Statistical inference for heavy tailed time series and vectors

Zhigang Tong

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics[1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

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

Abstract

In this thesis we deal with statistical inference related to extreme value phenomena. Specifically, if $\boldsymbol{X}$ is a random vector with values in d-dimensional space, our goal is to estimate moments of $\psi(\boldsymbol{X})$ for a suitably chosen function ψ when the magnitude of $\boldsymbol{X}$ is big. We employ the powerful tool of regular variation for random variables, random vectors and time series to formally define the limiting quantities of interests and construct the estimators. We focus on three statistical estimation problems: (i) multivariate tail estimation for regularly varying random vectors, (ii) extremogram estimation for regularly varying time series, (iii) estimation of the expected shortfall given an extreme component under a conditional extreme value model. We establish asymptotic normality of estimators for each of the estimation problems. The theoretical findings are supported by simulation studies and the estimation procedures are applied to some financial data.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Rafal Kulik for his important guidance, time and valuable remarks. Thanks for giving me such an invaluable opportunity to work on extreme value theory using the tools of regular variation with you. Your insight and clear explanations have always helped me get through the difficult questions at hand.

Thanks are also due to Professors Raluca Balan, Debbie Dupuis, Mohamedou Ould Haye and Mahmoud Zarepour for spending time going through my thesis and raising many interesting questions. Thanks for all of you for pointing out the typos and mistakes in the thesis.

Last but not least, I would like to give my deepest gratitude to my wife Katie Yan for her consistent encouragement and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Extreme value theory (EVT) provides an asymptotically justified framework for the statistical modeling of rare events. For independent, identically distributed (i.i.d.) univariate random variables, the theory dates back to the 1920s and the celebrated Fisher-Tippett-Gnedenko theorem that describes possible limits of normalized maxima. The limiting distributions are called Weibull, Gumbel or Fréchet. The Fréchet distribution has regularly varying (i.e., power-law tail with a positive index α). This distribution and its maximum domain of attraction are suitable models for describing *heavy-tail phenomena*. Regular variation also provides a link between EVT and powerful probabilistic tools such as point processes and their weak convergence. The theoretical foundations triggered the development of suitable statistical methods, including the Peaks Over Threshold (POT) approach and Generalized Pareto Distribution (GPD) fitting. In the Fréchet domain case, different estimators of the tail index, including the celebrated Hill estimator, were studied. Although some statistical challenges remain, one can argue that EVT for i.i.d. random variables is well understood and summarized in research monographs like [18] or [12].

However, the situation is more complicated when dealing with dependent (stationary or non-stationary) or complex (such as multivariate or function-valued) data. In this thesis we deal with statistical inference related to extreme value phenomena. Specifically, if $\boldsymbol{X}$ is a random vector with values in $\mathbb{R}^d$, our goal is to estimate moments of $\psi(\boldsymbol{X})$ for a suitably chosen function ψ when the magnitude of $\boldsymbol{X}$ is big. Since $\boldsymbol{X}$ has values in $\mathbb{R}^d$, we have to clarify what does it mean that the vector is *big*. This will be achieved by a notion of regular variation of random variables, random vectors and time series. This powerful tool will allow us to formally define limiting quantities of interests and guide us through the estimation procedure.

The structure of the thesis is as follows:

- In Chapter 2, we provide some preliminary technical tools that are used throughout the thesis. We first introduce the notion of regular variation. We cite some important results such as the Uniform Convergence Theorem, Potter's bounds, Karamata's Theorem and Breiman's Lemma. We then extend the notion of the univariate regular variation to a multivariate case. We discuss the distinction between extremal independence and extremal dependence. Then we introduce the concept of regular variation for time series. We also consider the Conditional Extreme Value model (CEV). The CEV is introduced to deal with situations when the regular variation fails to capture extremal behaviour, like in the aforementioned extremally independent case.

  Most of the material presented in Chapter 2 is based on existing literature (e.g. [28]), which is clearly mentioned in the appropriate places. However, some of the results presented in Section 2.2 (a quasi-spectral decomposition) and of Section 2.4 (properties of the Conditional Extreme Value model) are new and can be found in the authors' papers, [26] and [27], respectively.

- In Chapter 3, we summarize several results on weak convergence of stochastic processes. We consider weak convergence and discuss weak convergence using entropy methods. The results here are not new and we based our presentation on, respectively, [5] and [31].

- In Chapter 4, we discuss the weak convergence for tail empirical processes, using the theory described in Chapter 3. The tail empirical processes is an extremal counterpart of the classical empirical processes. We prove the weak convergence for tail empirical processes based on i.i.d. random vectors. The results, as stated, are new, but they follow a relatively classical path. We also state a functional central limit theorem for tail empirical processes based on regularly varying time series which is adapted from [25]. Finally, we prove a functional central limit theorem for tail empirical processes based on i.i.d. extremally independent random vectors, which is new and based on the author's paper, [27].

- Chapter 5 deals with the statistical inference and is the main part of the thesis. These results are new and they can be found in the author's papers, [26], [27] and [30]. First, in Section 5.2, we discuss the estimation for i.i.d. vectors under extremal dependence. We propose an alternative nonparametric approach for estimating the conditional tail distribution that is based on the *quasi-spectral* decomposition. We argue that the estimation procedure based on the quasi-spectral representation may lead to an improvement in terms of efficiency or in terms of the conditions required to achieve asymptotic normality, as compared to other nonparametric methods. Then, in Section 5.3, we extend the ideas of the preceding sections to the problem of estimation for regularly varying time series with extremal dependence. We prove the asymptotic normality of extremogram estimators using deterministic and random levels. We present the examples of AR(1) and a stochastic recurrence equation and conduct simulation studies and data analysis. Finally, in Section 5.4, we proceed with estimation under extremal independence, specifically using the Conditional Extreme Value model developed in Chapter 2. We also estimate the Marginal Expected Shortfall. We construct an estimator and derive its asymptotic normality. We compare the applicability of our results with [8]. We argue that our approach is more applicable than in the latter reference.

Chapter 2

Preliminaries

In this chapter we discuss some preliminary technical tools that are needed in the thesis, in particular, regular variation and measures of extremal dependence.

In Section 2.1 we introduce the notion of regular variation. First, in Section 2.1.1 we define regular variation of deterministic functions and quote important results like the Uniform Convergence Theorem, Potter's bounds and Karamata's Theorem. The results in this section are not original and they can be found in standard literature like e.g. [7] and [28]. Then, in Section 2.1.2 we introduce the notion of vague convergence and its relationship to weak convergence. Again, the results are taken from the literature, e.g. [28]. In Section 2.1.3 we consider regularly varying random variables. We connect regular variation to vague convergence and the latter to weak convergence of conditional probabilities. Then, we establish different moment bounds and tail asymptotics, like Potter's bounds and Breiman's Lemma. The results are not original (see e.g. [28]), however, we re-prove some results to illustrate the techniques used.

In Section 2.2 we extend the notion of the univariate regular variation to multivariate regular variation. We include an important distinction between extremal independence and extremal dependence. The regular variation implies the quasi-spectral decomposition. This decomposition has not yet appeared in the literature in the present form, however, we note that in the time series context this decomposition is known as the spectral tail process; see [3] for details.

In Section 2.3 we introduce the concept of regular variation for time series. As in the multivariate case, we link regular variation to quasi-spectral decomposition. Several examples (AR(1), solutions to Stochastic Recurrence Equations and threshold ARCH) are given. The contents are again based on the existing literature.

In Section 2.4 we consider the Conditional Extreme Value model. This set-up was studied rigorously in [20], [10] and [23]. Connections of CEV to the marginal expected shortfall were mentioned in [27]. Some results in Section 2.4 are new and can be found in the latter reference.

Section 2.5 introduces the extremogram, the measure of extremal dependence for time series. We follow [11].

2.1 Univariate regular variation

In this section we introduce the notion of univariate regular variation of functions and random variables. Most of the material is well-known and is taken from standard references like [7] and [28]. Some of the results are re-proven to illustrate the techniques.

The structure of this section is as follows:

- Section 2.1.1 contains important results on regularly varying functions including the Uniform Convergence Theorem, Potter's bounds and Karamata's theorem.
- Section 2.1.2 deals with an important concept of vague convergence. Comparison to weak convergence is discussed.
- Section 2.1.3 contains results on regular variation of random variables. We link regular variation to vague convergence of measures, and the latter to weak

convergence of conditional distributions. Furthermore, important results like Potter's bounds and Breiman's Lemma are stated.

2.1.1 Regularly varying functions

Definition 2.1.1 *A function f defined on $[0, \infty)$ is said to be regularly varying at infinity if the limit*

$$\lim_{x \to +\infty} \frac{f(tx)}{f(x)} \tag{2.1.1}$$

exists for all $t > 0$.

With this definition, we can show that $f(x) = x^\gamma \log(x)$, $\gamma \in \mathbb{R}$, is a regularly varying function. Indeed

$$\lim_{x \to +\infty} \frac{f(tx)}{f(x)} = \lim_{x \to +\infty} \frac{(tx)^\gamma \log(tx)}{x^\gamma \log(x)} = t^\gamma \lim_{x \to +\infty} \frac{\log(t) + \log(x)}{\log(x)} = t^\gamma.$$

Similarly, one can also prove that functions such as $f(x) = x^\gamma \log(\log(x))$ and $f(x) = x^\gamma (\log(x))^\beta$, $\gamma, \beta \in \mathbb{R}$, are regularly varying.

Under the single assumption that the function f is measurable, the following Uniform Convergence Theorem characterizes the form of the possible limits and the nature of the convergence.

Theorem 2.1.2 *Let f be a positive measurable function defined on $[0, \infty)$ such that for all $t > 0$, the limit (2.1.1) exists. Then there exists $\gamma \in \mathbb{R}$ such that*

$$\lim_{x \to \infty} \frac{f(tx)}{f(x)} = t^\gamma , \tag{2.1.2}$$

for all $t \in (0, \infty)$ and the convergence is uniform on compact subsets of $(0, \infty)$. Moreover,

- *If $\gamma > 0$, the convergence is uniform on sets $[0, b]$, $0 \leq b < \infty$;*
- *if $\gamma < 0$, the convergence is uniform on sets $[b, \infty]$, $0 < b < \infty$;*

- *if $\gamma = 0$, the convergence is uniform on sets $[a, b]$, $0 < a \leq b < \infty$.*

The parameter γ is called **the index of regular variation**. If $\gamma = 0$, then the function f is called *slowly varying (at infinity)*. In fact, each regularly varying function f with index γ can be expressed as

$$f(x) = x^\gamma \ell(x) ,$$

where ℓ is slowly varying. The uniform convergence implies the following result.

Lemma 2.1.3 *Let a_n and b_n be sequences of nonnegative real numbers such that $a_n, b_n \to \infty$ and $\lim_{n\to\infty} a_n/b_n = c \in (0, \infty)$. Let f be a regularly varying function at infinity with index $\gamma \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \frac{f(a_n)}{f(b_n)} = c^\gamma .$$

Proof: Write $f(a_n) = f((a_n/b_n)b_n)$. The conclusion of the lemma is immediate by noting that for $t > 0$, $f(tb_n)/f(b_n) \to t^\gamma$ uniformly on $[a, b]$, $0 < a \leq b < \infty$; see Theorem 2.1.2. ∎

Remark 2.1.4 We can modify the result as follows. Let a_n and b_n be sequences of nonnegative real numbers such that $a_n, b_n \to 0$ and $\lim_{n\to\infty} a_n/b_n = c \in (0, \infty)$. Let f be a regularly varying function at zero with index $\gamma \in \mathbb{R}$. Then

$$\lim_{n\to\infty} \frac{f(a_n)}{f(b_n)} = c^\gamma .$$

We now state Potter's bounds.

Proposition 2.1.5 *Let ℓ be a positive function defined on $[0, \infty)$, slowly varying at infinity, locally bounded away from zero and ∞. Then for each $\epsilon > 0$, there exist positive constants c_ϵ, C_ϵ such that, for all $x \geq 1$ and $y \geq x$,*

$$c_\epsilon x^{-\epsilon} \leq \ell(x) \leq C_\epsilon x^\epsilon , \tag{2.1.3}$$

$$c_\epsilon \left(\frac{x}{y}\right)^\epsilon \le \frac{\ell(y)}{\ell(x)} \le C_\epsilon \left(\frac{y}{x}\right)^\epsilon . \tag{2.1.4}$$

We note that $\ell(x) = \log(x)$ does not fulfill the assumptions of the above proposition, since it explodes at zero. In such cases, Potter's bounds hold for x large enough. This suffices for most of the applications.

One of the most quoted properties of regularly varying function is "Karamata's theorem". It states that the integral of a regularly varying function is regularly varying, unless the index of regular variation is -1.

Theorem 2.1.6 *Let f be locally bounded on $[0,\infty)$, ultimately positive and regularly varying at infinity with index $\gamma \in \mathbb{R}$ and let $\beta \in \mathbb{R}$.*

(K1) If $\beta + \gamma < -1$, then $\int_1^\infty t^\beta f(t)\,\mathrm{d}t < \infty$ and

$$\int_x^\infty t^\beta f(t)\,\mathrm{d}t \sim -(\gamma+\beta+1)^{-1} x^\beta f(x) .$$

(K2) If $\beta + \gamma > -1$, then $\int_1^\infty t^\beta f(t)\,\mathrm{d}t = \infty$ and

$$\int_1^x t^\beta f(t)\,\mathrm{d}t \sim (\gamma+\beta+1)^{-1} x^\beta f(x) .$$

(K3) If $\gamma + \beta = -1$, then the function L defined by $L(x) = \int_1^x t^\beta |f(t)|\,\mathrm{d}t$ is slowly varying and $x^\beta f(x) = o(L(x))$.

Note: The proof is not original, the goal was to understand it and learn the technique.

Proof: The integrability properties in (K1) and (K2) follows from Potter's bounds. Since f is regularly varying with index γ, there exists a slowly varying function ℓ such that $f(x) = x^\gamma \ell(x)$. Consider the case $\gamma + \beta + 1 > 0$. Then,

$$\frac{1}{x^\beta f(x)} \int_1^x t^\beta f(t)\,\mathrm{d}t = \int_{1/x}^1 s^{\gamma+\beta} \frac{\ell(sx)}{\ell(x)}\,\mathrm{d}s .$$

By the uniform convergence theorem, if $\gamma+\beta > 0$, then the convergence of $s^{\gamma+\beta}\ell(sx)/\ell(x)$ to $s^{\gamma+\beta}$ is uniform on $[0,1]$ and thus the proof is concluded. If $\gamma + \beta \in (-1, 0]$, a little bit more care is required since the convergence is then uniform only on intervals $[a,1]$ with $a>0$. The trick is to "add and subtract" some $\epsilon>0$ small enough so that $\gamma+\beta+1-\epsilon>0$. Write then

$$\int_{1/x}^{1} s^{\gamma+\beta}\frac{\ell(sx)}{\ell(x)}\,\mathrm{d}s = \int_{1/x}^{1} s^{\gamma+\beta-\epsilon}\frac{s^{\epsilon}\ell(sx)}{\ell(x)}\,\mathrm{d}s \ .$$

Now the Uniform Convergence Theorem yields that $s^{\epsilon}\ell(sx)/\ell(x)$ converges to s^{ϵ} uniformly on $[0,1]$. Since $\gamma+\beta+1-\epsilon>0$, we can apply the bounded convergence theorem and we obtain

$$\lim_{x\to\infty}\int_{1/x}^{1} s^{\gamma+\beta}\frac{\ell(sx)}{\ell(x)}\,\mathrm{d}s = \int_{0}^{1} s^{\gamma+\beta-\epsilon}\lim_{x\to\infty}\frac{s^{\epsilon}\ell(sx)}{\ell(x)}\,\mathrm{d}s = \int_{0}^{1} s^{\gamma+\beta-\epsilon}s^{\epsilon}\,\mathrm{d}s = \frac{1}{\gamma+\beta+1} \ .$$

The case $\gamma+\beta\leq -1$ is similar (and simpler). Consider now the case $\gamma+\beta=-1$. We must prove that if ℓ is slowly varying at infinity, then $L(x)=\int_1^x s^{-1}\ell(s)\mathrm{d}s$ is also slowly varying and dominates $\ell(x)$. Assume without loss of generality that ℓ is positive. For $\epsilon>0$ and x large enough, we have, by the Uniform Convergence Theorem,

$$\frac{L(x)}{\ell(x)} = \int_{1/x}^{1}\frac{\ell(sx)}{\ell(x)}\frac{\mathrm{d}s}{s} \geq \int_{\epsilon}^{1}\frac{\ell(sx)}{\ell(x)}\frac{\mathrm{d}s}{s} \to \log(1/\epsilon) \ .$$

Thus, $\limsup_{x\to\infty} L(x)/\ell(x) \geq \log(1/\epsilon)$, and since ϵ is arbitrary, this implies that $\lim_{x\to\infty} L(x)/\ell(x)=\infty$. Next, to prove that L is slowly varying, note that for $t>0$, we have, by the Uniform Convergence Theorem,

$$L(tx) = L(x) + \ell(x)\int_{1}^{t}\frac{\ell(sx)}{\ell(x)}\frac{\mathrm{d}s}{s} \sim L(x) + \ell(x)\log(t) \ .$$

Since we have just proved that $\ell(x)=o(L(x))$, this shows that $L(tx)/L(x)\to 1$, i.e. L is slowly varying. ∎

2.1.2 Vague convergence on $\overline{\mathbb{R}}\setminus\{0\}$

A Radon measure on $\overline{\mathbb{R}}\setminus\{0\}$ is a measure on the Borel sets which is finite on relatively compact sets. The examples of relatively compact sets on $\overline{\mathbb{R}}\setminus\{0\}$ are $[a,b]$ for $a>0$,

$b \le \infty$ and $[-\infty, -1] \cup [1, \infty]$. On the other hand $[a, \infty)$ for $a > 0$ or $[a, b]$ for $a < 0$, $b \le \infty$ are not relatively compact sets. In short, a relatively compact set in $\overline{\mathbb{R}} \setminus \{0\}$ cannot contain the origin.

Definition 2.1.7 *A sequence of Radon measures $\{\nu_n, n \in \mathbb{N}\}$ on $\overline{\mathbb{R}} \setminus \{0\}$ is said to converge vaguely to a Radon measure ν, denoted by $\nu_n \xrightarrow{v} \nu$, if $\lim_{n\to\infty} \nu_n(A) = \nu(A)$ for all relatively compact Borel sets A which are continuity sets for ν, that is $\nu(\partial A) = 0$, where ∂A is the topological boundary of A.*

Vague convergence on $\overline{\mathbb{R}} \setminus \{0\}$ is characterized by the intervals $[x, \infty]$ and $[-\infty, -x]$ for $x > 0$ such that these intervals are continuity sets for the limiting measure ν. That is, it is enough to show that ν_n converges to ν on sets $[x, \infty]$ and $[-\infty, -x]$.

The vague convergence can also be characterized by functions. Denote $\nu_n(f) = \int f(x)\nu_n(\mathrm{d}x)$. The sequence ν_n converges vaguely to ν if and only if $\lim_{n\to\infty} \nu_n(f) = \nu(f)$ for all functions f which are continuous ν-almost everywhere with compact support.

The fact that f only needs to be continuous ν-almost everywhere is very important when dealing with $\pm\infty$. A function f defined on $(0, \infty]$ is continuous at infinity if and only if its restriction to $(0, \infty)$ has a limit at infinity. If the measure ν puts no mass at infinity (which will be the case of the measures of interest hereafter), then such a requirement is not needed.

Weak vs. vague convergence. Weak convergence is defined in terms of bounded continuous functions, whereas vague convergence is defined in terms of continuous functions with compact support only. This distinction makes both types of convergence significantly different. A sequence of finite measures may converge vaguely to an infinite measure, although the limiting measure must be finite on each compact set. If a sequence of probability measures converges weakly, then the limiting measure is a probability measure as well. If it converges only vaguely, then the limiting measure may not be a probability measure. For instance, it is possible that a sequence of probability measures converges vaguely, but not weakly, to the null measure.

2.1.3 Regularly varying random variables

Regular variation

Let X be a random variable. We assume that the so-called *tail balance* condition holds:

$$\lim_{x\to\infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} = p \in [0,1] \ , \tag{2.1.5}$$

for some $p \in [0,1]$. A consequence of the tail balance condition is that the right tail of $|X|$ is also regularly varying with index $-\alpha$ for some $\alpha > 0$. The parameter p is called the skewness of (the distribution of) X. If $p = 1$ then the left tail is said to be lighter than the right tail. This is the case, in particular, when X is non negative. If $p = 0$ then the right tail is lighter than the left tail. The tail balance condition excludes oscillating or other pathological behavior of the tail.

Definition 2.1.8 *A random variable X with distribution function F is said to be regularly varying with index $-\alpha$, $\alpha > 0$, if its survival function $\bar{F} = 1 - F$ is regularly varying at infinity with index $-\alpha$ and the tail balance condition (2.1.5) holds.*

We will then indifferently say that X is a regularly varying random variable with index of regular variation $-\alpha$ or with tail index α.

Vague convergence

The regular variation of a random variable X can be equivalently expressed in terms of vague convergence of measures on $\overline{\mathbb{R}} \setminus \{0\}$. To this purpose, we have the following theorem. To state it, we will call $\{c_n\}$ a *scaling sequence* if $c_n \to \infty$ if $n \to \infty$.

Theorem 2.1.9 *The random variable X is regularly varying if and only if there exist a scaling sequence $\{c_n\}$ and a non zero Radon measure ν, that has no mass at $\{\pm\infty\}$, such that the sequence of Radon measures*

$$n\mathbb{P}(c_n^{-1}X \in \cdot) \tag{2.1.6}$$

converges vaguely on $\overline{\mathbb{R}} \setminus \{0\}$ to ν.

Moreover, if c_n and c'_n are scaling sequences such that

$$n\mathbb{P}(c_n^{-1}X \in \cdot)\ , \quad n\mathbb{P}({c'_n}^{-1}X \in \cdot)$$

converge vaguely to non zero measures ν and ν' on $\overline{\mathbb{R}} \setminus \{0\}$, that have no mass at $\{\pm\infty\}$, then c_n/c'_n converges to a non zero finite limit and ν is proportional to ν'.

Note: The proof is not original, my goal was to understand it and learn the technique.

Proof: We prove that regular variation implies vague convergence for measures in (2.1.6). For simplicity we assume that X is positive.

Let X be a regularly varying random variable with distribution function F. The quantile function of F (or of X) is its left-continuous inverse $F^{\leftarrow}$, defined on $[0,1]$ by

$$F^{\leftarrow}(t) = \inf\{x : \ F(x) \geq t\}\ .$$

It can be shown that $\bar{F}$ is regularly varying at infinity with index $-\alpha < 0$, if and only if the function $\widetilde{Q}$ defined on $[1,\infty)$ by

$$\widetilde{Q}(t) = F^{\leftarrow}(1 - 1/t) \tag{2.1.7}$$

is regularly varying at infinity with index $1/\alpha$. The function $\widetilde{Q}$ will be called the upper quantile function of X, since by definition, for $t > 1$, $\widetilde{Q}(t)$ is the quantile of order $1 - 1/t$ of X. Equivalently, the function

$$Q(p) = F^{\leftarrow}(1 - p) \tag{2.1.8}$$

is regularly varying at zero with index $-1/\alpha$.

