposition B.2.3. *Let (X_t) be a sequence of univariate i.i.d. random variables. If X_1 is regularly varying, then it verifies the subexponential property:*

$$\lim_{x\to+\infty} \frac{\mathbb{P}(\sum_{t=1}^n X_t > x)}{\mathbb{P}(\max_{t=1,\dots,n} X_t > x)} \;=\; \lim_{x\to+\infty} \frac{\mathbb{P}(\sum_{t=1}^n X_t > x)}{n\,\mathbb{P}(X_1 > x)} \;=\; 1. \tag{B.2.4}$$

The family of stable distributions is characterized by (B.2.2) and also by the limiting property in (B.2.1). It also coincides with the distribution family satisfying the sum-stable property:

$$a_n^{-1}(\textstyle\sum_{t=1}^n X_t - b_n) \;\stackrel{d}{=}\; \xi, \tag{B.2.5}$$

for an i.i.d. univariate sequence (X_t), a random variable ξ, and sequences $a_n > 0$, $b_n \in \mathbb{R}$. Examples of stable distributions are the Gaussian distribution with $\alpha = 2$, the Cauchy distribution with $\alpha = 1$, and $\beta = 0$, and the Lévy distribution with $\alpha = 1/2$ and $\beta = 1$. In this cases, analytical expressions exist for their density functions; see Chapter 1 in [130]. However, in many cases analytical expressions does not exist and only asymptotic expansions are available; see Chapter 3 in [130]. Modeling with stable distributions is common in financial applications to represent stock returns with infinite variance [4, 116]. Similarly they are considered in signal process [128] and network traffic [142]. A literature review of these applications is also given in Chapter 2 [130].

B.2.2 Large deviations for i.i.d. observations

Central limit theory as in Theorem 5.2.1 summarizes the convergence of sums towards its median. Weaker results are also at hand. For example, the strong (weak) law of large numbers provides rates for almost sure (probability) convergence or divergence of mean corrected sums assuming moment conditions. Stronger results also exists refining the central limit theorem. For example, large deviation principles focus on the distribution of mean corrected sums on extremal sets. The main idea is to quantify the convergence rate of its tails by studying the large jumps from the median. Cramér's theory is standard for a tail approximation of finite variance observations. It can be found in a handful of probability textbooks [17, 66]; see also [59, 119] for a review.

For heavier tails, we recall the main large deviation principles derived in A.V. and S.V. Nagaev [125, 124]; and Cline and Hsing [33] see also [121].

Theorem B.2.4. *Consider (X_t) to be univariate i.i.d. random variables satisfying the tail-balance condition in (B.2.3) for p_-, p_+, and with (tail)-index $\alpha > 0$. Then,*

$$\lim_{n\to+\infty} \sup_{x\geq x_n} \Big|\frac{\mathbb{P}(\sum_{t=1}^n X_t - b_n > x)}{n\,\mathbb{P}(|X| > x)} - p_+\Big| + \Big|\frac{\mathbb{P}(\sum_{t=1}^n X_t - b_n \leq -x)}{n\,\mathbb{P}(|X| > x)} - p_+\Big| \;=\; 0. \tag{B.2.6}$$

such that $n\,\mathbb{P}(|X| > x_n) \to 0$ and

$$x_n = \begin{cases} \sqrt{cn\log n}, & \text{if } \alpha \in (2, +\infty), \\ n^{\frac{1}{\alpha}+\kappa} & \text{if } \alpha \in (1, 2], \\ n^{\frac{1}{\alpha}+\kappa} & \text{if } \alpha \in (0, 1], \end{cases} \quad , \quad b_n = \begin{cases} n\mathbb{E}[X_1] & \text{if } \alpha \in (2, +\infty), \\ n\mathbb{E}[X_1] & \text{if } \alpha \in (1, 2), \\ 0 & \text{if } \alpha \in (0, 1] \end{cases} \tag{B.2.7}$$

where c is a constant satisfying $c > \alpha - 2$, and we can take any $\kappa > 0$.

Our motivation for introducing this classical theory here is twofold. First, it is an interesting perspective to quantify risk, particularly risk due to the effect of aggregating observations. Ruin type probabilities are an example of this [59]. We can use large deviation principles to assess the threat of extreme events for regularly varying random variables with tail index $\alpha \in (0, 2)$, and the central limit theorem appears as an alternative to the Fisher-Tippet-Gnedenko for modeling the tails of a distribution. The main drawback from this approach, is the lack of analytical expressions of all but few stable laws. Though, the recent improvement of numerical approximation and computer's power motivate us to look back at stable laws [4, 131]. Furthermore, the regular variation framework establishes a close link between sums and maxima while studying extremes due to the subexponentiality property; see Proposition B.2.3, this suggests heavy-tails can also be modeled through sum-type functionals.

B.3 Notation

In terms of notation, to include the temporal coordinates, we write vectors $(\mathbf{x}_t)_{t=a,\cdots,b} = \mathbf{x}_{[a,b]}$ for any $a, b \in \mathbb{Z}$, and $a \leq b$. For $p \in (0, +\infty]$, $(\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$, we define the $p-$modulus function $\|\cdot\|_p : (\mathbb{R}^d)^{\mathbb{Z}} \to [0, +\infty]$ by $\|\mathbf{x}_t\|_p := (\sum_{t\in\mathbb{Z}} |\mathbf{x}_t|^p)^{1/p}$, and we define the sequential space ℓ^p as sequences with finite p-modulus, with the convention that $p = \infty$ refers to the supremum norm. For $p \in (0, +\infty]$, the p-modulus induce a distance d_p in ℓ^p. If $p \in [1, +\infty]$ the p-modulus defines a norm, and (ℓ^p, d_p) are separable Banach spaces. If $p \in (0, 1)$, then (ℓ^p, d_p) is still a separable metric space.