Define the sequence c_n by

$$c_n = \widetilde{Q}(n) \ . \tag{2.1.9}$$

Then it holds that

$$\lim_{n\to\infty} n\mathbb{P}(X > c_n) = 1 \ . \tag{2.1.10}$$

With this definition, the regular variation of F implies that, for all $x > 0$,

$$\lim_{x\to\infty} n\mathbb{P}(X > c_n x) = x^{-\alpha} \ .$$

Since vague convergence is characterized by the convergence on intervals $[x, \infty]$, this precisely means that the sequence of measures $n\mathbb{P}(X \in c_n \cdot)$ converges vaguely on $\overline{\mathbb{R}} \setminus \{0\}$ to the measure ν_α defined on $(0, \infty)$ by

$$\nu_\alpha(dy) = \alpha y^{-\alpha-1} \mathbb{1}_{\{y>0\}} \ , \tag{2.1.11}$$

that is,

$$n\mathbb{P}(X \in c_n \cdot) \xrightarrow{\mathrm{v}} \nu_\alpha \ . \tag{2.1.12}$$

∎

Vague convergence and conditional probabilities

Assume for simplicity that the random variable X is nonnegative and regularly varying. We have

$$n\mathbb{P}(c_n^{-1} X \in \cdot) \xrightarrow{\mathrm{v}} \nu_\alpha \ , \qquad n \to \infty \ . \tag{2.1.13}$$

The sequence c_n can be replaced by a regularly varying function c defined by $c(t) = c_{[t]}$ and

$$x\mathbb{P}\left(\frac{X}{c(x)} \in \cdot\right) \xrightarrow{\mathrm{v}} \nu_\alpha \ , \qquad x \to \infty \ .$$

The function c can be assumed to be non decreasing and thus we obtain equivalently

$$c^{\leftarrow}(x)\mathbb{P}\left(x^{-1}X \in \cdot\right) \xrightarrow{\text{v}} \nu_\alpha \;, \qquad x \to \infty \tag{2.1.14}$$

or

$$\frac{\mathbb{P}\left(x^{-1}X \in \cdot\right)}{\mathbb{P}(X > x)} \xrightarrow{\text{v}} \nu_\alpha \;, \qquad x \to \infty \;. \tag{2.1.15}$$

This can be rewritten in terms of conditional probabilities. For $y \geq 1$,

$$\lim_{x\to\infty} \mathbb{P}(X > xy \mid X > x) = y^{-\alpha} = \nu_\alpha((y, \infty]) \;. \tag{2.1.16a}$$

That is, the conditional limiting distribution on $[1, \infty)$ of X/x given $X > x$ is the standard Pareto distribution with tail index α.

Bounds and moments

We now state (and re-prove) a version of Potter's bounds for regularly varying random variables.

Lemma 2.1.10 (Potter's bounds) *Let X be a nonnegative random variable with a regularly varying right tail with index $-\alpha$, $\alpha > 0$. For each $\epsilon > 0$, there exists a constant C, such that for all $x \geq 0$ and all $y > 0$,*

$$\frac{\mathbb{P}(yX > x)}{\mathbb{P}(X > x)} \leq C(y \vee 1)^{\alpha+\epsilon} \;. \tag{2.1.17}$$

Note: The proof is not original, my goal was to understand it and learn the technique.

Proof: If $y \leq 1$, then $\mathbb{P}(yX > x) \leq \mathbb{P}(X > x)$ so the requested bound holds trivially with $C = 1$. Assume now that $y \geq 1$. Then, for $\epsilon > 0$, writing $\mathbb{P}(X > x) = x^{-\alpha}\ell(x)$, the Uniform Convergence Theorem yields

$$\frac{\mathbb{P}(yX > x)}{\mathbb{P}(X > x)} = y^\alpha \frac{\ell(x/y)}{\ell(x)} \leq y^{\alpha+\epsilon} \sup_{0<t\leq 1} \frac{\ell(tx)(tx)^\epsilon}{\ell(x)x^\epsilon} \leq C_\epsilon y^{\alpha+\epsilon} \;.$$

∎

Potter's bound can be used to prove the celebrated and useful Breiman's Lemma.

Theorem 2.1.11 (Breiman's Lemma) *Let X and Y be independent non negative random variables, such that X is regularly varying at infinity with index $-\alpha$, $\alpha > 0$ and there exists $\epsilon > 0$ such that $\mathbb{E}[Y^{\alpha+\epsilon}] < \infty$. Then XY is regularly varying with index $-\alpha$ and*

$$\lim_{x\to\infty} \frac{\mathbb{P}(XY > x)}{\mathbb{P}(X > x)} = \mathbb{E}[Y^\alpha] \ . \tag{2.1.18}$$

Note: The proof is not original, my goal was to understand it and learn the technique.

Proof: For $x > 0$, define the function $G_x(y) = \mathbb{P}(yX > x)/\mathbb{P}(X > x)$. Applying Lemma 2.1.10, for any $\epsilon > 0$ there exists a constant C such that for all $y > 0$,

$$G_x(y) \le C(1 \vee y)^{\alpha+\epsilon} \ . \tag{2.1.19}$$

By definition of regular variation, the sequence of functions G_x converges pointwise to the function $y \to y^\alpha$ as $x \to \infty$. If $\mathbb{E}[Y^{\alpha+\epsilon}] < \infty$, the bound (2.1.19) allows us to apply the bounded convergence theorem, which yields (2.1.18). ∎

We have proved a slightly stronger result than Breiman's Lemma, namely that the sequence of functions $G_x(y)$ converges to y^p in $L^p(F_Y)$ for any $p < 1+\epsilon/\alpha$, where F_Y is the distribution of Y.

Proposition 2.1.12 (Truncated Moments) *Let X be a random variable with distribution function F. If $\bar{F}$ is regularly varying at infinity with index $-\alpha$, $\alpha \ge 0$, then $\mathbb{E}[X^\beta] < \infty$ if $\beta < \alpha$ and $\mathbb{E}[X^\beta] = \infty$ if $\beta > \alpha$ and*

$$\lim_{x\to\infty} \frac{\mathbb{E}[X^\beta \mathbb{1}_{\{X\le x\}}]}{x^\beta \bar{F}(x)} = \frac{\alpha}{\beta-\alpha} \ , \ \beta > \alpha \ , \tag{2.1.20}$$

$$\lim_{x\to\infty} \frac{\mathbb{E}[X^\beta \mathbb{1}_{\{X>x\}}]}{x^\beta \bar{F}(x)} = \frac{\alpha}{\alpha-\beta}\ ,\ \alpha > \beta\ . \tag{2.1.21}$$

Moreover, $\mathbb{E}[X^\alpha]$ *may be finite or infinite, the function* $x \to \mathbb{E}[X^\alpha \mathbb{1}_{\{X\le x\}}]$ *is slowly varying at infinity and*

$$\lim_{x\to\infty} \frac{\mathbb{E}[X^\alpha \mathbb{1}_{\{X\le x\}}]}{x^\alpha \bar{F}(x)} = \infty\ . \tag{2.1.22}$$

We will use vague convergence to prove (2.1.20). The limit (2.1.21) is obtained similarly.

Note: The proof is not original, my goal was to understand it and learn the technique.

Proof: (**Vague convergence proof of (2.1.20)**)

Let ν_x be the measure defined on $(0,\infty]$ by $\nu_x((y,\infty]) = \mathbb{P}(X > xy)/\mathbb{P}(X > x)$. Then ν_x converges vaguely to ν_α on $(0,\infty]$ and

$$\frac{\mathbb{E}[X^\beta \mathbb{1}_{\{X\le x\}}]}{x^\beta \mathbb{P}(X > x)} = \int_0^1 v^\beta \nu_x(\mathrm{d}v)\ .$$

For $\epsilon > 0$, by vague convergence, it holds that

$$\int_\epsilon^1 v^\beta \nu_x(\mathrm{d}v) = \int_\epsilon^1 v^\beta \nu_\alpha(\mathrm{d}v) = \alpha \int_\epsilon^1 v^{\alpha-\beta-1}\mathrm{d}v = \frac{\alpha(1-\epsilon^{\beta-\alpha})}{\beta-\alpha}\ .$$

When $\epsilon \to 0$, the latter quantity converges to the required limit. Thus, there only remains to prove that

$$\lim_{\epsilon\to 0} \int_0^\epsilon v^\beta \nu_x(\mathrm{d}v) = 0\ . \tag{2.1.23}$$

By the Uniform Convergence Theorem, we have, for η such that $\beta - \alpha - \eta > 0$,

$$\begin{aligned}\int_0^\epsilon v^\beta \nu_x(\mathrm{d}v) &= \beta \int_0^\epsilon v^{\beta-\alpha-1-\eta} \frac{\ell(xv)(xv)^\eta}{\ell(x)x^\eta}\,\mathrm{d}v \\ &\le C_\epsilon \int_0^\epsilon v^{\beta-\alpha-1-\eta}\mathrm{d}v = O(\epsilon^{\beta-\alpha-\eta})\ .\end{aligned}$$

This yields (2.1.23) and concludes the proof. ∎

The limits (2.1.20) and (2.1.21) entail the following bounds. There exists a constant C_β such that, for all $x \geq 0$,

$$\mathbb{E}[X^\beta \mathbb{1}_{\{X \leq x\}}] \leq C_\beta x^\beta \bar{F}(x) \ , \quad \alpha < \beta \tag{2.1.24}$$

$$\mathbb{E}[X^\beta \mathbb{1}_{\{X > x\}}] \leq C_\beta x^\beta \bar{F}(x) \ , \quad \alpha > \beta \ . \tag{2.1.25}$$

2.1.4 Examples

- Standard Pareto. Let X be a Pareto random variable with distribution function $F(x) = 1 - x^{-\alpha}$, for $x \geq 1$. Then X is regularly varying with index $-\alpha$.

- Reciprocal of normal.

Lemma 2.1.13 *If $Y \sim N(0,1)$, then $X = |Y|^{-1/\beta}$, $\beta > 0$, is regularly varying with index $-\beta$.*

Proof: Clearly

$$\mathbb{P}(X > x) = \mathbb{P}\left(\frac{1}{|Y|^{1/\beta}} > x\right) = \mathbb{P}\left(|Y| < \frac{1}{x^\beta}\right) \ .$$

Let $z = 1/x^\beta$, then $z \to 0$ when $x \to \infty$. We have

$$\mathbb{P}\left(X > x\right) = \mathbb{P}(|Y| < z) = 2\mathbb{P}(0 < Y < z) \sim \frac{2}{\sqrt{2\pi}}\, z \quad \text{as } z \to 0 \ .$$

Hence,

$$P(X > x) \sim \frac{2}{\sqrt{2\pi}} x^{-\beta} \quad \text{as } x \to \infty \ .$$

∎

2.2 Multivariate regular variation

In this section we consider a multivariate regular variation, an extension of the previously studied univariate one. Specifically,

- In Section 2.2.1 we define the multivariate regular variation and discuss its consequences. The material is based on classical theory, e.g. [28];
- In Section 2.2.3 we introduce the Marginal Expected Shortfall (MES), one of the risk measures used in finance. MES is well defined under multivariate regular variation. However, it can vanish and this will motivate us later to consider Conditional Extreme Value model in Section 2.4.
- Section 2.2.4 deals with a quasi-spectral decomposition of regularly varying random vectors. This is a recent tool, developed in the context of time series by [3].

2.2.1 Definition

We start with the following definition of regularly varying random vectors. We denote $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^d$.

Definition 2.2.1 *A vector $\boldsymbol{X} = (X_1, \dots, X_d)$ in $\mathbb{R}^d$ is (multivariate) regularly varying if there exists a non zero Radon measure $\boldsymbol{\nu}_{\boldsymbol{X}}$ on $\overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$, called the exponent measure of $\boldsymbol{X}$, such that $\boldsymbol{\nu}_{\boldsymbol{X}}(\overline{\mathbb{R}}^d \setminus \mathbb{R}^d) = 0$ and a scaling sequence $\{c_n\}$ such that the measure $n\mathbb{P}(c_n^{-1}\boldsymbol{X} \in \cdot)$ converges vaguely on $\overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$ to the measure $\boldsymbol{\nu}_{\boldsymbol{X}}$, i.e.*

$$n\mathbb{P}(c_n^{-1}\boldsymbol{X} \in \cdot) \stackrel{v}{\to} \boldsymbol{\nu}_{\boldsymbol{X}} \ , \quad on\ \overline{\mathbb{R}}^d \setminus \{\mathbf{0}\} \ . \tag{2.2.1}$$

The limiting measure is homogeneous with some index $-\alpha$, $\alpha > 0$. Let C be a relatively compact set in $\overline{\mathbb{R}}^d$. For $y > 0$ denote $yC = \{yc : c \in C\}$. Then $\boldsymbol{\nu}_{\boldsymbol{X}}(yC) = y^{-\alpha}\boldsymbol{\nu}_{\boldsymbol{X}}(C)$. We call $-\alpha$ *the index of regular variation* of $\boldsymbol{X}$.

The above definition implies that either all marginal distributions are regularly varying with the same index $-\alpha$ or there exists at least one such marginal and others are lighter. In the latter case the limiting measure is degenerated in the sense that it is concentrated on a subspace of $\mathbb{R}^d$. For example, if $d = 2$ and X_1, X_2 are regularly varying with the indices $-\alpha, -\beta$, $\beta > \alpha$, then the limiting measure is concentrated on the axis that corresponds to X_1.

To cover the case of unequal margins, the concept of multivariate regular variation can be extended to:

Definition 2.2.2 *A vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$ is (multivariate) regularly varying if there exists a non zero Radon measure $\boldsymbol{\nu}_{\boldsymbol{X}}$ on $\overline{\mathbb{R}}^d \setminus \{\boldsymbol{0}\}$, called the exponent measure of $\boldsymbol{X}$, such that $\boldsymbol{\nu}_{\boldsymbol{X}}(\overline{\mathbb{R}}^d \setminus \mathbb{R}^d) = 0$ and a scaling vector $\{\boldsymbol{c}_n\}$, where $\boldsymbol{c}_n = (c_{n,1}, \ldots, c_{n,d})$, such that the measure $n\mathbb{P}(\boldsymbol{c}_n^{-1}\boldsymbol{X} \in \cdot)$ converges vaguely on $\overline{\mathbb{R}}^d \setminus \{\boldsymbol{0}\}$ to the measure $\boldsymbol{\nu}_{\boldsymbol{X}}$, i.e.*

$$n\mathbb{P}(\boldsymbol{c}_n^{-1}\boldsymbol{X} \in \cdot) \stackrel{\mathrm{v}}{\to} \boldsymbol{\nu}_{\boldsymbol{X}} \ , \quad \text{on } \overline{\mathbb{R}}^d \setminus \{\boldsymbol{0}\} \ . \tag{2.2.2}$$

Here, for two vectors $\boldsymbol{a} = (a_1, \ldots, a_d)$ and $\boldsymbol{b} = (b_1, \ldots, b_d)$, the operation $\boldsymbol{a}/\boldsymbol{b}$ is the coordinatewise division.

Here and in the sequel, we are focusing on the situation described in Definition 2.2.1. Now, we proceed as in the univariate case. The sequence c_n in Definition 2.2.1 can be replaced by a regularly varying function c defined by $c(x) = c_{[x]}$ and (2.2.1) becomes

$$x\mathbb{P}\left(\frac{\boldsymbol{X}}{c(x)} \in \cdot\right) \stackrel{\mathrm{v}}{\to} \boldsymbol{\nu}_{\boldsymbol{X}} \ , \qquad x \to \infty \ .$$

The latter convergence is equivalent to

$$c^{\leftarrow}(x)\mathbb{P}\left(\frac{\boldsymbol{X}}{x} \in \cdot\right) \stackrel{\mathrm{v}}{\to} \boldsymbol{\nu}_{\boldsymbol{X}} \ , \qquad x \to \infty \ . \tag{2.2.3}$$

Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. The homogeneity of the measure $\boldsymbol{\nu}_{\boldsymbol{X}}$ and the application of the convergence (2.2.3) to the set $\{\boldsymbol{x} : \|\boldsymbol{x}\| > y\}$ yields, for all $y > 0$,

$$\lim_{x\to\infty} c^{\leftarrow}(x)\mathbb{P}(\|\boldsymbol{X}\| > xy) = \boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| > y\}) = y^{-\alpha}\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1\}) \; .$$

Since $\boldsymbol{\nu}_{\boldsymbol{X}}$ is not identically zero, it must hold that $\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1\}) > 0$. This proves that $\|\boldsymbol{X}\|$ is regularly varying with index $-\alpha$ and

$$\frac{\mathbb{P}(x^{-1}\boldsymbol{X} \in \cdot)}{\mathbb{P}(\|\boldsymbol{X}\| > x)} \xrightarrow{\mathrm{v}} \frac{\boldsymbol{\nu}_{\boldsymbol{X}}(\cdot)}{\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| \geq 1\})} \; , \tag{2.2.4}$$

where again vague convergence holds on $\overline{\mathbb{R}}^d \setminus \{\boldsymbol{0}\}$. If the sequence c_n in (2.2.1) is chosen as $c_n = \widetilde{Q}_{\|\boldsymbol{X}\|}(n)$ (cf. (2.1.7)), then $\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| \geq 1\}) = 1$.

The vague convergence and the homogeneity property imply also that for each $i = 1, \ldots, d$ and $y > 0$ we have

$$\lim_{x\to\infty} \frac{\mathbb{P}(|X_i| > xy)}{\mathbb{P}(\|\boldsymbol{X}\| > x)} = \vartheta_i y^{-\alpha} \; ,$$

with

$$\vartheta_i = \frac{\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : |x_i| > 1\})}{\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : \|\boldsymbol{x}\| > 1\})} \; .$$

It may happen that $\vartheta_i = 0$ for some, but not all, indices i. This means that some of the components may have a lighter tail than the others, but there must exist at least one component with tail index α. For instance, if X_1 is a Pareto random variable and X_2 is has a standard exponential distribution (not necessarily independent of X_1), then (X_1, X_2) is regularly varying. The exponent measure is concentrated on the x axis, and $\vartheta_2 = 0$. (We described such a situation previously to motivate Definition 2.2.2).

However, if all the random variables X_i, $i = 1, \ldots, d$, have the same distribution, then necessarily $\vartheta_1 = \cdots = \vartheta_d > 0$ and all the subvectors of $\boldsymbol{X}$ are regularly varying with the same tail index. In such a case (2.2.4) yields

$$\boldsymbol{\nu}_{\boldsymbol{d},\boldsymbol{x}}(\cdot) := \frac{\mathbb{P}(x^{-1}\boldsymbol{X} \in \cdot)}{\mathbb{P}(|X_1| > x)} \xrightarrow{\mathrm{v}} \frac{\boldsymbol{\nu}_{\boldsymbol{X}}(\cdot)}{\boldsymbol{\nu}_{\boldsymbol{X}}(\{\boldsymbol{x} : |x_1| \geq 1\})} =: \boldsymbol{\nu}^*_{\boldsymbol{X}}(\cdot) \; . \tag{2.2.5}$$

Note that the restriction of $\boldsymbol{\nu_{d,x}}$ to $[1,\infty)\times\mathbb{R}^d$ is in fact a probability measure. Hence, the above vague convergence becomes weak convergence.

We finish this subsection with the important distinction between extremal independence and extremal dependence.

Definition 2.2.3 *If the limiting measure $\boldsymbol{\nu_X}$ in Definition 2.2.1 is concentrated on the axes we say that $\boldsymbol{X}$ has extremal independence. Otherwise, the vector has extremal dependence.*

Two typical situations for extremal independence are as follows. Let $d = 2$: If X_2 has lighter tail than X_1, then the limiting measure is concentrated on horizontal axis. If X_1 and X_2 have the same distribution but extremes of the same order do not occur together (for example when X_1 and X_2 are independent), then the limiting measure is concentrated on both axes.

In what follows we will need the following lemma.

Lemma 2.2.4 *Let $\boldsymbol{X}$ be a regularly varying random vector such that all components have the same distribution function F and are regularly varying with the tail index α. Let $\psi : \mathbb{R}^d \to \mathbb{R}_+$ be a homogenous function with index γ. Let C be a relatively compact set in $\overline{\mathbb{R}}^d \setminus \{\mathbf{0}\}$, $s > \epsilon > 0$ and assume that for some $\delta > 0$ and $x_0 > 0$,*

$$\sup_{x>x_0} \frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi^{1+\delta}\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\right] < \infty\ . \tag{2.2.6}$$

Then

$$\lim_{x\to\infty} \frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\right] = s^{\gamma-\alpha}\int_C \psi(\boldsymbol{v})\boldsymbol{\nu_X}(\mathrm{d}\boldsymbol{v})\ .$$

Proof: Since ψ is homogenous, it is also continuous. We note that (2.2.6) implies that

$$\lim_{x\to\infty}\int_C \psi^{1+\delta}(\boldsymbol{v})\boldsymbol{\nu_X}(\mathrm{d}\boldsymbol{v}) < \infty\ . \tag{2.2.7}$$

Let $\gamma \geq 0$ (the proof for $\gamma < 0$ is analogous). For any $A > 0$ we have

$$\frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\right] =$$
$$= \frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\mathbb{1}_{\{\psi(\boldsymbol{X}/x)\leq A\}}\right] + \frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\mathbb{1}_{\{\psi(\boldsymbol{X}/x)> A\}}\right].$$

- First term: we can just use the vague convergence

$$\lim_{x\to\infty}\frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\psi(\boldsymbol{X}/x)\leq A\}}\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\right] =$$
$$\lim_{x\to\infty}\int_{sxC}\psi\left(\frac{\boldsymbol{v}}{x}\right)\mathbb{1}_{\{\psi(\boldsymbol{v}/x)\leq A\}}\frac{F(\mathrm{d}\boldsymbol{v})}{\bar{F}(x)} = \lim_{x\to\infty}\int_{C}\psi\left(s\boldsymbol{z}\right)\mathbb{1}_{\{\psi(s\boldsymbol{z})\leq A\}}\frac{F(sx\mathrm{d}\boldsymbol{z})}{\bar{F}(x)}$$
$$= s^{\gamma}\lim_{x\to\infty}\int_{C}\psi\left(\boldsymbol{z}\right)\mathbb{1}_{\{\psi(s\boldsymbol{z})\leq A\}}\boldsymbol{\nu}_{\boldsymbol{X}}(s\mathrm{d}\boldsymbol{z})$$
$$= s^{\gamma-\alpha}\int_{C}\psi(\boldsymbol{z})\mathbb{1}_{\{\psi(s\boldsymbol{z})\leq A\}}\boldsymbol{\nu}_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{z})$$

  and we let $A \to \infty$ which is allowed by (2.2.7).

- Second term: We have for $\delta > 0$,

$$\frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\mathbb{1}_{\{\psi(\boldsymbol{X}/x)> A\}}\right] \leq A^{-\delta}\frac{1}{\bar{F}(x)}\mathbb{E}\left[\psi^{1+\delta}\left(\frac{\boldsymbol{X}}{x}\right)\mathbb{1}_{\{\boldsymbol{X}\in sxC\}}\right]$$

  and the term vanishes by letting first $x \to \infty$, then using (2.2.6) and finally letting $A \to \infty$.

∎

2.2.2 Tail Dependence Coefficient

Consider a bivariate case (X_1, X_2). Assume that (X_1, X_2) is regularly varying in the sense of Definition 2.2.1 with the exponent measure $\boldsymbol{\nu}_{\boldsymbol{X}}$. Then the following limit exists:

$$\lim_{x\to\infty}\frac{\mathbb{P}(X_1 > x, X_2 > x)}{\mathbb{P}(X_1 > x)} = \lim_{x\to\infty}\mathbb{P}(X_2 > x \mid X_1 > x) \in [0,1]\,. \tag{2.2.8}$$

As mentioned above, the limit is nonzero in case of extremal dependence, while it vanishes in case of extremal independence. Again, as mentioned before, extremal independence can occur in several way, in particular, when X_2 has lighter tail than X_1. In the latter situation, it is however possible that the following limit does not vanish:

$$\lambda_{\mathrm{TDC}} := \lim_{p \to 0} \mathbb{P}(X_2 > Q_{X_2}(p) \mid X_1 > Q_{X_1}(p)) \ , \tag{2.2.9}$$

where recall that for a random variable X, $Q_X(p) = F_X^{\leftarrow}(1-p)$; cf. (2.1.8). The above quantity is called the **Tail Dependence Coefficient**. If $\lambda_{\mathrm{TDC}} = 0$ then we will say that X_1 and X_2 are *asymptotically independent*, otherwise if $\lambda_{\mathrm{TDC}} \in (0, 1]$, then they are *asymptotically dependent*.

Summarizing, if $\boldsymbol{X} = (X_1, X_2)$ is regularly varying, then

- $\boldsymbol{X}$ is extremally independent if the exponent measure in the Definition 2.2.1 is concentrated on the axes and in particular the limit in (2.2.8) vanishes;
- $\boldsymbol{X}$ is asymptotically independent if the tail dependence coefficient is zero;
- It is possible that the vector is extremally independent, but asymptotically dependent, like in the case of $X_2 = X_1^{\phi}$, $\phi \in (0, 1)$;
- The concept of the tail dependence coefficient can be extended to arbitrary random variables, not only those that are regularly varying.

2.2.3 Marginal Expected Shortfall

Financial institutions are required to calculate *risk* measures (credit risk, systemic risk, operating risk among others). One of the typical approaches is to evaluate Value-At-Risk of a loss variable X with a distribution function F_X, that is the quantile

$Q_X(p) = F_X^{\leftarrow}(1-p)$ of the loss distribution function, when p is small. The Value-At-Risk has some properties that are not desirable, like not being a coherent risk measure (cf. [1]). Furthermore, it does not measure severity of the loss, when it occurs. Alternatively, one can consider the conditional value-at-risk, which in the case of continuous random variables is equivalent to Expected Shortfall (ES) defined (if it exists) as

$$\mathbb{E}[X \mid X > Q_X(p)] \,.$$

This idea can be extended to multivariate portfolios. We will consider the bivariate case $d = 2$ only and write (X, Y) for (X_1, X_2). In particular, Marginal Expected Shortfall (MES) is defined as the expected loss of an equity given the occurrence of an extreme loss in the aggregated portfolio. In this set-up, the previously mentioned random variable X denotes the portfolio loss, while the new random variable Y describes the loss from one particular equity (another possible interpretation is that X and Y denote a loss of the entire financial market and of the particular company, respectively). In mathematical terms we can write MES as

$$\theta(p) := \mathbb{E}[Y \mid X > Q_X(p)] \,. \tag{2.2.10}$$

Under the assumptions of Lemma 2.2.4 (see also Corollary 2.4.4 below in a more general context of the Conditional Extreme Value model) we have

$$\lim_{p\to 0} \mathbb{E}\left[\frac{Y}{Q_X(p)} \mid X > Q_X(p)\right] = \int_1^\infty \int_0^\infty v\boldsymbol{\nu}(\mathrm{d}v, \mathrm{d}u) \,, \tag{2.2.11}$$

where we write $\boldsymbol{\nu}$ for the limiting exponent measure of the vector (X, Y) as in Definition 2.2.1.

As mentioned before, there exist two fundamentally different cases: either the exponent measure $\boldsymbol{\nu}$ is concentrated on the axes or it is not. The former means extremal independence and the latter extremal dependence. In case of extremal dependence the right hand side of (2.2.11) does not vanish and multivariate regular variation can be used to study MES. As for the extremal independence we consider again two situations: either the exponent measure is concentrated on both horizontal and vertical axes, like in case of two independent regularly varying random variables with the same distribution (let's call it the situation A), or $\boldsymbol{\nu}$ lives only on, say, the horizontal axis. The latter situation appears e.g. when Y has asymptotically lighter tail than X, like in the trivial situation $Y = X^{\phi}$, $\phi \in (0,1)$ (we call it the situation B). Regardless the situation, in case of extremal independence, the limit on the right hand side of (2.2.11) vanishes. We do not gain useful information about the expected shortfall from the regular variation of the vector (X,Y).

The assumption about the regular variation of (X,Y) can be modified as follows. Let F_Y and $F_Y^{\leftarrow}$ be the distribution and the quantile functions of Y. If Y is regularly varying with index $\alpha > 1$ and the limit

$$\lim_{p\to 0} \frac{1}{p}\mathbb{P}(X > F_X^{\leftarrow}(1-px), Y > F_Y^{\leftarrow}(1-py)) = R(x,y) \tag{2.2.12}$$

exists for $(x,y) \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$, then (cf. [8])

$$\lim_{p\to 0} \mathbb{E}\left[\frac{Y}{Q_Y(p)} \mid X > Q_X(p)\right] = \int_{x=0}^{\infty} R(x^{-\alpha},1)\mathrm{d}x \, . \tag{2.2.13}$$

We note that in (2.2.11) and (2.2.13) we have possibly a different scaling factors in the denominator. This approach can also possibly resolve the situation B described above, since (2.2.12) is in fact the tail dependence condition for the "standardized" variables $F_X(X)$, $F_Y(Y)$. That is, $R(1,1)$ is the tail dependence coefficient defined in (2.2.9).

However, it is still possible that $R(x,y) \equiv 0$. This would be typically the situation A described above. Such situations will be resolved in Section 2.4.1.

2.2.4 Quasi-spectral decomposition

In this section we assume for simplicity that all random variables are nonnegative. As in the univariate case, we can link vague convergence to weak convergence of conditional probabilities. For example, regular variation implies for $y \geq 1$ and A, B, relatively compact sets in $\overline{\mathbb{R}}^{i-1} \setminus \{\mathbf{0}\}$, $\overline{\mathbb{R}}^{d-i} \setminus \{\mathbf{0}\}$,

$$\lim_{x\to\infty} \mathbb{P}(x^{-1}\boldsymbol{X} \in A \times (y,\infty] \times B \mid X_i > x) = \frac{\boldsymbol{\nu_X}(A \times (y,\infty] \times B)}{\boldsymbol{\nu_X}(\overline{\mathbb{R}}^{i-1} \times (1,\infty] \times \overline{\mathbb{R}}^{d-i})} \ . \qquad (2.2.14)$$

In this spirit, regular variation implies the following *quasi-spectral decomposition.*

Proposition 2.2.5 *Let* $\boldsymbol{X}$ *be a regularly varying random vector with non-negative components and such that all the components are regularly varying with index* $-\alpha$*. Then conditionally on* $X_1 > x$*,*

$$x^{-1}(X_1, \ldots, X_d) \ , \qquad \left(\frac{X_1}{x}, \frac{X_2}{X_1}, \ldots, \frac{X_d}{X_1}\right)$$

converge in distribution as $x \to \infty$ *to random vectors* $(V_1, \ldots, V_d)$ *and* $(V_1, \Theta_2, \ldots, \Theta_d)$*, respectively where*

1. V_1 *has the standard Pareto distribution;*

2. $\Theta_j = V_j/V_1$, $j = 2, \ldots, d$ *and* $(\Theta_2, \ldots, \Theta_d)$ *is independent of* V_1*.*

Note: The proof follows the lines of [3].

Proof: Since $\boldsymbol{X}$ is regularly varying we have for $A \in \mathbb{R}^{d-1}$,

$$\lim_{x\to\infty} \frac{\mathbb{P}(x^{-1}\boldsymbol{X} \in (y,\infty] \times A)}{\mathbb{P}(X_1 > x)} = \frac{\boldsymbol{\nu_X}((y,\infty] \times A)}{\boldsymbol{\nu_X}((y,\infty] \times \mathbb{R}^{d-1})} \ .$$

If moreover $y \geq 1$, the left hand side becomes the conditional probability

$$\lim_{x\to\infty} \mathbb{P}(x^{-1}\boldsymbol{X} \in (y,\infty] \times A \mid X_1 > x) \ .$$

In other words, conditionally on $X_1 > x$, $x^{-1}\boldsymbol{X}$ converges weakly to a random vector, say $\boldsymbol{V} = (V_1, \dots, V_d)$. Therefore, for any $f : \mathbb{R}^d \to \mathbb{R}$ bounded and continuous we have

$$\lim_{x\to\infty} \mathbb{E}\left[f\left(x^{-1}\boldsymbol{X}\right) \mid X_1 > x\right] = \mathbb{E}[f(\boldsymbol{V})] .$$

Now, let $g : \mathbb{R}^d \to \mathbb{R}$ be bounded and continuous. Then

$$\mathbb{E}\left[g\left(\frac{X_1}{x}, \frac{X_2}{X_1}, \dots, \frac{X_d}{X_1}\right) \mid X_1 > x\right] = \mathbb{E}\left[f\left(\frac{X_1}{x}, \frac{X_2}{x}, \dots, \frac{X_d}{x}\right) \mid X_1 > x\right] ,$$

where $f(u_1, \dots, u_d) = g(u_1, u_2/u_1, \dots, u_d/u_1)$ is also bounded and continuous whenever $u_1 \geq 1$. Hence,

$$\lim_{x\to\infty} \mathbb{E}\left[g\left(\frac{X_1}{x}, \frac{X_2}{X_1}, \dots, \frac{X_d}{X_1}\right) \mid X_1 > x\right] = \mathbb{E}\left[f(V_1, \dots, V_d)\right] = \mathbb{E}\left[g(V_1, V_2/V_1, \dots, V_d/V_1)\right] .$$

In other words, conditionally on $X_1 > x$,

$$\left(\frac{X_1}{x}, \frac{X_2}{X_1}, \dots, \frac{X_d}{X_1}\right)$$

converges in distribution to $(V_1, V_2/V_1, \dots, V_d/V_1) = (V_1, \Theta_2, \dots, \Theta_d)$. It is obvious that V_1 has a standard Pareto distribution. We claim that V_1 is independent of $(\Theta_2, \dots, \Theta_d)$. Indeed, for $A_i \subseteq \mathbb{R}$, $i = 2, \dots, d$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{X_1}{x} > y, \frac{X_2}{X_1} \in A_2, \cdots, \frac{X_d}{X_1} \in A_d \mid X_1 > x\right) \\
&= \frac{\mathbb{P}\left(\frac{X_1}{x} > y, \frac{X_2}{X_1} \in A_2, \cdots, \frac{X_d}{X_1} \in A_d, X_1 > x\right)}{\mathbb{P}(X_1 > x)} \\
&= \frac{\mathbb{P}\left(X_1 > xy, \frac{X_2}{X_1} \in A_2, \cdots, \frac{X_d}{X_1} \in A_d\right)}{\mathbb{P}(X_1 > xy)} \frac{\mathbb{P}(X_1 > xy)}{\mathbb{P}(X_1 > x)} \\
&= \mathbb{P}\left(\frac{X_2}{X_1} \in A_2, \cdots, \frac{X_d}{X_1} \in A_d \mid X_1 > xy\right) \frac{\mathbb{P}(X_1 > xy)}{\mathbb{P}(X_1 > x)} \\
&\to \mathbb{P}\left(\frac{V_2}{V_1} \in A_2, \cdots, \frac{V_d}{V_1} \in A_d\right) \mathbb{P}(V_1 > y).
\end{aligned}$$

On the other hand,

$$\mathbb{P}\left(\frac{X_1}{x} > y, \frac{X_2}{X_1} \in A_2, \cdots, \frac{X_d}{X_1} \in A_d \mid X_1 > x\right)$$
$$\to \mathbb{P}\left(V_1 > y, \frac{V_2}{V_1} \in A_2, \cdots, \frac{V_d}{V_1} \in A_d\right).$$

Hence, $(\Theta_2 = V_2/V_1, \ldots, \Theta_d = V_d/V_1)$ and V_1 are independent. ∎

Remark 2.2.6 *Throughout the paper the quasi spectral-decomposition into V_1 and $(\Theta_2, \ldots, \Theta_d)$ is obtained by conditioning on X_1. We can condition on X_j for any j. Note however that for each different j we get different vectors $\boldsymbol{V}$ (that depend formally on j).*

Remark 2.2.7 *In extreme value theory it is a common practice to standardize marginals. Define*

$$U(x) = \left(\frac{1}{1-F}\right)^{\leftarrow}(x)\,.$$

Then $V_j = U^{\leftarrow}(X_j)$, $j = 1, \ldots, d$, are regularly varying with index 1. Under the conditions of Proposition 2.2.5, the decomposition of the limiting conditional law of $V_1, \ldots, V_d$ becomes V_1', $\Theta_2', \ldots, \Theta_d'$, where V_1' is the standard Pareto, while $\Theta_j' = \Theta_j^{\alpha}$, $j = 1, \ldots, d$.

Representation of conditional tail distribution

We use the quasi-spectral representation to express the conditional tail distribution. The proof follows closely [3].

Proposition 2.2.8 *Let $\boldsymbol{X}$ be a regularly varying random vector with non-negative components and such that all the components are regularly varying with index $-\alpha$. Then for $j_2, \ldots, j_l, j_{l+1}, \ldots, j_d \in \{1, \ldots, d\}$ and $y_j > 0$ we have*

$$\lim_{x\to\infty} \mathbb{P}(X_{j_{l+1}} > y_{j_{l+1}}x, \ldots, X_{j_d} > y_{j_d}x \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x)$$

$$= \frac{\mathbb{E}\left[\left(\frac{\Theta_{j_{l+1}}}{y_{j_{l+1}}}\right)^\alpha \wedge \cdots \wedge \left(\frac{\Theta_{j_d}}{y_{j_d}}\right)^\alpha \wedge \Theta_{j_2}^\alpha \wedge \cdots \Theta_{j_l}^\alpha \wedge 1\right]}{\mathbb{E}\left[\Theta_{j_2}^\alpha \wedge \cdots \Theta_{j_l}^\alpha \wedge 1\right]} . \tag{2.2.15}$$

Proof: Proposition 2.2.5 implies that for $y_1 \geq 1$, $y_2, \ldots, y_d > 0$,

$$\lim_{x\to\infty} \mathbb{P}(X_1 > y_1 x, \ldots, X_d > y_d x \mid X_1 > x) = \mathbb{P}(V_1 > y_1, \ldots, V_d > y_d)$$
$$= \mathbb{P}(V_1 > y_1, V_1\Theta_2 > y_2, \ldots, V_1\Theta_d > y_d)$$
$$= \alpha \int_{y_1 \vee 1}^{\infty} \mathbb{P}(\Theta_2 > y_2/u, \ldots, \Theta_d > y_d/u) u^{-\alpha-1} \mathrm{d}u$$
$$= \alpha \int_{y_1 \vee 1}^{\infty} \mathbb{P}\left(\left(\frac{\Theta_2}{y_2}\right)^\alpha > u^{-\alpha}, \ldots, \left(\frac{\Theta_d}{y_d}\right)^\alpha > u^{-\alpha}\right) u^{-\alpha-1} \mathrm{d}u$$
$$= \mathbb{E}\left[\left(\frac{1}{y_1} \wedge \frac{\Theta_2}{y_2} \wedge \cdots \frac{\Theta_d}{y_d}\right)^\alpha\right] .$$

Furthermore,

$$\mathbb{P}(X_{j_{l+1}} > y_{j_{l+1}} x, \ldots, X_{j_d} > y_{j_d} x \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x) =$$
$$= \frac{\mathbb{P}(X_{j_2} > x, \ldots, X_{j_l} > x, X_{j_{l+1}} > y_{j_{l+1}} x, \ldots, X_{j_d} > y_{j_d} x \mid X_1 > x)}{\mathbb{P}(X_{j_2} > x, \ldots, X_{j_l} > x \mid X_1 > x)}$$

and the result follows. ∎

We note that the numerator and the denominator in (2.2.15) can be expressed as limits. In particular, via Proposition 2.2.5, the numerator in (2.2.15) equals

$$\lim_{x\to\infty} \mathbb{E}\left[g\left(\frac{X_{j_2}}{X_1}, \ldots, \frac{X_{j_l}}{X_1}, \frac{X_{j_{l+1}}}{y_{j_{l+1}} X_1}, \ldots, \frac{X_{j_d}}{y_{j_d} X_1}\right) \mid X_1 > x\right]$$

with a bounded and continuous function $g(u_2, \ldots, u_d) = (u_2 \wedge \cdots \wedge u_d \wedge 1)^\alpha$. Consequently, for $y > 0$, and setting $(X_1, X_2) = (X, Y)$

$$\lim_{x\to\infty} \mathbb{P}(Y > yx \mid X > x) = \mathbb{E}\left[\left(\frac{\Theta_2}{y} \wedge 1\right)^\alpha\right] = \lim_{x\to\infty} \mathbb{E}\left[\left(\frac{Y}{yX} \wedge 1\right)^\alpha \mid X > x\right] . \tag{2.2.16}$$

This representation will be used for statistical inference.

Representation of conditional tail expectation

The proof of the following result is original.

Proposition 2.2.9 *Let $\boldsymbol{X}$ be a regularly varying random vector with non-negative components and such that all the components are regularly varying with index $-\alpha$. Assume moreover that for some $\delta > 0$ we have*

$$\sup_{x>0} \mathbb{E}\left[\left(\frac{X_{j_d}}{x}\right)^{1+\delta} \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x\right] < \infty \, . \tag{2.2.17}$$

Then

$$\mathbb{E}\left[\frac{X_{j_d}}{x} \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x\right] = \frac{\alpha}{\alpha-1} \frac{\mathbb{E}\left[\Theta_{j_d}\left(\Theta_{j_2} \wedge \cdots \wedge \Theta_{j_l} \wedge 1\right)^{\alpha-1}\right] \, .}{\mathbb{E}[\Theta_{j_1}^{\alpha} \wedge \ldots \wedge \Theta_{j_l}^{\alpha} \wedge 1]}.$$

Proof: We note first that (2.2.17) implies that $\alpha > 1$. Let $A \subseteq (0, \infty)$. Proposition 2.2.5 implies that as $x \to \infty$

$$\mathbb{E}\left[\frac{X_{j_d}}{x} \mathbb{1}_{\{X_{j_d} \le xA\}} \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x\right] \to \frac{\mathbb{E}[V_{j_d} \mathbb{1}_{\{1 < V_{j_d} \le A\}} \mathbb{1}_{\{V_1 > 1, V_{j_2} > 1, \ldots, V_{j_l} > 1\}}]}{\mathbb{E}[\Theta_{j_1}^{\alpha} \wedge \ldots \wedge \Theta_{j_l}^{\alpha} \wedge 1]} \, .$$

A computation similar to Corollary 2.2.8 yields that the numerator in the last expression is

$$\frac{\alpha}{\alpha-1} \mathbb{E}\left[\Theta_{j_d}\left(\Theta_{j_2} \wedge \cdots \wedge \Theta_{j_l} \wedge 1\right)^{\alpha-1}\right] \, .$$

Furthermore, (2.2.17) implies

$$\begin{aligned} &\lim_{A\to\infty} \limsup_{x\to\infty} \mathbb{E}\left[\frac{X_{j_d}}{x} \mathbb{1}_{\{X_{j_d} > xA\}} \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x\right] \\ &\le \lim_{A\to\infty} A^{-\delta} \limsup_{x\to\infty} \mathbb{E}\left[\left(\frac{X_{j_d}}{x}\right)^{1+\delta} \mid X_1 > x, X_{j_2} > x, \ldots, X_{j_l} > x\right] = 0 \, . \end{aligned}$$

∎

In particular, if $\alpha > 1$ then setting again $(X_1, X_2) = (X, Y)$,

$$\lim_{x\to\infty} \mathbb{E}\left[\frac{Y}{x} \mid X > x\right] = \frac{\alpha}{\alpha-1} \mathbb{E}\left[\Theta_2\right] = \frac{\alpha}{\alpha-1} \lim_{x\to\infty} \mathbb{E}\left[\frac{Y}{X} \mid X > x\right] =: \aleph_{\mathrm{CTE}} \tag{2.2.18}$$

and the limit is strictly positive in case of extremal dependence, that is when the limiting exponent measure $\boldsymbol{\nu}_{\boldsymbol{X}}$ in (2.2.1) is not concentrated on the axes.

Quasi-spectral vs. spectral decomposition

We recall the following well-known result (see [28, Theorem 6.1]). Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$ and let $\mathbb{S}^{d-1}$ be the associated unit sphere. A vector $\boldsymbol{X}$ in $\mathbb{R}^d$ is regularly varying with index $-\alpha$ if and only if there exists a constant $\vartheta > 0$, a probability measure $\boldsymbol{\Lambda}^*$ on $\mathbb{S}^{d-1}$ and a scaling sequence $\{c_n\}$ such that, for all $y > 0$,

$$n\mathbb{P}\left(\|\boldsymbol{X}\| > c_n y, \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|} \in \cdot\right) \xrightarrow{\mathrm{v}} \vartheta y^{-\alpha}\boldsymbol{\Lambda}^* \ , \ n \to \infty \ ,$$

where vague convergence holds on $(0,\infty] \times \mathbb{S}^{h-1}$. The crucial difference is that in the quasi-spectral decomposition the limiting vector does not lie on the unit sphere.

2.2.5 Examples

The following lemma shows how to generate regularly varying random vectors. The result can be considered as standard in theory of regularly varying random vectors, but we provide a proof of it.

Lemma 2.2.10 *Assume that*

- *R is regularly varying with index $-\alpha$;*
- *$\boldsymbol{Z} = (Z_1, Z_2)$ is such that $\mathbb{E}|Z_1|^{\alpha+\delta} + \mathbb{E}|Z_2|^{\alpha+\delta} < \infty$ for some $\delta > 0$;*
- *R and $\boldsymbol{Z}$ are independent.*

Then $\boldsymbol{X} = (X_1, X_2) = R\boldsymbol{Z}$ is regularly varying.