Our focus will be on shift-invariant functionals on ℓ^p, which is the case of the p-modulus. We define the quotient space $\tilde{\ell}^p = \ell^p/\sim$ by $(\mathbf{x}_t) \sim (\mathbf{y}_t)$ is and only if there exists $k \in \mathbb{Z}$ such that $x_{t-k} = y_t$, $t \in \mathbb{Z}$. We define the metric space $(\tilde{\ell}^p, \tilde{d}_p)$ by

$$\tilde{d}_p([\mathbf{x}_t], [\mathbf{y}_t]) = \inf_{k\in\mathbb{Z}} \{d_p(\mathbf{x}_{t-k}, \mathbf{y}_t) : (\mathbf{x}_t) \in [\mathbf{x}_t], (\mathbf{y}_t) \in [\mathbf{y}_t]\}, \tag{B.3.8}$$

for $[\mathbf{x}_t], [\mathbf{y}_t] \in \tilde{\ell}^p$. In what follows it will convenient to write elements $[\mathbf{x}_t]$ in $\tilde{\ell}^p$ as $(\mathbf{x}_t)$. We sometimes need to embed vectors $\mathbf{x}_{[a,b]}$ in the sequential space $(\mathbb{R}^d)^{\mathbb{Z}}/\sim$, we do this by assigning zeros to undefined coordinates. Details on the shift-invariant spaces as in (B.3.8) are deferred to [26, 9].

For matrices $\mathbf{A} \in \mathbb{R}^{d\times d}$, we define the operator norm as $|\mathbf{A}|_{op} := \sup_{|\mathbf{x}|=1} |\mathbf{A}\mathbf{x}|$. Truncation of $(\mathbf{x}_t) \in (\mathbb{R}^d)^{\mathbb{Z}}$ at the level ϵ, for $\epsilon > 0$, is written as $(\overline{\mathbf{x}}_t^{\varepsilon}) = (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|\leq\varepsilon})$, and $(\underline{\mathbf{x}_t}_{\epsilon}) = (\mathbf{x}_t \mathbb{1}_{|\mathbf{x}_t|>\varepsilon})$.

B.4 Assumptions

In what follows, we consider stationary time series $(\mathbf{X}_t)$ with heavy-tailed finite-cuts allowing for spatio-temporal dependencies. To extend the central limit theory and the large deviation principles to a stationary setting, we follow the ideas in [40]. We introduce next an anti-clustering condition to inhibit long-range dependence of extremes, a vanishing-small-values condition to control the sums of moderate values, and a mixing condition for statistical purposes.

Assume there exists sequences (x_n), (b_n), such that, $n\mathbb{P}(|\mathbf{X}_0| > x_n) \to 0$, $\mathbb{P}(|\mathbf{X}_0| > x_{b_n}) \to 1$, $n/b_n \to +\infty$, as n $\to +\infty$. In this chapter we consider the assumptions below, tailored for these sequences.

- **AC** : Assume for all $\delta > 0$,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{P}(\|\mathbf{X}_{[k,n]}\|_\infty > \delta\, x_n \,|\, |\mathbf{X}_0| > \delta\, x_n) = 0. \tag{B.4.9}$$

- $\mathbf{CS}_1$: Assume for all $\delta > 0$, $\epsilon > 0$, for $\alpha \in [1,2)$,

$$\lim_{\epsilon \downarrow 0} \limsup_{n\to+\infty} \frac{\mathbb{P}(|\sum_{t=1}^n |\overline{X_t/x_n}^\epsilon| - \mathbb{E}[|\overline{X_t/x_n}^\epsilon|]| > \delta)}{n\mathbb{P}(|X_0| > x_n)} = 0. \tag{B.4.10}$$

- $\mathcal{A}(x_{b_n})$: Assume for non-negative continuous bounded function $f : \mathbb{R} \to \mathbb{R}$, vanishing in a neighborhood of the origin,

$$\lim_{n\to+\infty} |\,\mathbb{E}[\exp\{-\textstyle\sum_{t=1}^n f(x_{b_n}^{-1} X_t)\}] - \mathbb{E}[\exp\{-\textstyle\sum_{t=1}^{b_n} f(x_{b_n}^{-1} X_t)\}]^{\lfloor n/b_n \rfloor}\,| \;=\; 0. \tag{B.4.11}$$

Remark B.4.1. *If the assumptions are required simultaneously, we assume the sequence (x_n) appearing in it coincides. The sequence (x_n) in* **AC** *and* $\mathbf{CS}_1$ *controls the extremes of the supremum norm $\|\mathbf{X}_{[0,n]}\|_\infty$ and of sum-type functionals, respectively. In this sense, it has a probabilistic interpretation. Instead, we introduce the sequence (b_n) in $\mathcal{A}(x_{b_n})$ for statistical purposes.*

The first condition has its roots in the work of Leadbetter [109] and Leadbetter *et.al* [110]. It is linked to the maxima approach in Chapter A, and is typically used to justify the extremal index heuristic from equation (A.2.7). The second and third conditions are vanishing-small-values and mixing conditions, respectively, and a version of them was already introduced in [40]. These two are tailored for studying sums of regularly varying innovations. We will use them in Chapter B to state classical central limit theory and large deviation principles for heavy-tailed time series. We discuss thoroughly these assumptions in Chapter 2 and verify them for classical models in Chapter 5.

B.5 Regularly varying time series

We recall next the definition of a stationary regularly varying time series following [10].

Definition B.5.1. *Let $(\mathbf{X}_t)$ be a stationary time series taking values in $(\mathbb{R}^d, |\cdot|)$. The time series is regularly varying with (tail)-index $\alpha > 0$ if for all $h = 0, 1, \ldots,$ there exists a random variable $(\Theta_t)_{[-h,h]}$ satisfying $|\Theta_0| = 1$ a.s. and*

$$\mathbb{P}(|\mathbf{X}_0| > tx, \mathbf{X}_{[-h,h]}/|\mathbf{X}_0| \in \cdot \mid |\mathbf{X}_0| > x) \xrightarrow{w} t^{-\alpha}\mathbb{P}(\Theta_{[-h,h]} \in \cdot), \quad x \to +\infty. \tag{B.5.12}$$