Proof: For simplicity, we assume that $Z_1 \stackrel{d}{=} Z_2$, and $Z_1, Z_2 \geq 0, R \geq 0$. To show that $\boldsymbol{X}$ is regularly varying we need to show that there exist a sequence c_n and a measure $\boldsymbol{\nu}_{\boldsymbol{X}}$ such that

$$n\mathbb{P}(c_n^{-1}(X_1, X_2) \in \cdot) \xrightarrow{\mathrm{v}} \boldsymbol{\nu}_{\boldsymbol{X}}(\cdot) \ .$$

Equivalently, since we assumed that the marginals are the same,

$$\frac{\mathbb{P}(x^{-1}(X_1, X_2) \in \cdot)}{\mathbb{P}(X_1 > x)} \xrightarrow{\mathrm{v}} \boldsymbol{\nu}^*_{\boldsymbol{X}}(\cdot)$$

for a measure $\boldsymbol{\nu}^*_{\boldsymbol{X}}(\cdot)$. The vague convergence is considered on $([0,\infty] \times [0,\infty]) \setminus \{\mathbf{0}\}$, hence it suffices to show that for $y_1, y_2 > 0$,

$$\frac{\mathbb{P}(x^{-1}(X_1, X_2) \in ([0, y_1] \times [0, y_2])^c)}{\mathbb{P}(X_1 > x)} \to \boldsymbol{\nu}^*_{\boldsymbol{X}}(([0, y_1] \times [0, y_2])^c) \ .$$

We have

$$\begin{aligned} &\frac{\mathbb{P}(x^{-1}(X_1, X_2) \in ([0, y_1] \times [0, y_2])^c)}{\mathbb{P}(X_1 > x)} \\ &= \frac{\mathbb{P}(X_1 > xy_1)}{\mathbb{P}(X_1 > x)} + \frac{\mathbb{P}(X_2 > xy_2)}{\mathbb{P}(X_1 > x)} - \frac{\mathbb{P}(X_1 > xy_1, X_2 > xy_2)}{\mathbb{P}(X_1 > x)} \ . \qquad (*) \end{aligned}$$

First, by Breiman's Lemma (see Theorem 2.1.11), since $\mathbb{E}Z_1^{\alpha+\delta} < \infty$, we have

$$\frac{\mathbb{P}(Z_1 R > xy_1)}{\mathbb{P}(Z_1 R > x)} \to y_1^{-\alpha} \ .$$

Similarly,

$$\frac{\mathbb{P}(Z_2 R > xy_2)}{\mathbb{P}(Z_2 R > x)} \to y_2^{-\alpha} \ ,$$

as $x \to \infty$. For the third term in (*) we have

$$\begin{aligned} \frac{\mathbb{P}(X_1 > xy_1, X_2 > xy_2)}{\mathbb{P}(X_1 > x)} &= \frac{\mathbb{E}[\mathbb{P}(R > \frac{xy_1}{Z_1}, R > \frac{xy_2}{Z_2} \mid Z_1, Z_2)]}{\mathbb{P}(X_1 > x)} \\ &= \frac{\mathbb{E}[\mathbb{P}(R > x(\frac{y_1}{Z_1} \vee \frac{y_2}{Z_2}) \mid Z_1, Z_2)]}{\mathbb{P}(X_1 > x)} = \mathbb{E}\left[\frac{\mathbb{P}(R > x(\frac{y_1}{Z_1} \vee \frac{y_2}{Z_2}) \mid Z_1, Z_2)}{\mathbb{P}(R > x)}\right] \frac{\mathbb{P}(R > x)}{\mathbb{P}(Z_1 R > x)} \\ &\to \mathbb{E}\left[\left(\frac{y_1}{Z_1} \vee \frac{y_2}{Z_2}\right)^{-\alpha}\right] \frac{1}{\mathbb{E}(Z_1^\alpha)} \ . \end{aligned}$$

We can exchange the limit with integration by Potter's bounds (see Proposition 2.1.5). Hence,

$$\boldsymbol{\nu}^*_{\boldsymbol{X}}([0, y_1] \times [0, y_2])^c) = y_1^{-\alpha} + y_2^{-\alpha} + \mathbb{E}\left[\left(\frac{y_1}{Z_1} \vee \frac{y_2}{Z_2}\right)^{-\alpha}\right] \frac{1}{\mathbb{E}(Z_1^\alpha)} \ .$$

∎

We apply Lemma 2.2.10 to the bivariate t distribution.

Example 2.2.11 Bivariate t. A two dimensional random vector $\boldsymbol{X} = (X_1, X_2)$ is said to have the bivariate t distribution with ν degrees of freedom, mean vector $\boldsymbol{\mu}$ and with $\boldsymbol{\Sigma}$ denoting the covariance matrix if $\boldsymbol{X}$ has the stochastic representation

$$\boldsymbol{X} = \boldsymbol{\mu} + \frac{\nu}{\sqrt{S}}\boldsymbol{Z} \ ,$$

where $S \sim \chi^2_\nu$ and $\boldsymbol{Z} \sim N(0, \boldsymbol{\Sigma})$ are independent. $\boldsymbol{X}$ is regularly varying with index ν. Indeed, $1/\sqrt{S}$ is regularly varying with index $-\nu$ which can be clearly seen for the case of $\nu = 1$ (see Lemma 2.1.13). Finally, the regular variation of the vector follows from Lemma 2.2.10.

2.3 Multivariate regular varying time series

In this section we introduce the concept of regular variation for time series. As in the multivariate case, we link regular variation to quasi-spectral decomposition following the lines of [3]. Several examples (AR(1), solutions to Stochastic Recurrence Equations, threshold ARCH) are given.

Assume that $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ is a strictly stationary regularly varying time series with values in $\mathbb{R}^d$. This means that all the finite-dimensional distributions are (multivariate) regularly varying with a positive tail index α; that is for all $h \geq 0$,

$$\frac{\mathbb{P}(x^{-1}(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_h) \in \cdot)}{\mathbb{P}(\|\boldsymbol{X}_0\| > x)} \stackrel{\text{v}}{\to} \boldsymbol{\nu_{0,h}}(\cdot) \tag{2.3.1}$$

for some non-zero Radon measure $\boldsymbol{\nu_{0,h}}$, where $\stackrel{\text{v}}{\to}$ denotes the vague convergence in $\bar{\mathbb{R}}^{d(h+1)} \setminus \{\boldsymbol{0}\}$, $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ and $\|\cdot\|$ is a vector norm on $\mathbb{R}^d$.

Tail process and spectral tail process. Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Since $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ is regularly varying, there exist processes $\{\boldsymbol{V}_j, j \in \mathbb{N}\}$ and $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ such that for each $h \geq 0$,

$$x^{-1}(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_h) \ ,$$

conditionally on $\{\|\boldsymbol{X}_0\| > x\}$, converges in distribution to $\boldsymbol{V}_0, \ldots, \boldsymbol{V}_h$, and

$$\frac{1}{\|\boldsymbol{X}_0\|}(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_h) ,$$

conditionally on $\{\|\boldsymbol{X}_0\| > x\}$, converges in distribution to $(\boldsymbol{\Theta}_0, \ldots, \boldsymbol{\Theta}_h)$. In particular, $\boldsymbol{V}_h = \|\boldsymbol{V}_0\|\boldsymbol{\Theta}_h$ for all $h \geq 1$, $\boldsymbol{V}_0$ and $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ are independent, while $\mathbb{P}(\|\boldsymbol{V}_0\| > y) = y^{-\alpha}$, $y \geq 1$. See [3]. We note that formally the limiting processes $\{\boldsymbol{V}_j, j \in \mathbb{N}\}$ and $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ depend on the choice of the norm $\|\cdot\|$. We call both processes the *tail* and the *spectral tail* processes, respectively. We note that the term *spectral* is not to be understood in the usual sense, since the random elements $\boldsymbol{\Theta}_j$ are not concentrated on the unit sphere in $\mathbb{R}^d$.

2.3.1 Examples

In this section we provide several examples of regularly varying time series. We give (without a proof) conditions for existence of the stationary solution as well as regular variation. The formulas for the tail and the spectral tail processes are given.

Example 2.3.1 Consider AR(1) process $X_j = \rho X_{j-1} + \varepsilon_j$, where, for simplicity, $\rho \in (0, 1)$ and $\{\varepsilon_j, j \in \mathbb{Z}\}$ are i.i.d., nonnegative and regularly varying random variables. The unique stationary solution exists and is given by

$$X_j = \sum_{k=0}^{\infty} \rho^k \varepsilon_{j-k} , \quad j \geq 0 .$$

The stationary solution is also multivariate regularly varying and the tail index of X_0 agrees with the tail index of ε_0. In particular, as $x \to \infty$,

$$\mathbb{P}(X_0 > x) \sim \frac{1}{1-\rho^\alpha}\mathbb{P}(\varepsilon_0 > x) ,$$

cf. [18, Lemma A3.26]. Furthermore, we have

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} = \rho^\alpha , \tag{2.3.2}$$

that is, the tail dependence coefficient equals ρ^α. The above formula means also that the tail process has the form $V_1 = V_0\rho^\alpha$, where V_0 is Pareto-distributed with index α. This also means that Θ_1 is degenerated, $\Theta_1 = \rho^\alpha$. In general, $\Theta_j = \rho^{j\alpha}$.

We verify (2.3.2):

$$\begin{aligned}
&\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} \\
&= \lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, \rho X_0 + \varepsilon_1 > x)}{\mathbb{P}(X_0 > x)} \\
&> \lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, \rho X_0 > x)}{\mathbb{P}(X_0 > x)} \\
&= \lim_{x\to\infty} \frac{\mathbb{P}(\rho X_0 > x)}{\mathbb{P}(X_0 > x)} = \rho^\alpha\,.
\end{aligned}$$

On the other hand, for $\delta \in (0, 1)$

$$\begin{aligned}
&\frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} \\
&= \frac{\mathbb{P}(X_0 > x, \rho X_0 + \varepsilon_1 > x, \varepsilon_1 > \delta x)}{\mathbb{P}(X_0 > x)} + \frac{\mathbb{P}(X_0 > x, \rho X_0 + \varepsilon_1 > x, \varepsilon_1 \le \delta x)}{\mathbb{P}(X_0 > x)} \\
&\le \frac{\mathbb{P}(X_0 > x)\mathbb{P}(\varepsilon_1 > \delta x)}{\mathbb{P}(X_0 > x)} + \frac{\mathbb{P}(X_0 > x, \rho X_0 + \delta x > x)}{\mathbb{P}(X_0 > x)} \\
&= \mathbb{P}(\varepsilon_1 > \delta x) + \frac{\mathbb{P}(X_0 > x, X_0 > \frac{x(1-\delta)}{\rho})}{\mathbb{P}(X_0 > x)}\,.
\end{aligned}$$

Choose δ so small so that $\frac{(1-\delta)}{\rho} > 1$. Letting $x \to \infty$, we get

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} \le \left(\frac{1-\delta}{\rho}\right)^{-\alpha}\,.$$

Now, letting $\delta \to 0$, we have

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} \le \rho^\alpha\,.$$

Therefore, (2.3.2) holds. Similarly, we have

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > xy_0, X_1 > xy_1)}{\mathbb{P}(X_0 > x)} = \left(y_0 \vee y_1\rho^{-1}\right)^{-\alpha}\,.$$

Example 2.3.2 Assume that $\{X_j\}$ is an AR(p) model

$$X_j = \phi_1 X_{j-1} + \cdots + \phi_p X_{j-p} + \varepsilon_j \ , \quad j \geq 0 \ ,$$

that satisfies the following conditions:

- the innovations $\{\varepsilon_j\}$ are i.i.d. and regularly varying with index α;
- the spectral radius of the matrix (the supremum among the absolute values of the elements in its spectrum)

$$\Sigma = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} ,$$

  is smaller than 1.
- if $\alpha \leq 2$, then $\sum_{i=1}^{p} |\phi_i|^q < 1$ for $q = \min\{1, \alpha\}$.

The AR(p) process can be embedded into an $\mathbb{R}^p$-valued vector-autoregressive Markov chain

$$\mathbb{Y}_j = \Sigma \mathbb{Y}_{j-1} + Z_j \tag{2.3.3}$$

with

$$\mathbb{Y}_j = (X_j, \ldots, X_{j-p+1})^T \ , \quad Z_j = (\varepsilon_j, 0, \ldots, 0)^T \ .$$

Since the spectral radius of Σ is smaller than 1, the stationary solution to (2.3.3) exists and is given by $\mathbb{Y}_j = \sum_{k=0}^{\infty} \Sigma^k \mathbb{Y}_{j-k}$. Since the innovation sequence $\{\varepsilon_j\}$ is regularly varying, the chain $\{\mathbb{Y}_j\}$ is also regularly varying with index α. The AR(p) process is recovered by taking $X_j = g(\mathbb{Y}_j)$ with $g(\boldsymbol{y}) = g(y_0, \ldots, y_p) = y_0$ and $\{X_j\}$ is also regularly varying. We refer to [14] for more details.

Example 2.3.3 Assume that $\{(A_j, B_j), j \in \mathbb{Z}\}$, are i.i.d. random vectors of nonnegative random variables and define

$$X_{j+1} = A_{j+1}X_j + B_{j+1} .$$

The stationary solution exists whenever $-\infty \le \mathbb{E}[\log A_0] < 0$ and $\mathbb{E}[\log^+ B_0] < \infty$; see [2]. If there exists $\alpha > 0$ such that $\mathbb{E}[A_0^\alpha] = 1$, $\mathbb{E}[A_0^\alpha \log^+ A_0] < \infty$ and $\mathbb{E}[B_0^\alpha] < \infty$, then the marginal distribution of the stationary solution is regularly varying by [21] and [19]. Moreover, the sequence $\{X_j\}$ is also regularly varying; see [2]. Furthermore,

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_0 > x, X_1 > x)}{\mathbb{P}(X_0 > x)} = \mathbb{E}[(A_0 \wedge 1)^\alpha] .$$

This defines the tail dependence coefficient between X_0 and X_1. We also have the formula for the spectral process:

$$\Theta_j = \prod_{i=1}^{j} A_i .$$

Example 2.3.4 Let $\xi \in \mathbb{R}$. Assume that $\{X_j\}$ is T-ARCH model

$$X_j = (b_{10} + b_{11}X_{j-1}^2)^{1/2}\varepsilon_j \mathbb{1}_{\{X_{j-1}<\xi\}} + (b_{20} + b_{21}X_{j-1}^2)^{1/2}\varepsilon_j \mathbb{1}_{\{X_{j-1}\ge\xi\}} , \qquad (2.3.4)$$

that satisfies the following conditions:

- $b_{ij} > 0$;
- the innovations $\{\varepsilon_j, j \in \mathbb{Z}\}$ are i.i.d. such that $\mathbb{E}[|\varepsilon_0|^\beta] < \infty$ for all $\beta > 0$;
- the innovations have a density f_Z not vanishing in a neighbourhood of zero and bounded;
- the Lyapunov exponent

  $$\gamma = p \log b_{11}^{1/2} + (1-p) \log b_{21}^{1/2} + \mathbb{E}[\log(|\varepsilon_0|)] ,$$

  where $p = \mathbb{P}(\varepsilon_0 < 0)$, is strictly negative;

- $(b_{11} \vee b_{21})^{q/2}\mathbb{E}[|\varepsilon_0|^q] < 1$.

Under the stated conditions, the Markov chain $\{X_j\}$ has a stationary solution, see [9, Theorem 2.2]. Moreover, the stationary distribution is regularly varying and the index of regular variation of X_0 is obtained by solving

$$b_{11}^{\alpha/2}\mathbb{E}[|\varepsilon_0|^\alpha \mathbb{1}_{\varepsilon<0}] + b_{21}^{\alpha/2}\mathbb{E}[|\varepsilon_0|^\alpha \mathbb{1}_{\varepsilon_0\geq 0}] = 1 \ ;$$

see again [9].

2.4 Conditional Extreme Value (CEV) model

Consider the case of $d = 2$ and assume for simplicity that all random variables are nonnegative. When the limiting measure in the definition of the multivariate regular variation (Definition 2.2.1) is concentrated on the axes, we have extremal independence and no information about joint behaviour of extremes can be inferred from $\boldsymbol{\nu}_{\boldsymbol{X}}$, where $\boldsymbol{X} = (X_1, X_2)$. In order to deal with this, we consider the so-called *Conditional Extreme Value* assumption, studied rigorously by [20], [10] and [23]. Since we consider the bivariate case only, we will write (X, Y) instead of (X_1, X_2).

Assumption 1 *Assume that X and Y are nonnegative random variables. There exists a scaling function b and Radon measure $\boldsymbol{\mu}$ on $(0,\infty] \times [0, +\infty]$ such that*

$$\frac{1}{\mathbb{P}(X > x)}\mathbb{P}\left(\left(\frac{X}{x}, \frac{Y}{b(x)}\right) \in \cdot\right) \stackrel{\text{v}}{\to} \boldsymbol{\mu} \ , \quad \text{as } u \to \infty \ , \tag{2.4.1}$$

on $(0,\infty] \times [0, +\infty]$ and for all $y_0 > 0$,

a. the measure $\boldsymbol{\mu}([y_0, \infty] \times \cdot)$ on $\mathbb{R}_+$ has no mass at $+\infty$;

b. the measure $\boldsymbol{\mu}([y_0, \infty] \times \cdot)$ on $\mathbb{R}_+$ is not concentrated at a point;

c. the measure $\boldsymbol{\mu}(\cdot \times \mathbb{R}_+)$ on $(0, \infty]$ has no mass at $+\infty$.

We say that (X, Y) *is CEV with the function* b.

The most important consequence of Assumption 1, is that the function b is regularly varying with index $\phi \in \mathbb{R}$ (see [20, Proposition 1] and [23].) To put emphasis on the regular variation of the function b, we call ϕ the **conditional scaling exponent** (see [23]). Furthermore, Assumption 1 implies that the variable X is regularly varying (with index α). On the other hand, there is no distributional assumption on Y.

These properties are the key features that we will use for the statistical inference. Let us collect several further comments on Assumption 1; see [20], [10] and [23].

- The conditions a-c are the *non-degeneracy* assumptions. For example, part c applies to the first coordinate only and can be recognized as the standard condition for the stated regular variation of X;

- If the random variables X and Y are independent, X is regularly varying with index α and Y has distribution F_Y, then $b(x) = 1$ and the limiting measure is the product measure, $\boldsymbol{\mu} = \nu_\alpha \times F_Y$, where $\nu_\alpha(\mathrm{d}x) = \alpha x^{-\alpha-1}\mathrm{d}x$, $x > 0$. It is possible that the limiting measure is the product measure even if X and Y are dependent, see Example 2.4.7;

- There is no general relationship between the multivariate regular variation (as stated in Definition 2.2.1) and Assumption 1. In particular, the multivariate regular variation with extremal independence (that is, when the measure $\boldsymbol{\nu_X}$ is concentrated on the axes) does not imply that Assumption 1 holds, see Example 2.4.9;

- However, if the vector (X, Y) is multivariate regularly varying in the sense of the Definition 2.2.1 and Assumption 1 holds, then:

  - For the scaling function we have $b(x) = x$ in case of extremal dependence. In this case the measures $\boldsymbol{\nu_X}$ and $\boldsymbol{\mu}$ coincide on $(0, \infty] \times [0, \infty]$;

– For the scaling function we have $b(x) = o(x)$ in case of extremal independence.

The following lemma can be thought as the extension of Lemma 2.2.4 and was proven in [23].

Lemma 2.4.1 *Let Assumption 1 hold. Let $g : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function. Assume moreover that there exists $\delta > 0$ such that*

$$\sup_{x \in \mathbb{R}_+} \mathbb{E}\left[\left|g^{1+\delta}\left(\frac{Y}{b(x)}\right)\right| \mid X > x\right] < \infty \,. \tag{2.4.2}$$

Then

$$\lim_{x\to\infty} \mathbb{E}\left[g\left(\frac{Y}{b(x)}\right) \mid X > x\right] = \int_1^\infty \int_0^\infty g(v)\boldsymbol{\mu}(\mathrm{d}u, \mathrm{d}v) \,. \tag{2.4.3}$$

The next lemma shows how CEV is preserved under marginal transformation of X. The result is original.

Lemma 2.4.2 *Let (X, Y) be CEV with function b. Let f be a strictly increasing continuous function such that $f^{\leftarrow}$ is regularly varying with index $1/\gamma$, $\gamma > 0$. Define $\widetilde{X} = f(X)$. Then $(\widetilde{X}, Y)$ is CEV with $\widetilde{b} = b \circ f^{\leftarrow}$.*

Proof: It suffices to prove that for $y_1, y_2 > 0$, the limit

$$\lim_{x\to\infty} \mathbb{P}(Y < \widetilde{b}(x)y_2, \widetilde{X} > xy_1 \mid \widetilde{X} > x)$$

exists and is finite. We have by setting $v = f^{\leftarrow}(x)$

$$\begin{aligned}
\mathbb{P}(Y < \widetilde{b}(x)y_2, \widetilde{X} > y_1x \mid \widetilde{X} > x) &= \frac{\mathbb{P}(Y < \widetilde{b}(x)y_2, \widetilde{X} > y_1x, \widetilde{X} > x)}{\mathbb{P}(\widetilde{X} > x)} \\
= \frac{\mathbb{P}(Y < b(f^{\leftarrow}(x))y_2, \widetilde{X} > x(y_1 \vee 1))}{\mathbb{P}(\widetilde{X} > x)} &= \frac{\mathbb{P}(Y < b(f^{\leftarrow}(x))y_2, X > f^{\leftarrow}(x(y_1 \vee 1)))}{\mathbb{P}(X > f^{\leftarrow}(x))} \\
= \frac{\mathbb{P}\left(Y < b(v)y_2, X > f^{\leftarrow}(x)\frac{f^{\leftarrow}(x(y_1\vee 1))}{f^{\leftarrow}(x)}\right)}{\mathbb{P}(X > v)} &= \frac{\mathbb{P}\left(Y < b(v)y_2, X > v\frac{f^{\leftarrow}(x(y_1\vee 1))}{f^{\leftarrow}(x)}\right)}{\mathbb{P}(X > v)} \,.
\end{aligned}$$

If $x \to \infty$, then $v = f^{\leftarrow}(x) \to \infty$. Since $f^{\leftarrow}$ is regularly varying with index γ, we know that $\frac{f^{\leftarrow}(x(y_1 \vee 1))}{f^{\leftarrow}(x)}$ converges to $c(y_1) = (y_1 \vee 1)^{1/\gamma}$. Then

$$\begin{aligned}
&\lim_{x\to\infty} \mathbb{P}(Y < \widetilde{b}(x)y_2, \widetilde{X} > y_1 x \mid \widetilde{X} > x) \\
&= \lim_{v\to\infty} \frac{\mathbb{P}(Y < b(v)y_2, X > vc(y_1))}{\mathbb{P}(X > v)} = \boldsymbol{\mu}((c(y_1), \infty), (0, y_2)) \ .
\end{aligned}$$

∎

2.4.1 Marginal Expected Shortfall

Assume again that X, Y are nonnegative. In Section 2.2.3 we consider the Marginal Expected Shortfall (cf. (2.2.10))

$$\theta(p) := \mathbb{E}[Y \mid X > Q_X(p)] \ . \tag{2.4.4}$$

We concluded that it vanishes under extremal independence. The CEV assumption allows us to resolve this issue. Instead of considering the limit (2.2.11), that is

$$\lim_{p\to 0} \mathbb{E}\left[\frac{Y}{Q_X(p)} \mid X > Q_X(p)\right] = \int_1^\infty \int_0^\infty v\boldsymbol{\nu}(\mathrm{d}v, \mathrm{d}u) \ ,$$

we will study

$$\lim_{p\to 0} \mathbb{E}\left[\frac{Y}{b(Q_X(p))} \mid X > Q_X(p)\right] \ ,$$

where $Q_X(p) = F_X^{\leftarrow}(1-p)$. In order to do this, we have to justify the existence and finiteness of the limit

$$\aleph_{\mathrm{CTE}}(q) := \lim_{p\to 0} \mathbb{E}\left[\frac{Y^q}{b^q(Q_X(p))} \mid X > Q_X(p)\right] \tag{2.4.5}$$

for some $q > 0$.

The following result is original contribution from the author.

Lemma 2.4.3 *Let Assumption 1 hold. Let $q > 0$. Assume that one of the following conditions is satisfied:*

a) There exists $\delta > 0$ such that

$$\limsup_{x\to\infty} \mathbb{E}\left[\left(\frac{Y}{b(x)}\right)^{q+\delta} \mid X > x\right] < \infty \ ;$$

b) The limiting measure $\boldsymbol{\mu}$ is the product measure and $\mathbb{E}[Y^q] < \infty$.

Then for $s \geq s_0 > 0$,

$$\lim_{x\to\infty} \frac{1}{\bar{F}_X(x)} \mathbb{E}\left(\frac{Y^q}{b^q(x)} \mathbb{1}_{\{X>xs\}}\right) = s^{q\phi-\alpha} \int_1^\infty \int_0^\infty v^q \boldsymbol{\mu}(\mathrm{d}u, \mathrm{d}v) \in (0,\infty) \tag{2.4.6}$$

and hence

$$\aleph_{\mathrm{CTE}}(q) = \int_1^\infty \int_0^\infty v^q \boldsymbol{\mu}(\mathrm{d}u, \mathrm{d}v) \ . \tag{2.4.7}$$

Conversely, if (2.4.6) holds then $q\phi < \alpha$.

As noted above, if (X, Y) is regularly varying with extremal dependence, then Assumption 1 is automatically fulfilled with $b(x) = x$. Hence, recalling that in this case the measures $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ coincide, we have

Corollary 2.4.4 *Assume that (X, Y) is regularly varying with the measure $\boldsymbol{\nu}$ that is not concentrated on the axes. If*

$$\limsup_{x\to\infty} \mathbb{E}\left[\left(\frac{Y}{x}\right)^{q+\delta} \mid X > x\right] < \infty \ ,$$

then for $s \geq s_0$

$$\lim_{x\to\infty} \frac{1}{\bar{F}_X(x)} \mathbb{E}\left(\frac{Y^q}{x^q} \mathbb{1}_{\{X>xs\}}\right) = s^{q-\alpha} \int_1^\infty \int_0^\infty v^q \boldsymbol{\nu}(\mathrm{d}u, \mathrm{d}v) \in (0,\infty) \ . \tag{2.4.8}$$

The existence of the limit in (2.4.7) is a key to estimation of $\theta(p)$ defined in (2.2.10). We conjecture that $q\phi < \alpha$ suffices for the finiteness of (2.4.6).

Proof: (**Proof of Lemma 2.4.3**)

We start by noting that since the random variables X and Y are nonnegative, the limit on the right hand side of (2.4.6) would vanish only when the limiting measure $\boldsymbol{\mu}$ is concentrated on the horizontal axis. This is prevented, however, by Assumption 1b).

Now, we proceed with the proof of finiteness. For a) the result follows from [23], while it is obvious under the assumption b). As for the converse, for $A > s_0$ set

$$I_A(q) = \int_0^\infty \int_A^\infty y^q \boldsymbol{\mu}(\mathrm{d}u, \mathrm{d}v) \; .$$

This quantity is finite by the assumption and its finiteness implies $\lim_{A\to\infty} I_A(q) = 0$. On the other hand,

$$\begin{aligned} I_A(q) &= \lim_{x\to\infty} \frac{1}{\bar{F}_X(x)} \mathbb{E}\left[\frac{Y^q}{b^q(x)} \mathbb{1}_{\{X > xA\}}\right] \\ &= \lim_{x\to\infty} \frac{\bar{F}_X(xA)}{\bar{F}_X(x)} \frac{b^q(xA)}{b^q(x)} \frac{1}{\bar{F}_X(xA)} \mathbb{E}\left[\frac{Y^q}{b^q(xA)} \mathbb{1}_{\{X > xA\}}\right] \\ &= \lim_{x\to\infty} \frac{\bar{F}_X(xA)}{\bar{F}_X(x)} \frac{b^q(xA)}{b^q(x)} \mathbb{E}\left[\frac{Y^q}{b^q(xA)} \mid X > xA\right] = A^{q\phi-\alpha} \aleph_{\mathrm{CTE}}(q) \; . \end{aligned} \tag{2.4.9}$$

Hence, $\lim_{A\to\infty} I_A(q) = 0$ if and only if $\alpha > q\phi$.

Finally, replacing A with s in (2.4.9) yields the right hand side in (2.4.6). ∎

2.4.2 Examples

Example 2.4.5 This example can be treated as a generic model for CEV. Assume that X, V are independent regularly varying random variables with index $\alpha > 0$. For $\phi \in (0, 1)$ define $Y = X^\phi V$. By Breiman's lemma (see e.g. Theorem 2.1.11), Y is also regularly varying with index α and tail equivalent to X. The pair (X, Y) is bivariate regularly varying with index $-\alpha$ and the exponent measure concentrated on both axes and hence extremally independent. Also, the tail dependence coefficient vanishes.

Since the sets $(s,\infty]\times[0,t]$ characterize the measure in $(0,\infty]\times[0,\infty]$, to verify that (X,Y) is CEV with the function $b(x)=x^\phi$, it suffices to calculate

$$\lim_{x\to\infty}\mathbb{P}(X>xs,Y\le x^\phi t)/\mathbb{P}(X>x)\ .$$

Regular variation of X entails

$$\lim_{x\to\infty}\frac{\mathbb{P}(X>xs,Y\le x^\phi t)}{\mathbb{P}(X>x)}=\lim_{x\to\infty}\left\{\frac{\mathbb{P}(X>xs)}{\mathbb{P}(X>x)}-\frac{\mathbb{P}(Y>x^\phi t,X>xs)}{\mathbb{P}(X>x)}\right\}$$
$$=s^{-\alpha}-\lim_{x\to\infty}\frac{\mathbb{P}(Y>x^\phi t,X>xs)}{\mathbb{P}(X>x)}\ .$$

Now, for the latter let $v=xs$, then

$$\lim_{x\to\infty}\frac{\mathbb{P}(Y>x^\phi t,X>xs)}{\mathbb{P}(X>x)}=\lim_{v\to\infty}\frac{\mathbb{P}(Y>v^\phi t/s^\phi,X>v)}{\mathbb{P}(X>v)}\lim_{x\to\infty}\frac{\mathbb{P}(X>xs)}{\mathbb{P}(X>x)}$$
$$=s^{-\alpha}\lim_{v\to\infty}\mathbb{P}(Y>v^\phi t/s^\phi\mid X>v)=s^{-\alpha}\int_1^\infty\mathbb{P}(V>z^{-\phi}t/s^\phi)\alpha z^{-\alpha-1}\mathrm{d}z\ .$$

The last equality above follows from calculation of the limiting conditional probabilities. Let F_X be the distribution function of X. Then

$$\lim_{v\to\infty}\mathbb{P}(Y>v^\phi t\mid X>v)=\lim_{v\to\infty}\frac{\mathbb{P}(Y>v^\phi t,X>v)}{\mathbb{P}(X>v)}$$
$$=\lim_{v\to\infty}\frac{1}{\mathbb{P}(X>v)}\mathbb{P}(X^\phi V>v^\phi t,X>v)=\lim_{v\to\infty}\frac{1}{\mathbb{P}(X>v)}\int_v^\infty\mathbb{P}(x^\phi V>v^\phi t)F_X(dx)$$
$$=\lim_{v\to\infty}\frac{1}{\bar{F}_X(v)}\int_1^\infty\mathbb{P}(V>z^{-\phi}t)F_X(v\mathrm{d}z)=\int_1^\infty\mathbb{P}(V>z^{-\phi}t)\alpha z^{-\alpha-1}\mathrm{d}z\ .$$

The above integral is finite since $\int_1^\infty\alpha z^{-\alpha-1}\mathrm{d}z=1$. We can interchange the limit with integration by dominated convergence theorem by noting that the measure $F_X(v\cdot)/\bar{F}_X(v)$ converges weakly on $[1,\infty)$ to the measure $\nu((z,\infty))=z^{-\alpha}$.

In summary, CEV is fulfilled with $b(x)=x^\phi$ and

$$\boldsymbol{\mu}((s,\infty]\times[0,t])=s^{-\alpha}\left(1-\int_1^\infty\mathbb{P}(V>z^{-\phi}t/s^\phi)\alpha z^{-\alpha-1}\mathrm{d}z\right)\ .$$

Example 2.4.6 As in [20, Section 6.2.1], let (U, V) be a bivariate regularly varying random vector with the exponent measure $\boldsymbol{\nu}$ that is not concentrated on the axes. Hence both U and V are regularly varying with index α and tail equivalent. As such, we can write

$$\frac{\mathbb{P}\left(u^{-1}(U, V) \in \cdot\right)}{\mathbb{P}(U > u)} \xrightarrow{\text{v}} \boldsymbol{\nu}$$

on $[0, \infty]^2 \setminus \{\mathbf{0}\}$, as $u \to \infty$. Let (U_i, V_i), $i = 1, 2$, be independent copies from (U, V). For $\phi, \theta \in (0, 1)$ define

$$(X, Y) = B(U_1, V_1^{\phi}) + (1 - B)(U_2^{\theta}, V_2) \ ,$$

where $\mathbb{P}(B = 0) = \mathbb{P}(B = 1) = 1/2$ and B is independent of U_i, V_i, $i = 1, 2$. Both X and Y are regularly varying with index α due to Breiman lemma and in fact $\mathbb{P}(X > x) \sim 0.5\mathbb{P}(U > x)$ as $x \to \infty$. Furthermore, (X, Y) is bivariate regularly varying, with extremal independence (even though (U, V) may have extremal dependence.) This implies that the tail dependence coefficient vanishes. CEV is satisfied and the scaling function is given by $b(x) = x^{\phi}$. Indeed, by repeating the argument in [20],

$$\begin{aligned}
&\lim_{x\to\infty} \frac{\mathbb{P}(X > xs, Y \le x^{\phi}t)}{\mathbb{P}(X > x)} = \lim_{x\to\infty} \left\{ \frac{\mathbb{P}(X > xs)}{\mathbb{P}(X > x)} - \frac{\mathbb{P}(Y > x^{\phi}t, X > xs)}{\mathbb{P}(X > x)} \right\} \\
&= s^{-\alpha} - \lim_{x\to\infty} \frac{\mathbb{P}(Y > x^{\phi}t, X > xs)}{\mathbb{P}(X > x)} = s^{-\alpha}\left\{1 - \boldsymbol{\nu}((1, \infty] \times (t^{1/\phi}/s, \infty])\right\} \\
&=: \boldsymbol{\mu}((s, \infty] \times [0, t]) \ .
\end{aligned}$$

In fact, only the part $B(U_1, V_1^{\phi})$ contributes to the limit. Indeed:

$$\begin{aligned}
&\frac{\mathbb{P}(Y > x^{\phi}t, X > xs)}{\mathbb{P}(X > x)} = 0.5\frac{\mathbb{P}(V_1^{\phi} > x^{\phi}t, U_1 > xs)}{\mathbb{P}(X > x)} + 0.5\frac{\mathbb{P}(V_2 > x^{\phi}t, U_2^{\theta} > xs)}{\mathbb{P}(X > x)} \\
&\sim \frac{\mathbb{P}(V > (xs)t^{1/\phi}/s, U > xs)}{\mathbb{P}(U > xs)} \frac{\mathbb{P}(U > xs)}{\mathbb{P}(U > x)} + O\left(\frac{\mathbb{P}(U^{\theta} > xs)}{\mathbb{P}(U > x)}\right)
\end{aligned}$$

as $x \to \infty$ and clearly the big O term vanishes since $\theta \in (0, 1)$.

Example 2.4.7 We present a modification of Example 7 from [10]. Consider a pair (X, Y) with the following distribution function:

$$F(x, y) = \frac{(1 - x^{-\alpha})(1 - y^{-\alpha})}{1 + (xy)^{-\alpha}} .$$

We note that this example is generated from the Ali-Mikhail-Haq copula with $\theta = -1$. The marginal distributions of X and Y are clearly Pareto with index $-\alpha$. CEV assumption is fulfilled with $b(x) = 1$ and the limiting measure is the product measure

$$\boldsymbol{\mu}((s, \infty] \times [0, t]) = s^{-\alpha}(1 - t^{-\alpha}) .$$

Example 2.4.8 Assume that X is a regularly varying random variable with index $\alpha > 0$. Let $\phi \in (0, 1)$ and assume that U is a nonnegative random variable, independent of X such that $\mathbb{E}[U^{\alpha/\phi+\delta}] < \infty$ for some $\delta > 0$. Define $Y = X^{\phi}U$. By Breiman's lemma, Y is also regularly varying with index α/ϕ. The pair (X, Y) is bivariate regularly varying with index α, with the exponent measure concentrated on the horizontal axes, and hence extremally independent. The main difference between the current situation and Example 2.4.5 is that the extremal independence here is linked to different tail behaviour of X and Y. CEV assumption 1 is satisfied with the scaling function given by $b(x) = x^{\phi}$.

Example 2.4.9 We now give an example, where the conditional laws do not exists. Consider a pair of dependent standard normal random variables ξ_1, ξ_2 and define $X = \mathrm{e}^{c\xi_1^2}$, $Y = \mathrm{e}^{c\xi_2^2}$ with $c < 1/2$. Then X and Y are regularly varying. In this case, extremal independence holds for the bivariate distributions, but a non trivial limiting conditional distribution for Y given $X > x$ does not exist. See [20, section 2.4]. In this case our approach using CEV is not applicable, however, due to extremal independence, the method of [8] cannot be used as well.

2.5 Extremogram

Assume that $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ is a strictly stationary regularly varying time series with values in $\mathbb{R}^d$. In [11] the authors introduced the extremogram as a measure of *extremal dependence* for the stationary regularly varying time series. Assume that $A, B \subseteq \bar{\mathbb{R}}^d$ are sets that are bounded away from zero and such that $\boldsymbol{\nu_{0,0}}(A) > 0$, $\boldsymbol{\nu_{0,h}}(\partial(A \times \bar{\mathbb{R}}^{d(h-1)} \times B)) = 0$. Then the following limit exists:

$$\rho_{AB}(h) = \lim_{x\to\infty} \mathbb{P}(\boldsymbol{X}_h \in xB \mid \boldsymbol{X}_0 \in xA) \in [0,1] \ .$$

We call the sequence $\rho_{AB}(h), h \geq 0$, the *extremogram* of $\{\boldsymbol{X}_j\}$.

Spectral tail process representation of the extremogram. For a norm $\|\cdot\|$ on $\mathbb{R}^d$ take $A = B = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| > 1\}$. Let $\{\boldsymbol{V}_j, j \in \mathbb{N}\}$ and $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ be the corresponding tail and spectral tail processes. Then the extremogram becomes

$$\begin{aligned}
&\lim_{x\to\infty} \mathbb{P}(\boldsymbol{X}_h \in xB \mid \boldsymbol{X}_0 \in xA) = \mathbb{P}(\boldsymbol{V}_h \in B) = \mathbb{P}(\|\boldsymbol{V}_0\|\boldsymbol{\Theta}_h \in B) \\
&= \mathbb{P}(\|\boldsymbol{V}_0\|\|\boldsymbol{\Theta}_h\| > 1) = \mathbb{E}\left[\int_{\{y:y>\|\boldsymbol{\Theta}_h\|^{-1}\}\cap\{y\geq 1\}} \alpha y^{-\alpha-1}\mathrm{d}y\right] = \mathbb{E}\left[\|\boldsymbol{\Theta}_h\|^\alpha \wedge 1\right] \ . \quad (2.5.1)
\end{aligned}$$

Thus, for the specific sets A and B the extremogram can be represented as the expectation of the spectral tail process. The choice of the sets A, B may seem to be restrictive, but suffices in most applications. Also, one can easily extend our considerations to the case of two different norms. For our choice of the sets we will write $\rho(h)$ to denote the corresponding extremogram. We note further that

$$\rho(h) = \mathbb{E}\left[\|\boldsymbol{\Theta}_h\|^\alpha \wedge 1\right] = \lim_{x\to\infty} \mathbb{E}\left[\left(\frac{\|\boldsymbol{X}_h\|}{\|\boldsymbol{X}_0\|} \wedge 1\right)^\alpha \mid \|\boldsymbol{X}_0\| > x\right] \ .$$

Tail Dependence Coefficient. Assume here that $\{X_j, j \in \mathbb{Z}\}$ is a univariate regularly sequence of nonnegative random variables. The marginal distribution function is denoted by F. The tail dependence coefficient at lag h is defined as

$$\tau(h) = \lim_{x\to\infty} \mathbb{P}(X_h > x \mid X_0 > x) \ .$$

Thus, $\tau(h) = \rho_{AB}(h)$ for $A = B = (1, \infty)$.

Chapter 3

Weak convergence

In this chapter we summarize several results on weak convergence of stochastic processes. In Section 3.1 we consider weak convergence in the space $\mathbb{D}$, the space of all real-valued functions that are right-continuous and have left limits, equipped with the Skorokhod J_1-topology. Then, in Section 3.2, we discuss weak convergence using entropy methods. The advantage of the latter is that it is in a sense *dimension-free*. The results here are not new, we base our presentation on, respectively, [5] and [6] (Section 3.1) as well as [31] (Section 3.2). We will use this theory in Chapter 4.

We start with the following classical discussion (see e.g. [6, Section 5]). Let S be an arbitrary metric space and let $\mathcal{S}$ be a class of Borel sets. Recall that a sequence $\{P_n\}$ of probability measures (defined on S) converges weakly to a probability measure P if for all continuous functions $f : S \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int f dP_n = \int f dP \ .$$

We write $P_n \Rightarrow P$. The above definition is not very easy to use, hence alternative criteria for weak convergence are needed. A sequence $\{P_n\}$ of probability measures is *relatively compact* if every subsequence $\{P_{n_i}\}$ contains a further subsequence $\{P_{n_{i(k)}}\}$ such that $P_{n_{i(k)}} \Rightarrow P$ for some probability measure P. We note that there is no guarantee that the limit is the same for all subsequences. However, if finite dimensional distributions of P_n converge weakly to P, then also the entire sequence converges weakly to P. Hence, relative compactness plus convergence of finite dimensional distributions yields weak convergence of probability measures.

The relative compactness is still not the property that is easy to work with. Instead, one can consider *tightness*. A sequence of probability measures $\{P_n\}$ is tight if for every $\epsilon > 0$ there exists a compact set $K \in \mathcal{S}$ such that $P_n(K) > 1 - \epsilon$. Due to *Prokhorov's theorem* (see e.g. [6, Theorem 5.1]), tightness implies relative compactness.

As a consequence, we conclude that tightness along with convergence of finite dimensional distributions yields weak convergence. Therefore, in what follows, we will provide sufficient conditions for tightness in two particular cases of the space S.

3.1 Weak convergence in the space of cadlag functions

Here, $S = \mathbb{D} = \mathbb{D}([0,1])$, the space of real-valued functions on $[0,1]$ that are right-continuous and have left-hand limits and $\mathcal{S} = \mathcal{D}$ is the associated Borel σ-field. For a probability measure $\mathbb{P}$ on $(\mathbb{D}, \mathcal{D})$, let T_P consist of those elements in $\mathbb{D}([0,1])$ for which the projection is continuous except at points forming a set of $\mathbb{P}$-measure 0. Let $\mathbb{X}_n$ and $\mathbb{X}$ be random elements with values in $\mathbb{D}$. Write $T_{\mathbb{X}}$ for $T_{\mathbb{P}}$, where $\mathbb{P}$ is the distribution of $\mathbb{X}$. We have the following theorem, that provides sufficient conditions for weak convergence. In particular, (3.1.1) below yields tightness, which according to the discussion above, implies relative compactness. See Theorem 13.5 in [6].

Theorem 3.1.1 *Suppose that*

$$(\mathbb{X}_n(t_1), \ldots, \mathbb{X}_n(t_k)) \xrightarrow{\mathrm{d}} (\mathbb{X}(t_1), \ldots, \mathbb{X}(t_k))$$

holds whenever $t_1, \ldots, t_k$ *all lie in* $T_{\mathbb{X}}$*; that is* $\mathbb{P}(\mathbb{X}(1) \neq \mathbb{X}(1-)) = 0$*; and that*

$$\mathbb{P}(|\mathbb{X}_n(t) - \mathbb{X}_n(t_1)| \geq \lambda, |\mathbb{X}_n(t_2) - \mathbb{X}_n(t)| \geq \lambda) \leq \frac{1}{\lambda^{2\gamma}}[g(t_2) - g(t_1)]^{2\beta} \tag{3.1.1}$$

for $t_1 \leq t \leq t_2$ *and* $n \geq 1$*, where* $\gamma \geq 0$*,* $\beta > \frac{1}{2}$*, and* g *is a nondecreasing, continuous function on* $[0,1]$*. Then* $\mathbb{X}_n$ *converges weakly to* $\mathbb{X}$ *in* $\mathbb{D}([0,1])$ *equipped with the Skorokhod* J_1*-topology.*

There is a more restrictive version of (3.1.1) involving moments, namely

$$\mathbb{E}\{|\mathbb{X}_n(t) - \mathbb{X}_n(t_1)|^{\gamma}\|\mathbb{X}_n(t_2) - \mathbb{X}_n(t)|^{\gamma}\} \leq [g(t_2) - g(t_1)]^{2\beta}\ . \tag{3.1.2}$$

Theorem 3.1.1 remains valid if equation (3.1.1) is extended as follows;

$$P\{|\mathbb{X}_n(t) - \mathbb{X}_n(t_1)| \geq \lambda, |\mathbb{X}_n(t_2) - \mathbb{X}_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}}[g_n(t_2) - g_n(t_1)]^{2\beta} \tag{3.1.3}$$

where the functions g_n are all monotone increasing or all monotone decreasing, and the sequence (g_n) converges (uniformly on compact intervals) to a monotone continuous function g. Likewise, g can be replaced by g_n in the moment condition (3.1.2).

Comment: If g_n is increasing (respectively, decreasing) for every n and the sequence (g_n) converges pointwise to a continuous function g (which must be increasing or decreasing, respectively), then the convergence is uniform on compact sets. Billingsley in [6] restricts attention to $\mathbb{D}([0,1])$, but the proof of tightness is almost identical for $\mathbb{D}([0,b])$ for $0 < b < \infty$. This is enough for convergence in $\mathbb{D}([0,\infty))$ (see [6], Theorem 16.8).

We provide an argument that g can be replaced with g_n, following closely the set-up of [5]. All references and notations below refer to that book. Note that the goal is to use Theorem 15.4 by showing that (15.11) holds: for every positive ϵ and η, there exists δ, $0 < \delta < 1$ and n_0 such that

$$P(\omega''(\mathbb{X}_n, \delta) \geq \epsilon) \leq \eta, \quad n \geq n_0. \tag{3.1.4}$$

We follow exactly the proof in [6], pp. 128-130, replacing g with g_n in (15.24), (15.26), (15.29) and (15.30), then (15.30) is now

$$P\{\omega''(\mathbb{X}_n, \delta) \geq \epsilon\} \leq \frac{2K}{\epsilon^{2\gamma}}[|g_n(1) - g_n(0)|][\omega_{g_n}(2\delta)]^{2\beta-1},$$

where K depends on ϵ and β, but not on n (see Theorem 12.5). Note that for any $t_1, t_2 \in [0,1]$, $|g_n(t_2) - g_n(t_1)| \leq |g(t_2) - g(t_1)| + 2\sup_{t\in[0,1]} |g_n(t) - g(t)|$. Therefore, letting $\rho_n = 2\sup_{t\in[0,1]} |g_n(t) - g(t)|$, we have

$$\omega_{g_n}(\delta) \leq \omega_g(\delta) + \rho_n, \tag{3.1.5}$$

and

$$P\{\omega''(\mathbb{X}_n, \delta) \geq \epsilon\} \leq \frac{2K}{\epsilon^{2\gamma}}[|g(1) - g(0)| + \rho_n][\omega_g(2\delta) + \rho_n]^{2\beta-1}.$$

Since $\rho_n \to 0$ as $n \to \infty$ (by the aforementioned uniform convergence) and $\omega_g(\delta) \to 0$ as $\delta \to 0$ (by continuity of g), given $\eta > 0$ we can choose n_0 and δ so that the RHS of (3.1.5) is less than η for all $n \geq n_0$. This proves (3.1.4) under the assumption (3.1.3).

3.2 Weak convergence via entropy conditions

In this section we discuss weak convergence of processes using the dimension-free entropy method. The material in this section is based on [31].