The time series (Θ_t) is a well defined object, following the Kolmogorov's consistency theorem, and we call it the spectral tail process of $(\mathbf{X}_t)$. The spectral tail process does not inherit the initial stationarity property. Instead, we deduce the time-change formula: for any $s, t \in \mathbb{Z}, s \le 0 \le t$ and for any measurable bounded function $f : (\mathbb{R}^d)^{t-s+1} \to \mathbb{R}$,

$$\mathbb{E}[f(\Theta_{s-i}, \ldots, \Theta_{t-i})\mathbb{1}(|\Theta_{-i}| \neq 0)] \;=\; \mathbb{E}[|\Theta_i|^\alpha f(\Theta_s/|\Theta_i|, \ldots, \Theta_t/|\Theta_i|)], \tag{B.5.13}$$

The spectral tail process was introduced in [10] as a tool for modeling multivariate regular variation of finite vectors of the time series; see Theorem 2.1 in [10]. We state this result below using ℓ^p-norms, $p > 0$, over time observations. In fact, for $p < 1$, $\|\cdot\|_p$ is not a norm but a modulus as in [152], but by Proposition 3.1. in [152] this still entails regular variation.

Lemma B.5.2. *Consider a regularly varying time series* $(\mathbf{X}_t)$ *of (tail)-index* $\alpha > 0$. *Then, for any* $p > 0$, $h \geq 0$, $\mathbf{X}_{[0,h]}$ *is regularly varying and satisfies*

$$\frac{\mathbb{P}(\|\mathbf{X}_{[0,h]}\|_p > tx, \mathbf{X}_{[0,h]}/\|\mathbf{X}_{[0,h]}\|_p \in \cdot)}{\mathbb{P}(|\mathbf{X}_0| > x)} \xrightarrow{w} c(p,h)\, t^{-\alpha}\mathbb{P}(\mathbf{Q}^{(p)}(h) \in \cdot), \quad x \to +\infty \tag{B.5.14}$$

where $c(p,h)$ *is a positive constant and* $\mathbf{Q}^{(p)}(h)$ *is the spectral measure of* $\mathbf{X}_{[0,h]}$ *for the* ℓ^p*–norm, i.e.,* $\|\mathbf{Q}^{(p)}(h)\|_p = 1$, *a.s., and*

$$\mathbb{P}(\mathbf{Q}^{(p)}(h) \in \cdot) = \frac{1}{c(p,h)} \sum_{k=0}^{h} \mathbb{E}\big[\big\|\frac{\Theta_{-k+[0,h]}}{\|\Theta_{-t+[0,h]}\|_\alpha}\big\|_p^\alpha \mathbb{1}\big(\frac{\Theta_{-k+[0,h]}}{\|\Theta_{-k+[0,h]}\|_p} \in \cdot\big)\big], \quad x \to +\infty. \tag{B.5.15}$$

Then, a stationary time series $(\mathbf{X}_t)$ *satisfies Definition B.5.1 if and only if, for all* $h \geq 0$, $\mathbf{X}_{[0,h]}$ *is multivariate regularly varying.*

Lemma B.5.2 is proven in Lemma 7.1 in [26]. From Lemma B.5.2 we deduce

$$c(p,h) = \sum_{k=0}^{h} \mathbb{E}[\|\Theta_{-k+[0,h]}/\|\Theta_{-t+[0,h]}\|_\alpha\|_p^\alpha]. \tag{B.5.16}$$

Notice, if $p = \alpha$, then $c(\alpha, h) = (h+1)$, regardless of the extremal temporal dependence. Indeed, the representation in (B.5.15) highlights the important role of the ℓ^α-norm.

In view of the time-change formula, the forward spectral tail process $(\Theta_{[0,+\infty]})$ completely describes the spectral tail process; see [10]. In particular, $c(p,h) = \sum_{k=0}^{h} \mathbb{E}[\|\Theta_{[0,k]}\|_p^\alpha - \|\Theta_{[1,k]}\|_p^\alpha]$, and further computations allow us to rewrite (B.5.15) in terms of the forward process. We keep in mind the correspondance with joint multivariate regular variation, to keep track of the ℓ^p spectral components: "$\Theta/\|\Theta\|_p$" determining the distribution of $\mathbf{Q}^{(p)}(h)$ by the change of norms relation. In Section B.7 we will be interested in taking $h \to +\infty$ in (B.5.15), for studying the extremal behavior of blocks $\mathbf{X}_{[0,n]}$ as $n \to +\infty$.

Literature review

From a theoretical point of view, Definition B.5.1 admits further equivalent characterizations. For example, Theorem 4.1. in [152] and Theorem 2.9. [48] both use a measure theory approach. Moreover, the time-change formula (B.5.13) has proven to be a useful tool to characterize the limit spectral tail measure. [96] shows a correspondence with measures verifying this property and the max-stable processes. Motivated by Palm theory, the spectral tail measure defined by $(Y\Theta_t)$, for a (α)-Pareto random variable independent of (Θ_t), was also characterized in [139], highlighting the importance of the time-change type properties.

B.6 Central limit theory and clusters (of exceedances) in ℓ^∞

To study the overall clustering behavior, we want to let $h \to +\infty$ in (B.5.14) simultaneously as $x \to +\infty$. This can be done under **AC**. Condition **AC** allow us to define short periods with extramal behavior, that we call clusters, with finite expected length. Under **AC**, we will show in Chapter 2 the relation below holds

$$\mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_\infty > x_n) \xrightarrow{w} \mathbb{P}(Y\mathbf{Q}^{(\infty)} \in \cdot), \quad n \to +\infty. \tag{B.6.17}$$

where Y is independent of $\mathbf{Q}^{(\infty)}$, $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, $\|\mathbf{Q}^{(\infty)}\|_\infty = 1$ a.s., and convergence holds in $(\tilde{\ell}^\infty, \tilde{d}_\infty)$. Equation (B.6.17) gives a formal framework to define clusters. We call $\mathbf{Q}^{(\infty)}$ the spectral cluster (of exceedances). The cluster limit at the right-hand side of (B.6.17) was implicitly introduced in [40].