Let $\mathcal{X}$ be a measurable space. Let $\mathcal{G}$ be an arbitrary set and $\ell^\infty(\mathcal{G})$ be the set of bounded functions $G : \mathcal{G} \to \mathbb{R}$, equipped with the supremum norm $\|G\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |G(g)|$. A function G on $\ell^\infty(\mathcal{G})$ will be denoted by $G = \{G(g)\}_{g\in\mathcal{G}}$. The supremum norm on $\ell^\infty(\mathcal{G})$ induces a metric ρ_∞ defined as $\rho_\infty(G_1, G_2) = \|G_1 - G_2\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} |G_1(g) - G_2(g)|$. Let $\mathbb{X}_n, n \geq 1$ be a sequence of random elements defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathcal{X} = \ell^\infty(\mathcal{G})$. We denote $\mathbb{X}_n = \{\mathbb{X}_n(g)\}_{g\in\mathcal{G}}$.

Example 3.2.1 Let $\{X_j\}$ be a sequence of standard uniform i.i.d. random variables. Define

$$\mathbb{X}_n(t) = \mathbb{F}_n(t) \ , \quad t \in [0,1] \ ,$$

where $\mathbb{F}_n(t) = \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\{X_j \le t\}}$. Then $\mathbb{X}_n(\cdot)$ is a random element with values in $\ell^\infty(\mathcal{G})$, where $\mathcal{G}$ is the class of bounded functions defined on $[0,1]$. Later on we will use the centered and the scaled version of $\mathbb{X}_n$ defined as follows

$$\mathbb{X}_n(t) = \sqrt{n}\,\{\mathbb{F}_n(t) - t\} \ \ , \quad t \in [0,1] \ . \tag{3.2.1}$$

The weak convergence will be verified by proving finite dimensional convergence and *asymptotic tightness*, which generalizes tightness introduced at the beginning of this chapter (see [31, p.21]).

Definition 3.2.2 *Let $\{\mathbb{X}_n\}$ be a sequence of random maps indexed by a class of functions $\mathcal{G}$, with values in $\ell^\infty(\mathcal{G})$ equipped with the metric ρ_∞ induced by the supremum norm.*

The sequence is asymptotically tight if for each $\epsilon > 0$, there is a compact set $K \subset \ell^\infty(\mathcal{G})$ such that for any $\delta > 0$ we have

$$\limsup_{n\to\infty} \mathbb{P}(\mathbb{X}_n \notin K^\delta) < \epsilon \ ,$$

where $K^\delta = \{G \in \ell^\infty(\mathcal{G}) : \rho_\infty(G,K) < \delta\}$, where $\rho_\infty(G,K) = \inf_{G'\in K}\rho_\infty(G,G')$.

This definition is not very easy to work with. Assume that $\mathcal{G}$ is a compact semimetric space equipped with a metric ρ.

Definition 3.2.3 ([31], p.37) *Let ρ be a semimetric on $\mathcal{G}$. Let $\{\mathbb{X}_n\}$ be a sequence of random maps indexed by a class of functions $\mathcal{G}$, with values in $\ell^\infty(\mathcal{G})$. The sequence is asymptotically uniformly ρ-equicontinuous (in probability) if for every $\epsilon, \eta > 0$ there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{f,g\in\mathcal{G}:\rho(f,g)<\delta} |\mathbb{X}_n(f) - \mathbb{X}_n(g)| > \epsilon \right) < \eta \ .$$

In fact for the purpose of this thesis, these two concepts are equivalent ([31, Theorem 1.5.7]).

Proposition 3.2.4 *A sequence $\{\mathbb{X}_n\}$ is asymptotically tight if and only if $\{\mathbb{X}_n(g)\}$ is tight in $\mathbb{R}$ for every $g \in \mathcal{G}$ and there exists a semimetric ρ on $\mathcal{G}$ such that $(\mathcal{G}, \rho)$ is totally bounded and $\{\mathbb{X}_n\}$ is asymptotically uniformly ρ-equicontinuous (in probability).*

In order to prove tightness, we need to control the size of the class $\mathcal{G}$. This is done using *covering number* or *bracketing number.*

Definition 3.2.5 *Let $(\mathcal{G}, \|\cdot\|)$ be a normed space of real valued functions. Let $\epsilon > 0$. The covering number $N(\epsilon, \mathcal{G}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{G}$. The centers f of the balls need not to belong to $\mathcal{G}$, but it is required that $\|f\| < \infty$. The entropy (without bracketing) is the logarithm of the covering number.*

To define the bracketing number, let $g, h \in \mathcal{G}$, where $(\mathcal{G}, \|\cdot\|)$ is a normed space. The bracket $[g, h]$ is the set of all functions $f \in \mathcal{G}$ such that $g \leq f \leq h$. The functions g and h need not to belong to $\mathcal{G}$, but they have to have finite norms. The ϵ-bracket is a bracket $[g, h]$ such that $\|g - h\| \leq \epsilon$.

Definition 3.2.6 *Let $(\mathcal{G}, \|\cdot\|)$ be a normed space of real valued functions. Let $\epsilon > 0$. The bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the minimal number of ϵ-brackets needed to cover the set $\mathcal{G}$. The entropy with bracketing is the logarithm of the bracketing number.*

Example 3.2.7 Let $\mathcal{G}$ be a class of indicators of the form $\mathbb{1}_{[0,t]}$ or, equivalently, of the form $\mathbb{1}_{(t,1]}$, $t \in [0, 1]$. We equip $\mathcal{G}$ with the L^q-norm, $q \geq 1$:

$$\|f - g\| = \|f - g\|_{L^q} = \left(\int_0^1 |f(x) - g(x)|^q \mathrm{d}x \right)^{1/q} .$$

Let $\epsilon > 0$. If $\psi_t(x) = \mathbb{1}_{[0,t]}(x)$, then

$$\|\psi_t - \psi_{t-\epsilon^q}\|_{L^q} = \|\psi_t - \psi_{t+\epsilon^q}\|_{L^q} = \epsilon \ .$$

Hence, as the centers of the ϵ-balls we can choose the functions ψ_{t_j}, $t_j = j\epsilon^q$, $j = 1, \dots, J$, where J is the smallest integer such that $J\epsilon^q \geq 1$. This implies that the covering number $N(\epsilon, \mathcal{G}, L^q) = N(\epsilon, \mathcal{G}, \|\cdot\|_{L^q})$ is at most $1 + 1/(2\epsilon^q)$.

As for the bracketing number, we consider brackets $[\psi_{t_j}, \psi_{t_j+\epsilon^q}]$, where $t_j = (j-1)\epsilon^q$, $j = 1, \dots, [1/\epsilon^q] + 1$. Then $\|\psi_{t_j} - \psi_{t_j+\epsilon^q}\|_{L^q} = \epsilon$ and each function ψ_s, $s \in [0, 1]$, belongs to one of the brackets. Hence, the bracketing number $N_{[\,]}(\epsilon, \mathcal{G}, L^q)$ is at most $1 + 1/\epsilon^q$. The bracketing and the covering numbers are of the order ϵ^{-q} as $\epsilon \to 0$. □

From the above calculations we obtain of course $N_{[\,]}(2\epsilon, \mathcal{G}, L^q) \leq 1 + 1/(2\epsilon)^q$. Furthermore, if a function ψ_s belongs to the bracket $[\psi_{t_j}, \psi_{t_j+(2\epsilon)^q}]$, then it also belongs to a ball with a center at $\tilde{\psi}_{t_j} = (\psi_{t_j} + \psi_{t_j+(2\epsilon)^q})/2$ and radius ϵ. This yields

$$N(\epsilon, \mathcal{G}, L^q) \leq N_{[\,]}(2\epsilon, \mathcal{G}, L^q) \ . \tag{3.2.2}$$

However, the bound we obtained for the covering number $N(\epsilon, \mathcal{G}, L^q)$ (that is, $1 + 1/2\epsilon^q$) is greater than $1 + 1/(2\epsilon)^q$ unless $q = 1$. Hence, there is some room for improvement for the covering number. This improvement is achieved precisely by considering balls with centers at $\tilde{\psi}_{t_j}$. Note that those functions do not belong to $\mathcal{G}$, however, they still have finite L^q norms. We have

$$\|\psi_{t_j} - \tilde{\psi}_{t_j}\|_{L^q} = \|\psi_{t_j+(2\epsilon)^q} - \tilde{\psi}_{t_j}\|_{L^q} = \epsilon$$

and all functions ψ_s such that $\psi_{t_j} \leq \psi_s \leq \psi_{t_j+(2\epsilon)^q}$ belong to the ball with center at $\tilde{\psi}_{t_j}$ and radius ϵ. Consequently, by choosing centers at $\tilde{\psi}_{t_j}$, $t_j = j(2\epsilon)^q$, $j = 1, \dots, [1/(2\epsilon)^q] + 1$ we cover the whole space $\mathcal{G}$, yielding the covering number $N(\epsilon, \mathcal{G}, L^q)$ to be bounded by $1 + 1/(2\epsilon)^q$, in agreement with (3.2.2).

We extend the above calculations to indicator functions indexed by $t \in [0, 1]^d$.

Example 3.2.8 Let $\mathcal{G}$ be a class of functions $\psi_t(x) = \mathbb{1}_{[0,t]}(x)$, where $t = (t_1, t_2) \in [0,1]^2$ and $x \in \mathbb{R}^2$. We equip $\mathcal{G}$ with the norm

$$\|f - g\|_{L^q} = \left(\int_0^1 \int_0^1 |f(x_1, x_2) - g(x_1, x_2)|^q \mathrm{d}x_1 \mathrm{d}x_2\right)^{1/q} .$$

Then $\psi_t(x) \le \psi_s(x)$ for all x if and only if $t_1 \le s_1$ and $t_2 \le s_2$. Let $0 < \epsilon < 1$. Consider brackets $[\psi_{(t_1,t_2)}, \psi_{(t_1+\epsilon^q/2, t_2+\epsilon^q/2)}]$, where the points $t = (t_1, t_2)$ are chosen as

$$((i-1)\epsilon^q/2, (j-1)\epsilon^q/2), \qquad i, j = 1, \ldots, [2/\epsilon^q] + 1 .$$

For a given point $x \in [0,1]^2$, $\psi_{(t_1,t_2)}(x) - \psi_{(t_1+\epsilon^q/2, t_2+\epsilon^q/2)}(x) = 1$ if and only if x belongs to the union of two sets:

$$\{[t_1, t_1 + \epsilon^q/2] \times [0, t_2 + \epsilon^q/2]\} \bigcup \{[0, t_1 + \epsilon^q/2] \times [t_2, t_2 + \epsilon^q/2]\} ,$$

whose Lebesgue measure is bounded by ϵ^q. Consequently,

$$\|\psi_{(t_1,t_2)} - \psi_{(t_1+\epsilon^q/2, t_2+\epsilon^q/2)}\|_{L^q} \le \epsilon .$$

We need $(1 + 2/\epsilon^q)^2$ brackets to cover the entire set $\mathcal{G}$ and the number is of order ϵ^{-2q} as $\epsilon \to 0$. In the same spirit, the class of indicators $\mathbb{1}_{[0,t]}$, $t \in \mathbb{R}^d$, has the bracketing number at most $1/\epsilon^{qd}$.

The above examples involved indicators of sets $[0,t]$, $t \in [0,1]^d$ and the norm $\|\cdot\|_{L^q}$. We extend the above considerations to the whole real line. Let $\mathbb{P}$ be a probability measure and define

$$\|f - g\|_{L^q(\mathbb{P})} = \left(\int |f(x) - g(x)|^q \mathbb{P}(\mathrm{d}x)\right)^{1/q} .$$

This defines the norm on $L^q(\mathbb{P})$, the space of all functions g such that $\|g\|_{L^q(\mathbb{P})} < \infty$.

Example 3.2.9 Assume that $X_1, \ldots, X_n$ are i.i.d random variables with a common law $\mathbb{P}$ and the distribution function $F(x) = \mathbb{P}(\{\omega : X_1(\omega) \le x\})$. Define the empirical distribution function

$$\mathbb{F}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j \le t\}} , \qquad t \in (-\infty, \infty) . \tag{3.2.3}$$

Assume moreover that F is continuous and strictly increasing. Then

$$\|f-g\|_{L^q(F)} = \left(\int_0^1 |f(F^{\leftarrow}(y)) - g(F^{\leftarrow}(y))|^q \mathrm{d}y\right)^{1/q} .$$

If $\mathcal{G}$ is the class of indicators $\psi_t(x) = \mathbb{1}_{(-\infty,t]}(x)$, $t \in (-\infty,\infty)$, then $\psi_t(F^{\leftarrow}(y)) = \mathbb{1}_{[0,F(t)]}(y)$. It is obvious then that the bracketing and the covering numbers of $\mathcal{G}$ are the same as in Example 3.2.7.

3.2.1 Tightness via bracketing

The above introduced bracketing and covering numbers allow us to introduce the sufficient conditions for weak convergence. Specifically, we investigate conditions under which processes indexed by class $\mathcal{G}$ converges weakly, that is

$$\mathbb{X}_n \Rightarrow \mathbb{X} \qquad \text{in } \ell^\infty(\mathcal{G}) ,$$

where $\mathbb{X}$ is a tight Borel measurable element. We will state several theorems that, under different conditions, yield weak convergence of $\mathbb{X}_n$. In order to proceed, we introduce the envelope function:

$$\mathbf{G}(x) = \sup_{g\in\mathcal{G}} |g(x)| , \quad x \in \mathcal{X} .$$

As the supremum is taken over an uncountable set, the envelope function $\mathbf{G}$ need not to be measurable, hence its moments have to be evaluated in principle with respect to the so-called outer expectation. However, we will not address this issue here since it will not be important in the context of tail empirical processes. The next result is an adaptation of Theorem 2.5.6 in [31].

Theorem 3.2.10 (Tightness via Bracketing I) *Consider the process*

$$\mathbb{X}_n(g) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \{g(X_j) - \mathbb{E}[g(X_j)]\} \ , \quad g \in \mathcal{G} ,$$

where $\{X_j\}$ is a sequence of i.i.d. random elements with values in a measurable space $\mathcal{X}$ and a common law $\mathbb{P}$, and $\mathcal{G}$ is a collection of measurable, real valued functions defined on $\mathcal{X}$. Assume that $\mathbb{E}[\mathbf{G}^2(X_1)] < \infty$ and

$$\int_0^\infty \sqrt{\log N_{[\,]}(\epsilon, \mathcal{G}, L^2(\mathbb{P}))}\mathrm{d}\epsilon < \infty \ . \tag{3.2.4}$$

Then $\mathbb{X}_n$ is tight in $\ell^\infty(\mathcal{G})$.

Example 3.2.11 [Continuation of 3.2.8] For the class of indicators indexed by $t \in [0,1]^d$ equipped with the norm L^2, we recall that the bracketing number is at most $1 + 1/\epsilon^{2d}$ whenever $\epsilon \in (0,1)$. If $\epsilon > 1$ then we need one bracket to cover $\mathcal{G}$. Hence, it suffices to consider the integral (3.2.4) from 0 to 1. Since $\log(1+a) \le 1 + \log(a)$ for $a \ge 1$, we have by substitution $u = 1 - 2d \log \epsilon$,

$$\int_0^1 \sqrt{1 + \log(1/\epsilon^{2d})}\mathrm{d}\epsilon = \frac{1}{2d}\mathrm{e}^{1/(2d)} \int_1^\infty u^{1/2}\mathrm{e}^{-u/(2d)}\mathrm{d}u < \infty \ .$$

Therefore, the empirical process

$$\mathbb{X}_n(t) = \sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\{X_j \le t\}} - t\right\} \ , \quad t \in [0,1]$$

is tight.

Theorem 3.2.10 applies to empirical processes indexed by $\mathcal{G}$ based on i.i.d. random elements $\{X_j\}$. To deal with tail empirical processes later we need to extend this to arrays of independent, but not necessary identically distributed processes. In order to do this, we consider $\mathbb{X}_n$ to be

$$\mathbb{X}_n(g) = \sum_{i=1}^{m_n} \{\mathbb{X}_{n,i}(g) - \mathbb{E}\left[\mathbb{X}_{n,i}(g)\right]\} \ , \quad g \in \mathcal{G} \ ,$$

where $\mathbb{X}_{n,i}(g)$, $i = 1, \ldots, m_n$ are independent stochastic processes and $m_n \to \infty$ as $n \to \infty$. In this thesis, the processes $\mathbb{X}_{n,i}(g)$, $i = 1, \ldots, m_n$ will have the following form:

$$\mathbb{X}_{n,i}(g) = \sum_j g(\boldsymbol{X}_{n,j}) \ ,$$

where

- $\boldsymbol{X}_{n,j} = \boldsymbol{X}_j / u_n$, where $\boldsymbol{X}_j$ are random variables with values in $\mathbb{R}_+^d$ and u_n is a deterministic sequence, $u_n \to \infty$ as $n \to \infty$,
- the summation is taken over a particular subset of integers (that depends on i),
- $$g(x) = g_t(x) = \mathbb{1}_{\{x > t\}} \ , \quad x, t \in \mathbb{R}_+^d \ .$$

Instead of the bracketing number $N_{[\,]}(\epsilon, \mathcal{G}, L^2(\mathbb{P}))$ we define a new bracketing number as follows.

Definition 3.2.12 *For each $n \geq 1$, we define the bracketing number $N_{[\,]}(\epsilon, \mathcal{G}, L_n^2)$ as the minimal number of sets N_ϵ in a partition $\mathcal{G} = \bigcup_{j=1}^{N_\epsilon} \mathcal{G}_{\epsilon,j}^n$ of the index set $\mathcal{G}$ into sets $\mathcal{G}_{\epsilon,j}^n$ such that for all $j = 1, \ldots, N_\epsilon$ we have*

$$\sum_{i=1}^{m_n} \mathbb{E}\left[\sup_{f,g \in \mathcal{G}_{\epsilon,j}^n} |\mathbb{X}_{n,i}(f) - \mathbb{X}_{n,i}(g)|^2 \right] \leq \epsilon^2 \ . \tag{3.2.5}$$

With help of this definition we re-formulate the tightness criterion from Theorem 3.2.10 in the current setting. Recall that a metric space $(\mathcal{G}, \rho)$ is totally bounded if it can be covered by a finite number of balls of radius ϵ. Also, $\|\mathbb{X}_{n,i}\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |\mathbb{X}_{n,i}(g)|$. The following result is an adaptation of Theorem 2.11.9 from [31].

Theorem 3.2.13 (Tightness via bracketing II) *For each n, let $\mathbb{X}_{n,1}, \ldots, \mathbb{X}_{n,m_n}$ be independent stochastic processes indexed by a totally bounded metric space $(\mathcal{G}, \rho)$. Suppose that*

(E1) The set $\mathcal{G}$ consists of functions g such that $\mathbb{E}[g^2(\mathbb{X}_{n,1})] < \infty$ for each $n \geq 1$. The envelope function

$$\mathbf{G}(x) = \sup_{g \in \mathcal{G}} |g(x)|$$

is finite for all x.

(E2) For every $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{m_n} \mathbb{E}\left[\|\mathbb{X}_{n,i}\|_{\mathcal{G}} \mathbb{1}_{\|\mathbb{X}_{n,i}\|_{\mathcal{G}}>\epsilon}\right] = 0 . \tag{3.2.6}$$

(E3) It holds

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{f,g:\rho(f,g)<\epsilon} \sum_{i=1}^{m_n} \mathbb{E}\left[(\mathbb{X}_{n,i}(f) - \mathbb{X}_{n,i}(g))^2\right] = 0 . \tag{3.2.7}$$

(E4) For every sequence $\epsilon_n \downarrow 0$,

$$\lim_{n\to\infty} \int_0^{\epsilon_n} \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L_n^2)} \mathrm{d}\epsilon = 0 . \tag{3.2.8}$$

Then the sequence $\sum_{i=1}^{m_n} (\mathbb{X}_{n,i} - \mathbb{E}[\mathbb{X}_{n,i}])$ is asymptotically uniformly ρ-equicontinuous in $\ell^\infty(\mathcal{G})$.

Example 3.2.14 [Continuation of Examples 3.2.1, 3.2.7, 3.2.8 and 3.2.11] Consider the case $\mathbb{X}_{n,j}(g) = n^{-1/2} g(X_j)$, where $\{X_j\}$ are i.i.d. standard uniform random variables. If $\mathcal{G}$ is the class of indicators $\psi_t = \mathbb{1}_{[0,t]}$, $t \in [0,1]$, and $m_n = n$, then obviously $\sum_{i=1}^{m_n} (\mathbb{X}_{n,i}(g) - \mathbb{E}[\mathbb{X}_{n,0}(g)]) = \mathbb{X}_n(t)$, the empirical process defined in (3.2.1). We equip $\mathcal{G}$ with the L^2-norm, so that $\rho(f,g) = \|f - g\|_{L^2}$. If $\rho(f,g) \to 0$ then the condition (3.2.7) holds:

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[(f(X_i) - g(X_i))^2\right] = \mathbb{E}[(f(X_1) - g(X_1))^2] = \rho(f,g) \to 0 .$$

As for the condition (3.2.8) we choose the partition $\mathcal{G}_{\epsilon,j}^n = \mathcal{G}_{\epsilon,j}$ as brackets $[\psi_{(j-1)\epsilon^2}, \psi_{j\epsilon^2}]$, $j = 1, \ldots, [1/\epsilon^2] + 1$. Then

$$\mathbb{E}\left[\sup_{f,g\in\mathcal{G}_{\epsilon,j}} |f(X_1) - g(X_1)|^2\right] = \mathbb{E}\left[\mathbb{1}_{\{X_1\in[0,\epsilon^2]\}}\right] = \epsilon^2 .$$

Hence, the bracketing number $N_{[]}(\epsilon, \mathcal{G}, L_n^2)$ agrees with $N_{[]}(\epsilon, \mathcal{G}, L^2)$ and is at most $1 + [1/\epsilon^2]$. The condition (3.2.8) is fulfilled.

3.2.2 Tightness via random entropy

An alternative way to verify tightness for a sequence of independent arrays is via random entropy.

Definition 3.2.15 *Let $\mathbb{X}_{n,1},\ldots,\mathbb{X}_{n,m_n}$ be stochastic processes indexed by $\mathcal{G}$. We define a random semi-metric by*

$$d_n^2(f,g)=\sum_{i=1}^{m_n}\left(\mathbb{X}_{n,i}(f)-\mathbb{X}_{n,i}(g)\right)^2\ . \tag{3.2.9}$$

The following result is an adaptation of Theorem 2.11.1 in [31].

Theorem 3.2.16 *For each n, let $\mathbb{X}_{n,0},\ldots,\mathbb{X}_{n,m_n}$ be independent stochastic processes indexed by a totally bounded metric space $(\mathcal{G},\rho)$. Suppose*

(F1) The maps

$$(x_1,\ldots,x_{n_m})\to\sup_{f,g\in\mathcal{G},\rho(f,g)<\delta}\sum_{i=1}^{m_n}e_i\left|\mathbb{X}_{n,i}(f)-\mathbb{X}_{n,i}(g)\right|^l$$

are measurable for every $\delta>0$, every vector $(e_1,\ldots,e_{m_n})\in\{-1,0,1\}^{m_n}$, every $n\in\mathbb{N}$ and both $l=1,2$.