Using (B.6.17), and borrowing the ideas in [40], we can model extremal observations as independent clusters (of exceedances), under adequate mixing assumptions. Two things are crucial about this approach: the distribution of the cluster (of exceedances) itself obtained from the limit in (B.6.17), and the probability of hitting a cluster, i.e., $\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_\infty > x_n)$. This modeling approach was used in [40] to extend the central limit theorem of regularly varying observations to a time-dependent setting. We recall the main result in [40], that we formulate with the notation from Theorem 3.6. in [9].

Theorem B.6.1. *Consider $(\mathbf{X}_t)$ to be a regularly varying time series. Let (x_n) be a sequence satisfying $\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_\infty > x_n) \to 0$, and* **AC**, *$\mathcal{A}(x_{b_n})$; see Section B.4. Then, the following point processes convergence holds*

$$N_n = \textstyle\sum_{t=1}^n \epsilon_{x_{b_n}^{-1} X_t} \xrightarrow{d} N = \sum_{i=1}^\infty \sum_{j\in\mathbb{Z}} \epsilon_{\Gamma_i Q_{tj}^{(\infty)}}, \quad n \to +\infty, \tag{B.6.18}$$

where $(\mathbf{Q}_i^{(\infty)})$, $i = 1, 2, \dots$, is an i.i.d. sequence distributed as the spectral clusters (of exceedances); see (B.6.17), and $\sum_{t=1}^\infty \epsilon_{\Gamma_t}$ is a Poisson point process on $(0, \infty)$, with intensity measure $\theta_{|\mathbf{X}|} d(-y^{-\alpha})$, independent of the i.i.d. sequence $(\mathbf{Q}_i^{(\infty)})$, $i = 1, 2, \dots$. The constant $\theta_{|\mathbf{X}|}$ is the extremal index of $(|\mathbf{X}_t|)$ and satisfies

$$\mathbb{P}(\|\mathbf{X}_{[0,b_n]}\|_\infty > x_{b_n}) \sim \theta_{|\mathbf{X}|}\, b_n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty. \tag{B.6.19}$$

Finally, arguing as in [40], we can use a continuous mapping argument to deduce from equation (B.6.18) a central limit theorem for the mean corrected partial sums of regularly varying time series. We recall this result in the next theorem, proven in Theorem 3.1. in [40].

Theorem B.6.2. *Consider $(\mathbf{X}_t)$ to be a regularly varying time series. Assume the assumptions of Theorem B.6.1 hold. Denoting $(\mathbf{Q}_i^{(\infty)})$, for $i = 1, 2, \dots$ an i.i.d. sequence distributed as the spectral cluster (of exceedances), then*

- *If $\alpha \in (0, 1)$, $\sum_{t=1}^n X_t / x_{b_n} \xrightarrow{d} \xi$, as $n \to +\infty$, and $\xi = \sum_{i=1}^\infty \sum_{j\in\mathbb{Z}} \Gamma_i Q_{ij}^{(\infty)}$ is stable distributed.*

- *If $\alpha \in [1,2)$, and $\mathbf{CS}_1$ holds too, $\sum_{t=1}^{n} X_t/x_{b_n} - \mathbb{E}[\overline{X_t/x_{b_n}}^1] \xrightarrow{d} \xi$, as $n \to +\infty$, and ξ is stable distributed.*

Actually, [40] uses a vanishing-small-values condition stronger than $\mathbf{CS}_1$ instead to obtain the result. We proof in Chapter 4 that under **AC**, $\mathbf{CS}_1$, and $\mathcal{A}(x_{b_n})$, the result of Theorem B.6.2 holds true.

B.6.1 Literature Review

Cluster (of exceedances) are further reviewed in [10, 9, 41]. Central limit theory for time-dependent random variables was originally investigated in [40], from a point process point of view, but also in [97, 98] using telescopic sums arguments. Renormalized partial sums of regularly varying time series have been re-considered more recently in [91, 5, 123]. Functional Donsker-type limit theorem for re-scaled random walks are shown in [8, 9], using point process tools and a refinement of Theorem B.6.1.

B.7 Large deviation principles and cluster processes in ℓ^p

We consider $(\mathbf{X}_t)$ to be a regularly varying time series with (tail)-index $\alpha > 0$. We review Lemma 3.6. in [25] together with and Proposition 3.1. in [26] to state the Proposition below. The result appeals to the theory in [96].

Proposition B.7.1. *Consider a stationary regularly varying time series $(\mathbf{X}_t)$ with (tail)-index $\alpha > 0$, and spectral tail process (Θ_t). Then, the following statements are equivalent:*

i) $\|\Theta\|_\alpha < +\infty$ *a.s.*

ii) $\lim_{t\to+\infty} |\Theta_t| = 0$, *a.s.,*

*Moreover, $i), ii)$ hold under **AC** and in this case*

$$\mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_\infty > x_n) \xrightarrow{w} c(\infty)^{-1}\mathbb{E}\big[\tfrac{\|\Theta\|_\infty^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}(Y\Theta/\|\Theta\|_\infty \in \cdot)\big], \quad n \to +\infty \tag{B.7.20}$$

where Y is independent of (Θ_t), $\mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, and convergence holds in $(\tilde{\ell}^\infty, \tilde{d}_\infty)$. Also, $c(\infty) = \mathbb{E}[\|\Theta\|_\infty^\alpha/\|\Theta\|_\alpha^\alpha]$.

Remark B.7.2. *Moreover, under the assumptions of Thoerem B.6.1, the extremal index $\theta_{|\mathbf{X}|}$ of $(|\mathbf{X}_t|)$ is well defined; see (A.2.7), and $c(\infty) = \mathbb{E}[\|\Theta\|_\infty^\alpha/\|\Theta\|_\alpha^\alpha] = \theta_{|\mathbf{X}|}$.*

Our new approach consists in deriving new large deviation principles for blocks of consecutive observations $\mathbf{X}_{[0,n]}$, embedded in $(\tilde{\ell}^p, \tilde{d}_p)$. In this setting, we think the tail of $\mathbf{X}_{[0,n]}$ as the blocks distribution when its ℓ^p–norm its large. Then we consider a sequence (x_n) such that

$$\|\mathbf{X}_{[0,n]}\|_p \;=\; (\textstyle\sum_{t=1}^{n} |\mathbf{X}_t/x_n|^p)^{1/p} \;\xrightarrow{\mathbb{P}}\; 0, \quad n \to +\infty$$

and model $\mathbf{X}_{[0,n]}/x_n$ on the event $\{\|\mathbf{X}_{[0,n]}\|_p > x_n\}$. To study ℓ^p–norms, notice $|\mathbf{X}_1|^p$ is also regularly varying of index α/p, therefore the large deviations of ℓ^p–norms can be studied through large deviations of sums with heavy-tailed innovations. We built on the approach of large deviations of sums from [123, 121], extending Theorem B.2.4 to a time-dependent setting. We will work under assumptions similar to those considered in [121], but tailored for the sums of p-powers setting.

To sum up, we show in Chapter 2 that assuming **AC**, a small-vanishing-conditions similar to $\mathbf{CS}_1$ tailored for p–powers sums, and

$$c(p) \;:=\; \mathbb{E}[\|\Theta\|_p^\alpha/\|\Theta\|_\alpha^\alpha] < +\infty, \tag{B.7.21}$$

then,

$$\mathbb{P}(x_n^{-1}\mathbf{X}_{[0,n]} \in \cdot \mid \|\mathbf{X}_{[0,n]}\|_p > x_n) \xrightarrow{w} c(p)^{-1}\mathbb{E}\big[\tfrac{\|\Theta\|_p^\alpha}{\|\Theta\|_\alpha^\alpha}\mathbb{1}(Y\Theta/\|\Theta\|_p \in \cdot)\big], \tag{B.7.22}$$

$$\xrightarrow{w} \mathbb{P}(Y\mathbf{Q}^{(p)} \in \cdot), \quad n \to +\infty, \tag{B.7.23}$$

where Y is independent of $\mathbf{Q}^{(p)}, \mathbb{P}(Y > y) = y^{-\alpha}$, for $y > 1$, $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s., and convergence takes place in $(\tilde{\ell}^p, \tilde{d}_p)$. Moreover,

$$\mathbb{P}(\|\mathbf{X}_{[0,n]}\|_p > x_n) \sim c(p)\, n\, \mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty, \tag{B.7.24}$$

The relation (B.7.22) allows us to define cluster processes $\mathbf{Q}^{(p)} \in \ell^p$, satisfying $\|\mathbf{Q}^{(p)}\|_p = 1$ a.s. This will be the main large deviations result we derive in Chapter 2.

Moreover, $p \mapsto c(p)$ is a non-decreasing function such that $c(\alpha) = 1$ and $c(\infty)$ is the candidate extremal index; see Remark B.7.2. These constants can also be obtained letting $h \to +\infty$ in $c(p, h)$ defined in (B.5.16). We study the advantages of this approach for cluster inference in Chapters 2, 3, and for high quantiles inference in Chapter 4.

Finally, we argue that under **AC**, and motivated by the equivalence between $i), ii)$, a good candidate to capture the spectral cluster shape is $\Theta/\|\Theta\|_\alpha$ which we can recover letting $p = \alpha$ in (B.7.22).

B.7.1 Literature review

Pursuing the approach in Theorem B.2.4, we can study the tail of mean corrected partial sums from (X_t), a regularly varying time series satisfying the tail balance condition (B.2.3). We study regular variation, but the fundamental property is subexponenial distributions as defined in (B.2.4). This broader setting was already treated in [124, 125, 33] for i.i.d. observations, and it is considered in [118] for stationary subexponential observation. The case of regularly varying time series has received more attention, and we restrict to it in our work. Mikosch and Wintenberger [123, 121] give general and sufficient conditions on the dependence structure of (X_t) to state a similar result as in Theorem B.2.4. We refer to Theorem 3.1. in [121] for a generalization of Nagaev's theorem in the stationary regularly varying setting. They study suitable regions $\Lambda_n \subseteq (x_n, +\infty)$ such that if

(X_t) is univariate regularly varying, then for some $c \in [0, +\infty)$

$$\lim_{n\to+\infty} \sup_{x\in\Lambda_n} \Big| \frac{\mathbb{P}(\sum_{t=1}^n X_t > x)}{n\mathbb{P}(|X_t| > x)} - c \Big| \;=\; 0. \tag{B.7.25}$$

The extremal regions Λ_n here considered are analogous to the ones from Theorem B.2.4 for many classical examples. For Markov chains, they prove a version of Theorem B.2.4 over regions $x_n \le x \le c_n$ instead of $x \ge x_n$, since verification can be is easier to handle using their drift-condition $\mathbf{DC}_p$. Their result is general and builds on telescopic sum ideas already appearing in [97, 98]. Concrete examples have also been studied independently. For example, the case of linear processes was studied in [120], and stationary solutions to the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$, for an i.i.d. sequence (A_t, B_t) were studied in [90, 91, 106], for $\mathbb{E}[A^\alpha] < 1$ and B regularly varying of tail index $\alpha > 0$. For large deviation's under the Kesten's conditions we refer to [24].

B.8 Cluster inference

Inferring the behavior of the cluster processes through summary statistics is what we call cluster inference. Consider $(\mathbf{X}_t)$ to be a regularly varying time series taking values in $(\mathbb{R}^d, |\cdot|)$. Cluster inference has been typically addressed using the clusters (of exceedances) framework. In this setting, we start by dividing the sample into $m_n = \lfloor n/b_n \rfloor$ disjoint blocks as

$$\mathbf{X}_{[1,b_n]}, \mathbf{X}_{[b_n+1,2b_n]}, \ldots, \mathbf{X}_{[n-b_n+1,n]}, \tag{B.8.26}$$

then a large threshold is fixed and only blocks with at least one exceedance are kept leading to an empirical sample of clusters (of exceedances). The goal is then to compute the average clustering effect captured by these extremal block. This approach is formalized in [52] who assess the estimation performance.

For example, interpretation of the extremal index through cluster features is the basis for inference procedures from the early 1990s [160]. We can estimate the extremal index in this way exploiting the relation (B.6.19). This idea motivated the so-called blocks estimators [88, 161], based on counts of exceedances per cluster, and also the runs estimator [161, 65], based on interexceedances lengths. Overall, the main goal of cluster inference is to capture the properties of extremal clusters. The extremal index has been typically understood as a summary of the clustering of exceedances effect.