(F2) For every $\epsilon>0$,

$$\lim_{n\to\infty}\sum_{i=1}^{m_n}\mathbb{E}\left[\|\mathbb{X}_{n,i}\|_{\mathcal{G}}^2\mathbb{1}_{\|\mathbb{X}_{n,i}\|_{\mathcal{G}}>\epsilon}\right]=0\ . \tag{3.2.10}$$

(F3) It holds:

$$\lim_{\epsilon\to0}\limsup_{n\to\infty}\sup_{f,g:\rho(f,g)<\epsilon}\sum_{i=1}^{m_n}\mathbb{E}\left[\left(\mathbb{X}_{n,i}(f)-\mathbb{X}_{n,i}(g)\right)^2\right]=0\ . \tag{3.2.11}$$

(F4) For any $\tau>0$,

$$\lim_{\delta\to0}\lim_{n\to\infty}\mathbb{P}^*\left(\int_0^\delta\sqrt{\log N(\epsilon,\mathcal{G},d_n)}\mathrm{d}\epsilon>\tau\right)=0\ . \tag{3.2.12}$$

Then the sequence $\sum_{i=1}^{m_n}\left(\mathbb{X}_{n,i}-\mathbb{E}\left[\mathbb{X}_{n,0}\right]\right)$ is asymptotically uniformly ρ-equicontinuous.

Random entropy or bracketing. We comment on different conditions that appear above.

- (F2) is a strengthened version of (E2) and also implies the Lindeberg condition.
- (F3) is the same as (E3).
- The measurability condition (F1) holds whenever the processes $\{\mathbb{X}_{n,i}(g), g \in \mathcal{G}\}$ are separable, which is the case of all processes considered in this thesis.
- Condition (F4) is always fulfilled in case of finite dimensional index set, which is the case of this thesis ($\mathbb{R}^2$).

Chapter 4

Tail empirical processes (TEP)

4.1 Introduction

The goal of this chapter is to establish results on weak convergence for tail empirical processes, using the theory described in Chapter 3. These results will be used in Chapter 5, where we deal with statistical inference. Although the results for i.i.d. vectors presented here (as stated) are new, their proofs are relatively straightforward and are provided for completeness only. On the other hand, for time series we quote a result from [25].

We first discuss the empirical processes for i.i.d. regularly varying vectors in Section 4.2. We prove weak convergence of tail empirical processes. The proof is standard, but we provide it for completeness. In Section 4.2.2, we consider the weak convergence of tail empirical processes based on random levels. Interestingly, the limiting behaviour changes as compared to deterministic levels. In Section 4.3 we state the functional central limit theorem for tail empirical processes based on regularly varying time series. The result is adapted from [25]. Finally, in Section 4.4 we prove the functional limit theorem for tail empirical processes based on i.i.d. extremally independent random vectors. The result is new (see [27]), however, the proof is relatively straightforward.

We introduce some terminology. We say that $\{u_n\}$ is a scaling sequence if $u_n \to \infty$. We say that $\{r_n\}$ is an intermediate sequence if $r_n \to \infty$ and $r_n/n \to 0$.

Notational convention: In Section 4.2 we use G and $\mathbb{G}$ for the random elements and processes that appear there; in Section 4.3 we use M and $\mathbb{M}$, while in Section 4.4 we use T and $\mathbb{T}$.

4.2 TEP for i.i.d. vectors: extremal dependence

In this section we deal with tail empirical processes for i.i.d. regularly varying random vectors. First, in Section 4.2.1 we state and prove a result on weak convergence (Theorem 4.2.1). The proof is not novel, but we provide it for completeness, using the theory described in Section 3.1. Then, in Section 4.2.2 we extend Theorem 4.2.1 to random levels, by replacing the unspecified scaling sequence u_n with the user-chosen order statistics (see Theorem 4.2.6). This method is based on [29], [22] and [25]. Interestingly, in the latter case the limit changes as compared to deterministic levels.

4.2.1 Weak convergence of tail empirical process

Assume that (X, Y) is a regularly varying random vector with index $-\alpha$ and the same marginals. For simplicity, we shall consider nonnegative random variables only. We denote the limiting measure that appears in the definition of the regular variation (cf. Definition 2.2.1) simply by $\boldsymbol{\nu}$. Let $F = F_X$ be the marginal distribution of X (and so of Y). We note here that we changed the notation from Chapter 2, where a bivariate vector was denoted by (X_1, X_2). This change is needed purely from the notational point of view, so that we have a simple way to write samples from the vector (X, Y).

When dealing with extreme observations, we are often interested in estimating

$$\mathbb{E}\left[\psi\left(\frac{X}{x}, \frac{Y}{x}\right) \mid (X, Y) \in xC\right] , \tag{4.2.1}$$

where $\psi : \mathbb{R}^2 \to \mathbb{R}$, C is a suitably chosen subset of $\overline{\mathbb{R}}^2 \setminus \{\mathbf{0}\}$ and x is large. Of course, the function and the set have to be selected in such a way that the expectation is finite.

Assume that we have an i.i.d. sample (X_j, Y_j), $j = 1, \ldots, n$, from the distribution of (X, Y).

Let u_n be a deterministic scaling sequence, that is the sequence such that $u_n \to \infty$ and $n\bar{F}(u_n) \to \infty$. For $s_0 > 0$, define the tail empirical function

$$\widetilde{G}_n(s; \psi, C) = \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n} \psi\left(\frac{X_j}{u_n}, \frac{Y_j}{u_n}\right) \mathbb{1}_{\{(X_j, Y_j) \in s u_n C\}} , \qquad s \geq s_0 , \tag{4.2.2}$$

and

$$G_n(s; \psi, C) = \mathbb{E}[\widetilde{G}_n(s; \psi, C)] .$$

If ψ is homogeneous with index γ then Lemma 2.2.4 implies

$$G(s; C, \psi) \equiv \lim_{n\to\infty} G_n(s; C, \psi) = s^{\gamma-\alpha} \int_C \psi(v_1, v_2) \boldsymbol{\nu}(\mathrm{d}v_1, \mathrm{d}v_2) , \tag{4.2.3}$$

whenever ψ satisfies the integrability condition (2.2.6). Also, define

$$\widetilde{G}_n(s) , \quad G_n(s) , \quad G(s) = s^{-\alpha} \tag{4.2.4}$$

to be $\widetilde{G}_n(s; \psi, C)$, $G_n(s; \psi, C)$ and $G(s; \psi, C)$, respectively, for the function $\psi \equiv 1$ and the set $C = \{(x_1, x_2) : x_1 > 1\}$.

Consider the tail empirical process

$$\mathbb{G}_n(s; \psi, C) = \sqrt{n\bar{F}(u_n)} \left\{\widetilde{G}_n(s; \psi, C) - G_n(s; \psi, C)\right\} , \quad s \geq s_0 . \tag{4.2.5}$$

Also, define $\mathbb{G}_n(\cdot)$ to be the process $\mathbb{G}_n(\cdot;\psi,C)$ for the function $\psi \equiv 1$ and the set $C = \{(x_1,x_2) : x_1 > 1\}$.

The main result of this section is the following weak convergence for the tail empirical function.

Theorem 4.2.1 *Let $s_0 > 0$. Assume that (X_j, Y_j), $j = 1,\dots,n$, are i.i.d. regularly varying random vectors with index $-\alpha$ and the same marginals. Assume moreover that:*

(a) *We have $u_n \to \infty$ and $n\bar{F}(u_n) \to \infty$;*

(b) *The function ψ is homogenous of order $\gamma \in \mathbb{R}$;*

(c) *For $0 < s_0 \le s \le t$ we have $tC \subseteq sC$;*

(d) *There exists $\delta > 0$ such that $\int_C |\psi(v_1,v_2)|^{2+\delta}\boldsymbol{\nu}(\mathrm{d}v_1,\mathrm{d}v_2) < \infty$;*

then

$$(\mathbb{G}_n(\cdot), \mathbb{G}_n(\cdot;\psi,C)) \Rightarrow (\mathbb{G}(\cdot), \mathbb{G}(\cdot;\psi,C)) \tag{4.2.6}$$

in $\mathbb{D}([s_0,\infty)) \times \mathbb{D}([s_0,\infty))$, where $\mathbb{G}(\cdot)$, $\mathbb{G}(\cdot;\psi,C)$ are Gaussian processes with the covariance functions

$$\mathrm{Cov}(\mathbb{G}(s), \mathbb{G}(t)) = (s \vee t)^{-\alpha} ,$$

$$\mathrm{Cov}(\mathbb{G}(s;\psi,C), \mathbb{G}(t;\psi,C)) = (s \vee t)^{2\gamma-\alpha} \int_C \psi^2(v_1,v_2)\boldsymbol{\nu}(\mathrm{d}v_1,\mathrm{d}v_2) .$$

Proof of Theorem 4.2.1. The proof is relatively standard, but we provide it for completeness. We start with the central limit theorem. Multivariate convergence follows by the Cramer-Wold device. We prove the result only for $\mathbb{G}_n(\cdot;\psi,C)$ (see Remark 4.2.4). In what follows (X,Y) has the same distribution as any of (X_j,Y_j). Also, since the set C and the function ψ are fixed, in our notation we omit a dependence on them, unless it is necessary.

In the proof, w.l.o.g. we assume that $\psi \geq 0$.

Lemma 4.2.2 *Under the conditions of Theorem 4.2.1, for each $s \geq s_0$, $\mathbb{G}_n(s;\psi,C)$ converges in distribution to a centered normal random variable.*

Proof: We prove the central limit theorem by checking Lindeberg's conditions. Let

$$Z_{n,j} = \frac{1}{\sqrt{n\bar{F}(u_n)}}\left\{\psi\left(\frac{X_j}{u_n},\frac{Y_j}{u_n}\right)\mathbb{1}_{\{(X_j,Y_j)\in su_nC\}} - \mathbb{E}\left[\psi\left(\frac{X_j}{u_n},\frac{Y_j}{u_n}\right)\mathbb{1}_{\{(X_j,Y_j)\in su_nC\}}\right]\right\}$$

so that $\mathbb{G}_n(s;\psi,C) = \sum_{j=1}^n Z_{n,j}$. Clearly, $\mathbb{E}[Z_{n,j}] = 0$. Furthermore,

$$\begin{aligned}\mathrm{Var}(\mathbb{G}_n(s;\psi,C)) = &\frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\psi^2\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in su_nC\}}\right]\\ &- \bar{F}(u_n)\left(\frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\psi\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in su_nC\}}\right]\right)^2 .\end{aligned}$$

Since $\bar{F}(u_n) \to 0$ as $n \to \infty$, Lemma 2.2.4 implies that the first term dominates and $\lim_{n\to\infty}\mathrm{Var}(\mathbb{G}_n(s;\psi,C))$ exists.

Furthermore, noting that for arbitrary $\delta > 0$ and any random variable $\mathbb{1}_{\{|Y|>c\}} \leq |Y|^\delta/c^\delta$, we have

$$\begin{aligned}\mathbb{E}[Z_{n,j}^2\mathbb{1}_{\{|Z_{n,j}|>\delta\}}] &\leq \frac{1}{(n\bar{F}(u_n))^{\delta/2}}\mathbb{E}[|Z_{n,j}|^{2+\delta}]\\ &\leq \frac{1}{(n\bar{F}(u_n))^{1+\delta/2}}\mathbb{E}\left[\psi^{2+\delta}\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in su_nC\}}\right]\end{aligned}$$

and hence

$$\sum_{j=1}^n \mathbb{E}[Z_{n,j}^2\mathbb{1}_{\{|Z_{n,j}|>\delta\}}] \leq (n\bar{F}(u_n))^{-\delta/2}\left\{\frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\psi^{2+\delta}\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in su_nC\}}\right]\right\} .$$

Using Lemma 2.2.4 and since $\delta > 0$, the expression on the right hand side converges to 0. ∎

Lemma 4.2.3 *Under the conditions of Theorem 4.2.1 the sequence of processes* $\{\mathbb{G}_n(\cdot;\psi,C)\}$, $n\geq 1$, *is tight in* $\mathbb{D}([s_0,\infty))$ *equipped with the Skorokhod* J_1 *topology.*

Proof: For $s_0<s<t$, define $(s,t]u_nC=(su_nC)\setminus(tu_nC)$ and

$$U_{n,j}(s)=\psi\left(\frac{X_j}{u_n},\frac{Y_j}{u_n}\right)\mathbb{1}_{\{(X_j,Y_j)\in su_nC\}}\ ,\qquad U^*_{n,j}(s)=U_{n,j}(s)-\mathbb{E}[U_{n,j}(s)]\ ,$$

$$U_{n,j}(s,t)=U_{n,j}(s)-U_{n,j}(t)\ ,\qquad U^*_{n,j}(s,t)=U^*_{n,j}(s)-U^*_{n,j}(t)\ ,$$

$$g_n(s;m)=\frac{1}{\bar F(u_n)}\mathbb{E}\left[|U_{n,j}(s)|^m\right]\ ,\qquad g_n(s,t;m)=g_n(s;m)-g_n(t;m)\ .$$

Then

$$\mathbb{G}_n(s;\psi,C)-\mathbb{G}_n(t;\psi,C)=\frac{1}{\sqrt{n\bar F(u_n)}}\sum_{j=1}^n U^*_{n,j}(s,t)\ .$$

We use Theorem 3.1.1. For $s_0<s_1<t<s_2$ we have

$$\begin{aligned}
&\mathbb{E}\left[|\mathbb{G}_n(s_1;\psi,C)-\mathbb{G}_n(t;\psi,C)|^2|\mathbb{G}_n(t;\psi,C)-\mathbb{G}_n(s_2;\psi,C)|^2\right]\\
&=\frac{1}{(n\bar F(u_n))^2}\sum_{j=1}^n\mathbb{E}[\big(U^*_{n,j}(s_1,t)U^*_{n,j}(t,s_2)\big)^2]\\
&\quad+\frac{1}{(n\bar F(u_n))^2}\sum_{\substack{i,j\\ i\neq j}}^n\mathbb{E}[\big(U^*_{n,i}(s_1,t)\big)^2]\mathbb{E}[\big(U^*_{n,j}(t,s_2)\big)^2]\ .
\end{aligned}\tag{4.2.7}$$

By noting that for $s_1<t<s_2$ we have $U_{n,j}(s_1,t)U_{n,j}(t,s_2)=0$, we evaluate

$$\begin{aligned}
&\big(U^*_{n,j}(s_1,t)U^*_{n,j}(t,s_2)\big)^2=(U_{n,j}(s_1,t)-\mathbb{E}[U_{n,j}(s_1,t)])^2(U_{n,j}(t,s_2)-\mathbb{E}[U_{n,j}(t,s_2)])^2\\
&=U^2_{n,j}(s_1,t)\mathbb{E}^2[U_{n,j}(t,s_2)]+U^2_{n,j}(t,s_2)\mathbb{E}^2[U_{n,j}(s_1,t)]\\
&\quad-2U_{n,j}(s_1,t)\mathbb{E}[U_{n,j}(s_1,t)]\mathbb{E}^2[U_{n,j}(t,s_2)]-2U_{n,j}(t,s_2)\mathbb{E}[U_{n,j}(t,s_2)]\mathbb{E}^2[U_{n,j}(s_1,t)]\\
&\quad+\mathbb{E}^2[U_{n,j}(s_1,t)]\mathbb{E}^2[U_{n,j}(t,s_2)]\ ,
\end{aligned}$$

so that since $\psi \geq 0$,

$$\frac{1}{\bar{F}^2(u_n)}\mathbb{E}\left[\left(U^*_{n,j}(s_1,t)U^*_{n,j}(t,s_2)\right)^2\right] = \bar{F}(u_n)g_n(s_1,t;2)g_n^2(t,s_2;1)$$
$$+ \bar{F}(u_n)g_n(t,s_2;2)g_n^2(s_1,t;1)\ .$$

Next, we deal with the term in (4.2.7). For $s < t$ we have by Hölder inequality

$$\begin{aligned}
&\mathbb{E}[(U^*_{n,j}(s,t))^2] = \mathbb{E}[(U_{n,j}(s) - \mathbb{E}U_{n,j}(s) - (U_{n,j}(t) - \mathbb{E}U_{n,j}(t)))^2] \\
&= \mathbb{E}[(U_{n,j}(s) - U_{n,j}(t))^2] + 2(\mathbb{E}[U_{n,j}(s) - \mathbb{E}U_{n,j}(t)])^2 + (\mathbb{E}[U_{n,j}(s) - U_{n,j}(t)])^2 \\
&\leq \mathbb{E}[(U_{n,j}(s) - U_{n,j}(t))^2] + 3(\mathbb{E}[U_{n,j}(s) - U_{n,j}(t)])^2 \\
&\leq 4\mathbb{E}[(U_{n,j}(s) - U_{n,j}(t))^2]\ .
\end{aligned}$$

Hence, the term in (4.2.7) is bounded by

$$\begin{aligned}
&\frac{1}{(n\bar{F}(u_n))^2}\mathbb{E}[(U^*_{n,1}(s_1,t))^2]\mathbb{E}[(U^*_{n,1}(t,s_2))^2] \\
&\leq \frac{1}{\bar{F}(u_n)}\left(4\mathbb{E}\left[\psi^2\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in(s_1,t]u_nC\}}\right]\right) \\
&\quad\times\frac{1}{\bar{F}(u_n)}\left(4\mathbb{E}\left[\psi^2\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in(t,s_2]u_nC\}}\right]\right) \\
&\leq \frac{1}{\bar{F}^2(u_n)}\left(4\mathbb{E}\left[\psi^2\left(\frac{X}{u_n},\frac{Y}{u_n}\right)\mathbb{1}_{\{(X,Y)\in(s_1,s_2]u_nC\}}\right]\right)^2 \\
&= 16g_n^2(s_1,s_2;2)\ .
\end{aligned}$$

We used here the assumption that $tC \subseteq sC$ for $s < t$. The tightness follows since $g_n(s;2)$ converges uniformly to the continuous function, and by using Theorem 3.1.1. ∎

Remark 4.2.4 Note that

$$\mathbb{G}_n(s;\psi,C) = \sqrt{n\bar{F}(u_n)}\left\{\int \psi(v_1,v_2)\mathbb{1}_{\{(v_1,v_2)\in sC\}}\bar{\nu}_n(\mathrm{d}v_1,\mathrm{d}v_2)\right\}\ , \tag{4.2.8}$$

where $\bar{\nu}_n = \nu_n - \mathbb{E}[\nu_n]$ and ν_n is a random measure

$$\nu_n(\cdot) = \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n} \delta_{(X_j/u_n, Y_j/u_n)}(\cdot) \ .$$

We can write $a_1\mathbb{G}_n(s;\psi,C)+a_2\mathbb{G}_n(s)$ in the form (4.2.8), replacing $\psi(v_1,v_2)\mathbb{1}_{\{(v_1,v_2)\in sC\}}$ with

$$a_1\psi(v_1,v_2)\mathbb{1}_{\{(v_1,v_2)\in sC\}} + a_2\mathbb{1}_{\{v_1>s,v_2>0\}} \ .$$

This in turn implies that the tail empirical processes $\mathbb{G}_n(\cdot;\psi,C)$ and $\mathbb{G}_n(\cdot)$ converge jointly.

4.2.2 Tail empirical process with random levels

To apply the weak convergence established in Theorem 4.2.1 one needs to choose u_n. The sequence depends on the marginal distribution which is unknown. Hence, we consider the tail empirical process with random levels. This method was considered in [29] and [22].

The second issue is that the centering in the tail empirical process (4.2.5) is $G_n(s;\psi,C)$ not its limit $G(s;\psi,C)$. This will be handled by an appropriate "no-bias" condition.

To proceed, choose a sequence $k = k_n$ such that $k \to \infty$ and $k/n \to 0$ and define u_n by $k = n\bar{F}(u_n)$. Let $X_{n:1} \le X_{n:2} \le \cdots \le X_{n:n}$ be order statistics from X_j, $j = 1,\dots,n$. First, from Theorem 4.2.1 we conclude the following weak convergence. Let $G_n(s) = \bar{F}(su_n)/\bar{F}(u_n)$ (cf. (4.2.4)).

Proposition 4.2.5 *Assume that the conditions of Theorem 4.2.1 are satisfied. Furthermore, assume that the distribution function F is continuous and that*

$$\lim_{n\to\infty} G_n'(s) = -\alpha s^{-\alpha-1} \ , \tag{4.2.9}$$

uniformly in a neighborhood of 1. Then

$$\left(\sqrt{k}\left\{\frac{X_{n:n-k}}{u_n} - 1\right\}, \mathbb{G}_n(\cdot;\psi,C)\right) \Rightarrow (\alpha^{-1}\mathbb{G}(1), \mathbb{G}(\cdot;\psi,C))$$

in $\mathbb{R} \times \mathbb{D}([s_0, \infty))$.

We note that the normal convergence of the order statistics is standard (see e.g. [12, Theorem 2.4.1]), but we need to argue that the convergence holds jointly.

Proof: The argument is similar to that of [29].

- By Theorem 4.2.1 and the Skorokhod representation theorem, there exists a probability space, a sequence of processes $\{\widetilde{\mathbb{G}}_n(\cdot), \widetilde{\mathbb{G}}_n(\cdot;\psi, C)\}$ and processes $\widetilde{\mathbb{G}}(\cdot), \widetilde{\mathbb{G}}(\cdot;\psi, C)$ with the same distributions as, respectively, $\{\mathbb{G}_n(\cdot), \mathbb{G}_n(\cdot;\psi, C)\}$, $\mathbb{G}(\cdot)$ and $\mathbb{G}(\cdot;\psi, C)$, such that

$$\widetilde{\mathbb{G}}_n(\cdot) \to \widetilde{\mathbb{G}}(\cdot) \ , \qquad \widetilde{\mathbb{G}}_n(\cdot;\psi, C) \to \widetilde{\mathbb{G}}(\cdot;\psi, C) \tag{4.2.10}$$

  almost surely, uniformly on compact subsets of $[s_0, \infty)$. In what follows, for simplicity of notation we will write $\mathbb{G}_n(\cdot)$, $\mathbb{G}_n(\cdot;\psi, C)$, $\mathbb{G}(\cdot)$ and $\mathbb{G}(\cdot;\psi, C)$, without using the $\sim$ notation.

- Recall (4.2.4). Let $G_n^{\leftarrow}$ and $(\widetilde{G}_n)^{\leftarrow}$ be the right continuous inverses of G_n and $\widetilde{G}_n$, respectively. Then, $G_n^{\leftarrow}(1) = 1$, $(\widetilde{G}_n)^{\leftarrow}(1) = X_{n:n-k}/u_n$ and, since F is continuous, for all $s \in [\bar{F}(0)/\bar{F}(u_n), 0]$, $G_n(G_n^{\leftarrow}(s)) = s$.

- The (random) functions $\mathbb{G}_n$ and $\widetilde{G}_n^{\leftarrow}$ belong to $\mathbb{D}$. Furthermore, their almost sure limits $\mathbb{G}$ and $G^{\leftarrow}$ are continuous and $G^{\leftarrow}$ is strictly decreasing. Hence, the convergence (4.2.10) and Theorem 3.1 in [32] imply that

$$\mathbb{G}_n(G_n^{\leftarrow}(s)) = \sqrt{k}\left\{\widetilde{G}_n \circ G_n^{\leftarrow}(s) - s\right\} \to \mathbb{G}(G^{\leftarrow}(s))$$

  almost surely, uniformly on compact subsets of $[s_0, \infty)$.

- Vervaat Lemma ([12, Lemma A.0.2]) implies that

$$\sqrt{k}\left\{(\widetilde{G}_n \circ G_n^{\leftarrow})^{\leftarrow}(s) - s\right\} \to -\mathbb{G}(G^{\leftarrow}(s))$$

  almost surely, uniformly on compact subsets of $[s_0, \infty)$.

- Assumption (4.2.9) implies that G_n is continuous and strictly decreasing in a neighborhood of 1. Thus, there exists $\epsilon > 0$ such that $G_n \circ (\widetilde{G}_n)^{\leftarrow}(s) = (\widetilde{G}_n \circ G_n^{\leftarrow})^{\leftarrow}(s)$ for $s \in (1-\epsilon, 1+\epsilon)$ and
$$\sqrt{k}\left\{G_n \circ (\widetilde{G}_n)^{\leftarrow}(s) - s\right\} \to -\mathbb{G}(G^{\leftarrow}(s)) \ , \tag{4.2.11}$$
almost surely uniformly with respect to $s \in (1-\epsilon, 1+\epsilon)$.