We propose to use p-cluster processes with $p < \infty$ to infer the extremal behavior of the sample. We propose to keep the disjoint blocks in (B.8.26) whose ℓ^p–norm exceeds a large threshold instead. Recall, cluster processes are related through Equation (B.7.22) by a change of norms functional. In practice this means that in the case of cluster statistics, the same quantity can be estimated in different ways letting functionals g_p act on extremal ℓ^p-blocks for different p. This observation opens the road for improving cluster inference. We pursue this road on Chapter (2). For example, we mentioned the extremal index can be recovered from (B.6.19), but note it can also be recovered using extremal ℓ^α-blocks since $\theta_{|\mathbf{X}|} = c(\infty) = \mathbb{E}[\|\mathbf{Q}^{(\alpha)}\|_\infty^\alpha]$ from equation B.7.21; see Remark B.7.2. Investigation of cluster-based inference of the extremal index is discussed in detail in Chapter 3.

We can highlight two main advantages of addressing cluster inference via extremal ℓ^p-blocks with $p < +\infty$. First, as p decreases, the p–norm increases, and consequently, the probability of hitting a cluster increases when $c(p) < c(\infty)$ as can be seen from Equation (B.7.24). Second, in our framework, if $p = \alpha$,

$$\mathbb{P}(\|\mathbf{X}_{[0,b_n]}\|_\alpha > x_{b_n}) \sim b_n\,\mathbb{P}(|\mathbf{X}_0| > x_{b_n}), \quad n \to +\infty. \tag{B.8.27}$$

where the relation in (B.8.27) holds regardless of the time dependence dynamics. This invariance principle is advantageous since cluster inference no longer depends on $c(p)$.

B.8.1 literature review

From a theoretical point of view, it is already challenging to define clusters (of exceedances); see [65, 153, 154, 10, 9], and references there in. Concerning inference, the first theoretical setting for cluster-based statistics was established in [52]. The authors presented a global framework for showing asymptotic normality of cluster-based disjoint blocks estimators, and reviewed the example of the extremal index. Asymptotic normality of the sliding blocks estimators was then studied in [51]. Further, [28] showed, for cluster inference, the asymptotic variance of both disjoint and sliding blocks estimators coincide. We refer to [108] for a modern treatment of cluster inference. Already in [39, 50], the authors discuss how to improve cluster inference using the index of regular variation of the time series. However, all the literature previously mentioned deals uniquely with clusters (of exceedances).

B.9 Time series models

We investigate two examples of stationary regularly varying time series.

B.9.1 Linear model

Let $(\mathbf{Z}_t)$ be an i.i.d. $\mathbb{R}^d$-valued sequence, and independent of (Φ_j), which is a time series of matrices in $\mathbb{R}^{d\times d}$. We consider the linear model $(\mathbf{X}_t)$, defined as the solution to the equation

$$\mathbf{X}_t \quad = \quad \sum_{t\in\mathbb{Z}} \Phi_j \mathbf{Z}_{t-j}, \quad t \in \mathbb{Z}. \tag{B.9.28}$$

and the following holds

i) $(\mathbf{Z}_t)$ is an i.i.d. sequence, regularly varying of index $\alpha > 0$, centered if $\alpha > 1$, and independent of (Φ_j).

ii) $\mathbb{E}[\big(\|\Phi_j\|_{(\alpha-\epsilon)\wedge 2}\big)^{\alpha+\epsilon}] < +\infty$, for some $\epsilon > 0$, *iii)* $\mathbb{E}[\|\Phi_j\Theta_0^Z\|_\alpha^\alpha] > 0$, where Θ_0^Z is independent of (Φ_j) and is distributed as the spectral measure of Z_1 and $|\Theta_0^{\mathbf{Z}}| = 1$ a.s..

Conditions $i), ii), iii)$, guarantee, respectively, the heavy-tailedness,the existence of the stationary linear model in (B.9.28), and of a non-null exponent measure by an application of the multivariate Breiman's lemma; see [6]. More precisely, recalling Theorem 15.1.2 in [108], we can state the following result

Proposition B.9.1. *Let $(\mathbf{X}_t)$ be the stationary solution in (B.9.28) satisfying $i), ii), iii)$. Then, it is regularly varying of index $\alpha > 0$ with spectral tail (Θ_t) given by*

$$\Theta_t \;=\; \frac{\Phi_{t+J}}{|\Phi_j\Theta_0^{\mathbf{Z}}|}\,\Theta_0^{\mathbf{Z}}, \quad t \in \mathbb{Z},$$

where (Φ_j) is independent of Θ_0^Z, and J is a random-shift variable taking values in $\mathbb{Z}$, depending on $(\Theta_0^{\mathbf{Z}}, (\Phi_j))$, such that $\mathbb{P}\big(J = j \,|\, (\Theta_0^{\mathbf{Z}}, (\Phi_j))\big) = |\Phi_j\Theta_0^{\mathbf{Z}}|^\alpha / \mathbb{E}[\|\Phi_j\Theta_0^Z\|_\alpha^\alpha]$.

Classical examples defined by (B.9.28) are the m_0-dependent processes, where $|\Phi_j| = 0$ for all except finitely many indexes in (B.9.28). We can also study *autoregressive moving-average* ARMA processes as solutions to (B.9.28) (cf. [18]).

Example B.9.2. *Let (X_t) be the $AR(1)$ model, such that $X_t = \varphi X_{t-1} + Z_t$, $t \in \mathbb{Z}$,$|\varphi| < 1$ and (Z_t) is an i.i.d. regularly varying sequence with (tail)-index $\alpha > 0$, then*

$$\Theta_t \;=\; \begin{cases} \varphi^t\,\Theta_0^Z, & t \geq -J, \\ 0, & t < -J. \end{cases} \tag{B.9.29}$$

and $\mathbb{P}(J = j \,|\, \Theta_0^Z) = |\varphi^j|^\alpha(1 - |\varphi|^\alpha)$, for $j \in \mathbb{Z}$. Moreover, it has extremal index $\theta_{|\mathbf{X}|} = (1 - |\varphi|^\alpha)$.

B.9.2 Affine stochastic recurrence equation under Kesten's assumptions

Let $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$ be an i.i.d. sequence of non-negative random $d \times d$ matrices with generic element $\mathbf{A}$, and random vectors with generic element $\mathbf{B}$, taking values in $\mathbb{R}^d$. We consider $(\mathbf{X}_t)$ to be the causal solution to the equation:

$$\mathbf{X}_t \;=\; \mathbf{A}_t\mathbf{X}_{t-1} + \mathbf{B}_t, \qquad t \in \mathbb{Z}, \tag{B.9.30}$$

Kesten's celebrated paper [103], shows sufficient conditions for the existence of a stationary solution to (B.9.30), admitting the causal representation

$$\mathbf{X}_t \;=\; \sum_{i \geq 0} \mathbf{A}_{t-i+1} \dots \mathbf{A}_t\,\mathbf{B}_{t-i}, \qquad t \in \mathbb{Z}, \tag{B.9.31}$$

where the summand for $i = 0$ is $\mathbf{B}_t$, for $i = 1$ is $\mathbf{A}_t\mathbf{B}_{t-1}$, and so on; see [6, 24] and references therein. We quote below theorem 2.1. in [6] to explicit these conditions.

Theorem B.9.3. *(Existence of stationary solution) Assume* $\mathbb{E}\big[\log^+ |\mathbf{A}|_{op}\big] < +\infty$*,* $\mathbb{E}\big[\log^+ |\mathbf{B}|\big] < +\infty$*, and assume the Lyapunov exponent for the sequence* $(\mathbf{A}_t)$ *is negative. Then, the solution in* (B.9.31) *converges a.s. and is the unique strictly stationary causal solution of* (B.9.30)*.*

Concerning heavy-tailedness, under the so-called Kesten's conditions, $(\mathbf{X}_t)$ can admit a stationary regularly varying solution even when innovations $((\mathbf{A}_t, \mathbf{B}_t))$, are light-tailed as first noticed in [103]. Extremes are due to the products of arbitrary length in (B.9.31); see [13] for a review, [74] for the Goldie and Kesten's conditions for univariate innovations, and [6, 7] for a multivariate treatment of Equation (B.9.30). Under these Kesten's conditions, the index of regular variation of $(\mathbf{X}_t)$, denoted $\alpha > 0$, is the unique solution to

$$\lim_{n\to+\infty} n^{-1} \log \mathbb{E}\big[|\mathbf{A}_n \cdots \mathbf{A}_1|_{op}^{\alpha}\big] \;=\; 0. \tag{B.9.32}$$

Equation (B.9.30) is key in econometric to model squared ARCH(p) and the volatility of GARCH(p, q) processes. We review Theorem 2.4. with Corollary 2.7. in [6] to state exact conditions for heavy-tailed solutions $(\mathbf{X}_t)$ from light-tailed innovations.

Theorem B.9.4. *(Heavy-tailedness) Consider* $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$ *to be an i.i.d. sequence of matrices* $\mathbf{A}$ *in* $\mathbb{R}^{d\times d}$ *with non-negative entries, and random vectors* $\mathbf{B}$ *in* $\mathbb{R}^d$*-valued with non-negative entries such that* $\mathbf{B} \neq 0$ *a.s. Assume also the conditions below hold*

1. *For some* $\kappa > 0$*,* $\mathbb{E}[|\mathbf{A}|_{op}^{\kappa}] < 1$*.*

2. $\mathbf{A}$ *has no zero rows a.s.*

3. *The set* Γ *in* (B.9.33) *generates a dense group on* $\mathbb{R}$*.*

$$\Gamma \;=\; \{\ln |\mathbf{a}_n \cdots \mathbf{a}_1|_{op} : n \geq 1, \mathbf{a}_n \cdots \mathbf{a}_1 > 0,\ \mathbf{a}_n, \ldots, \mathbf{a}_1 \text{ are in the support of } \mathbf{A}\text{'s law }\}. \tag{B.9.33}$$

4. *There exists* $\kappa_1 > 0$*,* $\mathbb{E}\big[(\min_{i=1,\ldots,d} \sum_{t=1}^{d} A_{ij})^{\kappa_1}\big] \geq d^{\kappa_1/2}$ *and* $\mathbb{E}[|\mathbf{A}|_{op}^{k_1} \ln^+ |\mathbf{A}|_{op}] < +\infty$*.*

Then, there exists a unique solution $\alpha > 0$ *to* (B.9.32)*, there exists a unique strictly stationary causal solution* $(\mathbf{X}_t)$ *to* (B.9.32)*. Moreover, if* $\mathbb{E}[|\mathbf{B}|^{\alpha}] < +\infty$ *and either* $d = 1$ *or* $d > 1$ *and* $\alpha > 0$ *is not an even integer, then the finite-dimensional distributions of* $(\mathbf{X}_t)$ *are regularly varying of index* $\alpha > 0$*.*

Moreover, under the assumptions of Theorem B.9.4, $\mathbf{X}_0$ admits a spectral measure Θ_0 and its forward spectral tail measure is given by

$$\Theta_t \;=\; \mathbf{A}_t \cdots \mathbf{A}_1 \Theta_0, \quad t \in \mathbb{N}_{\geq 1}.$$

where $(\mathbf{A}_t)$ is a sequence of i.i.d. random matrices with generic element $\mathbf{A}$; see [6, 10].

Example B.9.5. *Let* $\log A_t \stackrel{d}{=} N_t - 0.5$ *where* N_t *denotes a centered and reduced Gaussian random variable. Then, the causal stationary solution* (X_t)*, satisfying* $X_t = AX_{t-1}, t \in \mathbb{Z}$*, is regularly varying with index* $\alpha = 1$ *and* $\mathbb{E}[(A)^{1-\delta}] < 1$ *for all* $\delta > 0$*.*

List of Figures

List of Tables

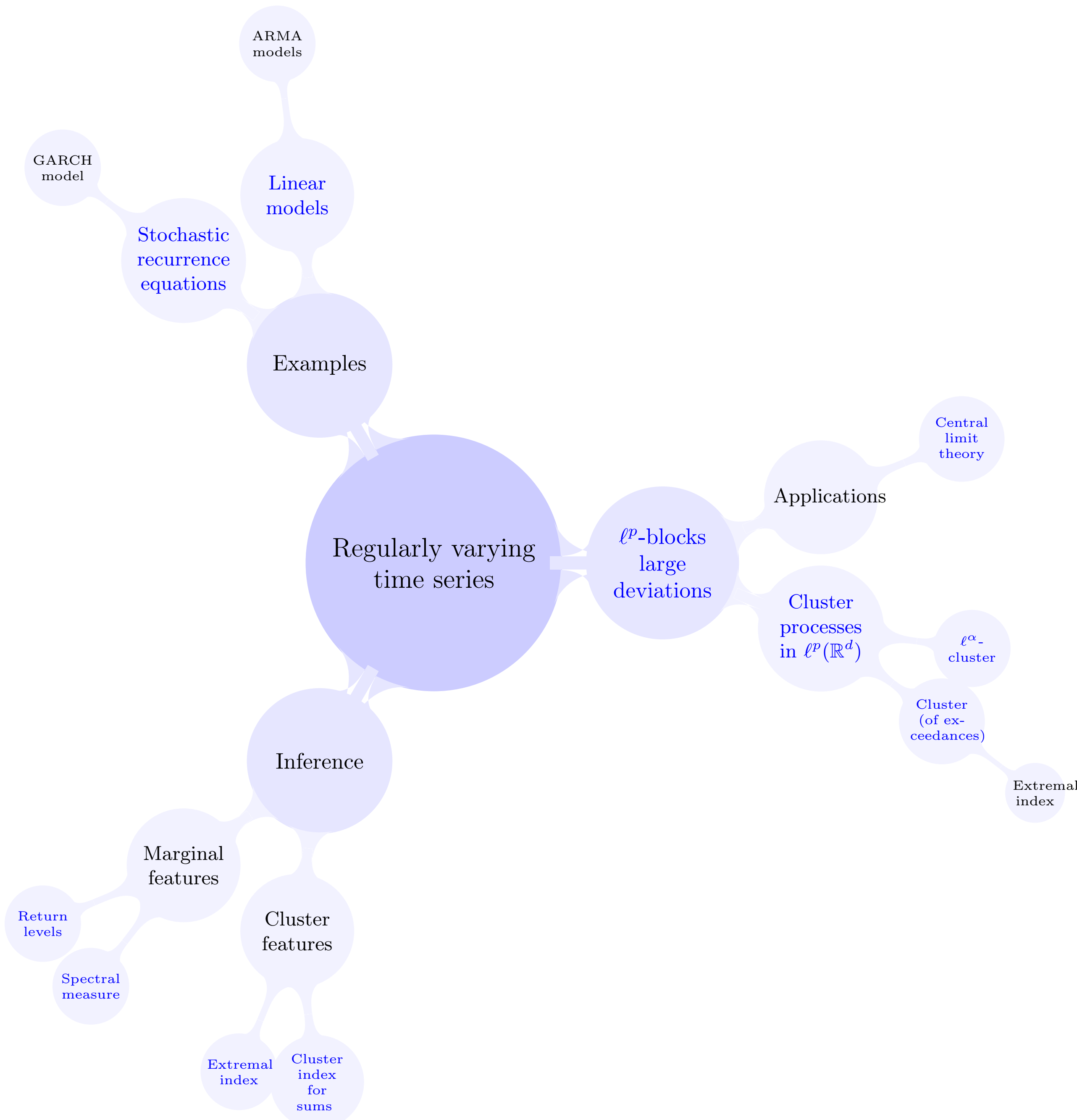


Figure 5.1: Mindmap of topics and links

Abstract

Assessing the time dependence of multivariate extremes for heavy rainfall modeling

Nowadays, it is common in environmental sciences to use extreme value theory to assess the risk of natural hazards. In hydrology, rainfall amounts reach high-intensity levels frequently, which suggests modeling heavy rainfall from a heavy-tailed distribution. In this setting, risk management is crucial for preventing the outrageous economic and societal consequences of flooding or landsliding. Furthermore, climate dynamics can produce extreme weather lasting numerous days over the same region. However, even in the stationary setting, practitioners often disregard the temporal memories of multivariate extremes. This thesis is motivated by the case study of fall heavy rainfall amounts from a station's network in France. Its main goal is twofold. First, it proposes a theoretical framework for modeling time dependencies of multivariate stationary regularly varying time series. Second, it presents new statistical methodologies to thoughtfully aggregate extreme recordings in space and time.

To achieve this plan, we consider consecutive observations, or blocks, and analyze their extreme behavior as their ℓ^p-norm reaches high levels, for $p > 0$. This consideration leads to the theory of p-clusters, which model extremal ℓ^p-blocks. In the case $p = \infty$, we recover the classical *cluster (of exceedances)*. For $p < \infty$, we built on large deviations principles for heavy-tailed observations. Then, we study in depth two setups where p-cluster theory appears valuable. First, we design disjoint blocks estimators to infer statistics of p-clusters, e.g., the *extremal index*. Actually, p-clusters are linked through a change of norms functional. This relationship opens the road for improving cluster inference since we can now estimate the same quantity with different choices of p. We show cluster inference based on $p < \infty$ is advantageous compared to the classical $p = \infty$ strategy in terms of bias. Second, we propose the stable sums method for high return levels inference. This method enhances marginal inference by aggregating extremes in space and time using the ℓ^α-norm, where $\alpha > 0$ is the (tail) index of the series. In simulation, it appears to be robust for dealing with temporal memories and it is justified by the α-cluster theory.

Keywords[1]: *Extreme value theory, regularly varying time series, large deviations, environmental time series*

[1] *Primary 60G70 Secondary 60F10 62G32 60F05 60G57 60G55 60F99 60J10 62M10*