- Since $k \to \infty$ and $(\widetilde{G}_n)^{\leftarrow}(1) = X_{n:n-k}/u_n$, (4.2.11) implies that $G_n(X_{n:n-k}/u_n)$ converges almost surely to 1. Since $G(1) = 1$ and G_n converges uniformly to G in a neighborhood of 1, this implies that $X_{n:n-k}/u_n$ converges almost surely to 1.

- By Taylor's expansion, there exists ς_n such that $|\varsigma_n - 1| \le |(\widetilde{G}_n)^{\leftarrow}(1) - 1|$ and
$$\begin{aligned} G_n((\widetilde{G}_n)^{\leftarrow}(1)) - 1 &= G_n((\widetilde{G}_n)^{\leftarrow}(1)) - G_n(G_n^{\leftarrow}(1)) \\ &= G_n'(\varsigma_n)\left\{(\widetilde{G}_n)^{\leftarrow}(1) - G_n^{\leftarrow}(1)\right\} \\ &= G_n'(\varsigma_n)\left\{X_{n:n-k}/u_n - 1\right\} \ . \end{aligned} \tag{4.2.12}$$

- Thus, (4.2.9), (4.2.11) and (4.2.12) yield that
$$\sqrt{k}\left\{\frac{X_{n:n-k}}{u_n} - 1\right\} \to \frac{1}{\alpha}\mathbb{G}(1) \ , \tag{4.2.13}$$
almost surely.

- Since the convergences $\mathbb{G}_n(\cdot;\psi,C) \to \mathbb{G}(\cdot;\psi,C)$ and (4.2.13) hold almost surely, they hold jointly. Coming back to the original probability space, we obtain the joint weak convergence.

∎

Furthermore, we impose the following no-bias condition:
$$\lim_{n\to\infty} \sup_{s>s_0} \sqrt{k}|G_n(s;\psi,C) - G(s;\psi,C)| = 0 \ . \tag{4.2.14}$$

This condition has to be verified for each problem separately. This leads to the following empirical processes

$$\widehat{\mathbb{G}}_n(s;\psi,C) = \sqrt{k}\left\{\widehat{G}_n(s;\psi,C) - G_n(s;\psi,C)\right\}$$

$$\widehat{\widehat{\mathbb{G}}}_n(s;\psi,C) = \sqrt{k}\left\{\widehat{\widehat{G}}_n(s;\psi,C) - G_n(s;\psi,C)\right\} ,$$

where

$$\widehat{G}_n(s;\psi,C) = \frac{1}{k}\sum_{j=1}^{n}\psi\left(\frac{X_j}{u_n},\frac{Y_j}{u_n}\right)\mathbb{1}_{\{(X_j,Y_j)\in sX_{n:n-k}C\}} \tag{4.2.15}$$

and

$$\widehat{\widehat{G}}_n(s;\psi,C) = \frac{1}{k}\sum_{j=1}^{n}\psi\left(\frac{X_j}{X_{n:n-k}},\frac{Y_j}{X_{n:n-k}}\right)\mathbb{1}_{\{(X_j,Y_j)\in sX_{n:n-k}C\}} .$$

Theorem 4.2.6 *Assume that the conditions of Theorem 4.2.1 are satisfied. Furthermore, assume that the distribution function F is continuous, (4.2.9) holds and $(\mathrm{d}/\mathrm{d}s)G(s;\psi,C)$ exists and is finite in a neighborhood of 1. Then*

$$\widehat{\mathbb{G}}_n(s;\psi,C) \Rightarrow \mathbb{G}(s;\psi,C) + s^{\gamma-\alpha}\frac{1}{\alpha}G'(1;\psi,C)\mathbb{G}(1) , \tag{4.2.16}$$

and

$$\widehat{\widehat{\mathbb{G}}}_n(s;\psi,C) \Rightarrow \mathbb{G}(s;\psi,C) + \frac{1}{\alpha}s^{\gamma-\alpha}G'(1;\psi,C)\mathbb{G}(1) - \frac{\gamma}{\alpha}s^{\gamma-\alpha}G(1;\psi,C)\mathbb{G}(1) \tag{4.2.17}$$

in $\mathbb{D}([s_0,\infty))$. If moreover (4.2.14) is satisfied, then the centering $G_n(s;\psi,C)$ can be replaced with its limit $G(s;\psi,C)$.

Proof: If (4.2.14) holds, it is obvious that the centering G_n can be replaced with G. We work under this condition. Note that $\widehat{G}_n(s;\psi,C) = \widetilde{G}_n(sX_{n:n-k}/u_n;\psi,C)$, where $\widetilde{G}_n$ and $\widehat{G}_n$ are the tail empirical functions defined in (4.2.2) and (4.2.15), respectively. Then, by the homogeneity property (4.2.3),

$$\begin{aligned}\widehat{\mathbb{G}}_n(s;\psi,C) &= \mathbb{G}_n(sX_{n:n-k}/u_n;\psi,C) \\ &\quad + s^{\gamma-\alpha}\sqrt{k}\left\{G(X_{n:n-k}/u_n;\psi,C) - G(1;\psi,C)\right\} =: I_1(s) + s^{\gamma-\alpha}I_2(s)\ .\end{aligned}$$

By Proposition 4.2.5

$$\sqrt{k}\left\{\frac{X_{n:n-k}}{u_n} - 1\right\} \xrightarrow{\mathrm{d}} \frac{1}{\alpha}\mathbb{G}(1)\ , \tag{4.2.18}$$

jointly with $\mathbb{G}_n(\cdot;\psi,C)$. In particular, $X_{n:n-k}/u_n$ converges in probability to 1. Thus, by Theorem 4.2.1, the term I_1 converges weakly to $\mathbb{G}(\cdot;\psi,C)$, while by the delta method the term $I_2(s)$ converges weakly to

$$\frac{1}{\alpha}G'(1;\psi,C)\mathbb{G}(1)\ .$$

This finishes the proof of (4.2.16). Furthermore,

$$\begin{aligned}\widehat{\widehat{\mathbb{G}}}_n(s;\psi,C) &= \left(\frac{X_{n:n-k}}{u_n}\right)^{-\gamma}\mathbb{G}_n(sX_{n:n-k}/u_n;\psi,C) \\ &\quad + \sqrt{k}\left(\frac{X_{n:n-k}}{u_n}\right)^{-\gamma}\left\{G_n(sX_{n:n-k}/u_n;\psi,C) - G(sX_{n:n-k}/u_n;\psi,C)\right\} \\ &\quad + \sqrt{k}\left\{\left(\frac{X_{n:n-k}}{u_n}\right)^{-\gamma} - 1\right\}(sX_{n:n-k}/u_n)^{\gamma-\alpha}\,G(1;\psi,C) \\ &\quad + \sqrt{k}s^{\gamma-\alpha}\left\{G(X_{n:n-k}/u_n;\psi,C) - G(1;\psi,C)\right\} \\ &\quad + \sqrt{k}\left\{G(s;\psi,C) - G_n(s;\psi;C)\right\} \\ &= \left(\frac{X_{n:n-k}}{u_n}\right)^{-\gamma} I_1(s) + J_1(s) + s^{\gamma-\alpha}J_2(s) + s^{\gamma-\alpha}I_2(s) + J_3(s)\ .\end{aligned}$$

Again, by Theorem 4.4.1 and $X_{n:n-k}/u_n \xrightarrow{\mathrm{p}} 1$, the first term converges weakly to $\mathbb{G}(\cdot;\psi,C)$. The second term as well as $J_3(s)$ vanish by (4.2.14). Furthermore, the delta method, the first order Taylor expansion of $G(\cdot;\psi,C)$ around 1 and (4.2.18) yield that $s^{\gamma-\alpha}(J_2(s)+I_2(s))$ converges to

$$-\frac{\gamma}{\alpha}s^{\gamma-\alpha}G(1;\psi,C)\mathbb{G}(1) + \frac{1}{\alpha}s^{\gamma-\alpha}G'(1;\psi,C)\mathbb{G}(1)\ .$$

The convergence (4.2.17) is proven. ∎

4.3 TEP for time series

In this section we consider a multivariate regularly varying time series $\{\boldsymbol{X}_j\}$ with values in $\mathbb{R}^d$. See Section 2.3 for the precise framework. Let $\|\cdot\|$ be a vector norm on $\mathbb{R}^d$, $d \geq 1$. Let F be the distribution function of $\|\boldsymbol{X}_0\|$. Consider the sequence $u_n \to \infty$ such that $n\bar{F}(u_n) \to \infty$. Let $\psi : \mathbb{R}^{2d} \to \mathbb{R}$. Fix $s_0 > 0$ and for $s \geq s_0$ define the tail empirical functions

$$\widetilde{M}_n(s;h) = \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n-h} \mathbb{1}_{\{\|\boldsymbol{X}_j\| > su_n\}} \mathbb{1}_{\{\|\boldsymbol{X}_{j+h}\| > su_n\}} , \tag{4.3.1}$$

$$\widetilde{M}_n^{\psi}(s;h) = \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n-h} \psi\left(\frac{\boldsymbol{X}_j}{u_n}, \frac{\boldsymbol{X}_{j+h}}{u_n}\right) \mathbb{1}_{\{\|\boldsymbol{X}_j\| > su_n\}} . \tag{4.3.2}$$

Set

$$M_n(s;h) = \mathbb{E}\left[\widetilde{M}_n(s;h)\right] , \quad M_n^{\psi}(s;h) = \mathbb{E}\left[\widetilde{M}_n^{\psi}(s;h)\right] ,$$

and consider the tail empirical processes

$$\mathbb{M}_n(s;h) = \sqrt{n\bar{F}(u_n)} \left\{\widetilde{M}_n(s;h) - M_n(s;h)\right\} , \tag{4.3.3}$$

$$\mathbb{M}_n^{\psi}(s;h) = \sqrt{n\bar{F}(u_n)} \left\{\widetilde{M}_n^{\psi}(s;h) - M_n^{\psi}(s;h)\right\} . \tag{4.3.4}$$

Furthermore, under the multivariate regular variation (as defined in Section 2.2) the following limits exist:

$$M(s;h) = \lim_{n\to\infty} M_n(s;h) = \lim_{x\to\infty} \frac{\mathbb{P}(\|\boldsymbol{X}_0\| > sx, \|\boldsymbol{X}_h\| > sx)}{\mathbb{P}(\|\boldsymbol{X}_0\| > x)} , \tag{4.3.5}$$

$$\begin{aligned} M^{\psi}(s;h) &= \lim_{n\to\infty} M_n^{\psi}(s;h) \\ &= \lim_{x\to\infty} \frac{1}{\mathbb{P}(\|\boldsymbol{X}_0\| > x)} \mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}_0}{x}, \frac{\boldsymbol{X}_h}{x}\right) \mathbb{1}_{\{\|\boldsymbol{X}_0\| > sx\}}\right] . \end{aligned} \tag{4.3.6}$$

The latter limit requires a moment condition on ψ, however, we shall assume that ψ is bounded.

We proceed as follows. In Section 4.3.1 we introduce the β-mixing (absolute regularity). In Section 4.3.2 we adapt results on weak convergence of tail empirical processes for time series from [25] to the present setting. The main statement of that section is Proposition 4.3.2.

4.3.1 Weak dependence condition

In order to establish limit theorems for stationary time series, we need some extra *weak dependence* conditions. The suitable one is β-mixing (absolute regularity). For arbitrary σ-fields $\mathcal{F}$ and $\mathcal{G}$ define

$$\beta(\mathcal{F},\mathcal{G}) = \sup \sum_i \sum_j |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \ ,$$

where the supremum is taken over all partitions A_i of $\mathcal{F}$ and B_j of $\mathcal{F}$. (Note that sometimes $\beta(\mathcal{F},\mathcal{G})$ is defined with the coefficient 1/2, but it is not relevant for our purposes.)

Definition 4.3.1 *Let $\{\boldsymbol{X}_j\}$ be a stationary sequence. Let $\mathcal{F}_i^j = \sigma(\boldsymbol{X}_i, \dots, \boldsymbol{X}_j)$. Define $\beta_n = \sup_j \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{\infty}^{j+n})$. We say that the sequence is β-mixing if $\beta_n \to 0$.*

4.3.2 Weak convergence of tail empirical processes

We adapt results on weak convergence of tail empirical processes for time series from [25] to the present setting. The results there are in turn an adaptation of the theory in [16].

We consider the classical anticlustering condition. Recall that $s_0 > 0$ is fixed. For sequences $\{u_n\}$ and $\{r_n\}$, $u_n \to \infty$, $r_n \to \infty$, we will say that Condition $\mathcal{S}(u_n, r_n)$

holds if for every $s, t \geq s_0$

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{\bar{F}(u_n)} \sum_{L<|j|\leq r_n} \mathbb{E}\left[\mathbb{1}_{\{\|\boldsymbol{X}_0\|>su_n\}} \mathbb{1}_{\{\|\boldsymbol{X}_j\|>u_n t\}}\right] = 0\ . \tag{$\mathcal{S}(u_n, r_n)$}$$

To describe the covariance structure of the limiting processes, we introduce the following quantities. Recall the definition of the measures $\boldsymbol{\nu_{0,h}}$, $h \geq 0$ from (2.3.1). For $\boldsymbol{x} = (\boldsymbol{x}_0, \ldots, \boldsymbol{x}_h) \in \mathbb{R}^{d(h+1)}$ define

$$\begin{aligned} c_0(s,t) &= \lim_{n\to\infty} \frac{\mathbb{P}(\|\boldsymbol{X}_0\| > (s\vee t)u_n, \|\boldsymbol{X}_h\| > (s\vee t)u_n)}{\bar{F}(u_n)} \\ &= \int_{\mathbb{R}^{d(h+1)}} \mathbb{1}_{\{\|\boldsymbol{x}_0\|>(s\vee t)\}} \mathbb{1}_{\{\|\boldsymbol{x}_h\|>(s\vee t)\}}\, \boldsymbol{\nu_{0,h}}(\mathrm{d}\boldsymbol{x})\ , \end{aligned}$$

while for $j \geq 1$ and $\boldsymbol{y} = (\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{j+h}) \in \mathbb{R}^{d(h+j)}$,

$$\begin{aligned} c_j(s,t) &= \lim_{n\to\infty} \frac{\mathbb{P}(\|\boldsymbol{X}_0\| > su_n, \|\boldsymbol{X}_h\| > su_n, \|\boldsymbol{X}_j\| > tu_n, \|\boldsymbol{X}_{j+h}\| > tu_n)}{\bar{F}(u_n)} \\ &= \int_{\mathbb{R}^{d(h+j)}} \mathbb{1}_{\{\|\boldsymbol{y}_0\|>s\}} \mathbb{1}_{\{\|\boldsymbol{y}_h\|>s\}} \mathbb{1}_{\{\|\boldsymbol{y}_j\|>t\}} \mathbb{1}_{\{\|\boldsymbol{y}_{j+h}\|>t\}}\, \boldsymbol{\nu_{0,j+h}}(\mathrm{d}\boldsymbol{y})\ . \end{aligned}$$

Define the covariance function

$$C(s,t) = c_0(s,t) + \sum_{j=1}^{\infty} \{c_j(s,t) + c_j(t,s)\}\ . \tag{4.3.7}$$

Likewise, for a bounded function ψ define

$$\begin{aligned} c_0^{\psi}(s,t) &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\psi^2\left(\frac{\boldsymbol{X}_0}{u_n}, \frac{\boldsymbol{X}_h}{u_n}\right) \mathbb{1}_{\{\|\boldsymbol{X}_0\|>(s\vee t)u_n\}}\right]}{\bar{F}(u_n)} \\ &= \int_{\mathbb{R}^{d(h+1)}} \psi^2(\boldsymbol{x}_0, \boldsymbol{x}_h) \mathbb{1}_{\{\|\boldsymbol{x}_0\|>s\vee t\}}\, \boldsymbol{\nu_{0,h}}(\mathrm{d}\boldsymbol{x})\ , \end{aligned}$$

$$\begin{aligned} c_j^{\psi}(s,t) &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\psi\left(\frac{\boldsymbol{X}_0}{u_n}, \frac{\boldsymbol{X}_h}{u_n}\right) \psi\left(\frac{\boldsymbol{X}_j}{u_n}, \frac{\boldsymbol{X}_{j+h}}{u_n}\right) \mathbb{1}_{\{\|\boldsymbol{X}_0\|>su_n\}} \mathbb{1}_{\{\|\boldsymbol{X}_j\|>tu_n\}}\right]}{\bar{F}(u_n)} \\ &= \int_{\mathbb{R}^{d(j+h)}} \psi(\boldsymbol{y}_0, \boldsymbol{y}_h)\psi(\boldsymbol{y}_j, \boldsymbol{y}_{j+h}) \mathbb{1}_{\{\|\boldsymbol{y}_0\|>s\}} \mathbb{1}_{\{\|\boldsymbol{y}_j\|>s\}}\, \boldsymbol{\nu_{0,j+h}}(\mathrm{d}\boldsymbol{y}) \end{aligned}$$

and

$$C^{\psi}(s,t) = c_0^{\psi}(s,t) + \sum_{j=1}^{\infty}\{c_j^{\psi}(s,t) + c_j^{\psi}(t,s)\} \ . \tag{4.3.8}$$

Proposition below is the promised adaptation of [25].

Proposition 4.3.2 *Let $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ be a strictly stationary regularly varying sequence with values in $\mathbb{R}^d$, $\{u_n\}$ be a scaling sequence and $\{r_n\}$ be an intermediate sequence such that Condition $\mathcal{S}(u_n, r_n)$ holds. Assume that the sequence $\{\boldsymbol{X}_j, j \in \mathbb{Z}\}$ is absolutely regular (i.e. beta-mixing) with coefficients $\{\beta_n, n \geq 1\}$ and there exists a sequence $\{\ell_n\}$ such that*

$$\ell_n \to \infty \ , \ \ell_n/r_n \to 0 \ , \ \lim_{n\to\infty} n\beta_{\ell_n}/r_n = 0 \ . \tag{4.3.9}$$

Assume that the distribution function F of $\|\boldsymbol{X}_0\|$ is continuous and

$$\lim_{n\to\infty} n\bar{F}(u_n) = \infty \ , \quad \lim_{n\to\infty} r_n\bar{F}(u_n) = 0 \ . \tag{4.3.10}$$

Furthermore, suppose that ψ is bounded. Then, for each $s_0 > 0$, the processes $\mathbb{M}_n$, $\mathbb{M}_n^{\psi}$ defined in (4.3.3)-(4.3.4) converge weakly to the centered Gaussian processes $\mathbb{M}$, $\mathbb{M}^{\psi}$ with the covariance functions C and C^{ψ} defined in, respectively, (4.3.7)-(4.3.8). The convergence holds jointly.

4.4 TEP for i.i.d. vectors: extremal independence

The set-up of this section is the same as of Section 2.4. The random vector (X, Y) with the marginal distributions F_X and F_Y fulfills the Assumption 1. The main result of this section is Proposition 4.4.1, where we prove weak convergence of tail empirical processes.

4.4.1 Tail empirical process

Let (X_j, Y_j) are i.i.d. copies of (X, Y). Recall that the random variables X_j, Y_j are nonnegative. Let $u_n \to \infty$ be such that $n\bar{F}_X(u_n) \to \infty$ and $s_0 \in (0,1)$. Define the bivariate tail empirical distribution function (hereafter TED) by

$$\widetilde{T}_n(s,t) = \frac{1}{n\bar{F}_X(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n s, Y_j \le b(u_n)t\}} \ , \ s \ge s_0, t > 0 \ ,$$

$$\widetilde{T}_n(s,\infty) = \frac{1}{n\bar{F}_X(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n s\}} \ , \ s \ge s_0 \ .$$

Obviously $\widetilde{T}_n(s,\infty)$ agrees with $\widetilde{G}_n(s)$ in (4.2.4). Furthermore, define

$$T_n(s,t) = \frac{\mathbb{P}(X > u_n s \ , \ Y \le b(u_n)t)}{\mathbb{P}(X > u_n)} \ , \quad s \ge s_0, t > 0 \ , \tag{4.4.1}$$

$$T_n(s,\infty) = \frac{\mathbb{P}(X > u_n s)}{\mathbb{P}(X > u_n)} \ , \quad s \ge s_0 \ . \tag{4.4.2}$$

Then the functions $T_n(s,t)$ and $T_n(s,\infty)$ defined in (4.4.1)-(4.4.2) can be written as $T_n(s,t) = \mathbb{E}[\widetilde{T}_n(s,t)]$, $T_n(s,\infty) = \mathbb{E}[\widetilde{T}_n(s,\infty)]$. Define further

$$T(s,t) = \lim_{n\to\infty} T_n(s,t) \ , \quad T(s) = \lim_{n\to\infty} T_n(s,\infty) \ . \tag{4.4.3}$$

Again, $T(s)$ agrees with $G(s)$ in (4.2.4). Consider the tail empirical process

$$\mathbb{T}_n(s,t) = \sqrt{n\bar{F}_X(u_n)} \left\{ \widetilde{T}_n(s,t) - T_n(s,t) \right\}, \quad s \ge s_0, t > 0 \ . \tag{4.4.4}$$

The main result of this section is the following weak convergence for the tail empirical processes.

We need the following assumption.

Assumption 2 *Let $s_0 \in (0,1)$. The function $T(\cdot,\cdot)$ defined in (4.4.3) is continuous on $[s_0,\infty) \times \mathbb{R}_+$.*

Proposition 4.4.1 *Let Assumptions 1 and 2 hold. Then the sequence $\mathbb{T}_n$ converges weakly to a Gaussian process $\mathbb{T}$ with the covariance function*

$$\mathrm{Cov}(\mathbb{T}(s,t), \mathbb{T}(s',t')) = T(s \vee s', t \wedge t') \ .$$

The proof is relatively standard, but we provide it for completeness. Similar proof in a much more involved context of univariate time series can be found in [24]. We start with the central limit theorem. Multivariate convergence follows by the Cramer-Wold device.

Lemma 4.4.2 *Let Assumptions 1 holds. For each $s \geq s_0$, $t > 0$, $\mathbb{T}_n(s,t)$ converges in distribution to a centered normal random variable and*

$$\lim_{n\to\infty} \mathrm{Cov}(\mathbb{T}_n(s,t), \mathbb{T}_n(s',t')) = T(s \vee s', t \wedge t') \; .$$

Proof: We prove the central limit theorem by checking Lindeberg's conditions. Let

$$Z_{n,j}(s,t) = \frac{1}{\sqrt{n\bar{F}_X(u_n)}} \mathbb{1}_{\{X_j > u_n s, Y_j \leq b(u_n)t\}} \; , \quad \bar{Z}_{n,j}(s,t) = Z_{n,j}(s,t) - \mathbb{E}[Z_{n,j}(s,t)] \; , \tag{4.4.5}$$

so that $\mathbb{T}_n(s,t) = \sum_{j=1}^n \bar{Z}_{n,j}(s,t)$. Clearly, $\mathbb{E}[\bar{Z}_{n,j}(s,t)] = 0$. Furthermore,

$$\mathrm{Var}(\mathbb{T}_n(s,t)) = \frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\mathbb{1}_{\{X > u_n s, Y_1 \leq b(u_n)t\}}\right] - \bar{F}_X(u_n) \left(\frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\mathbb{1}_{\{X > u_n s, Y_1 \leq b(u_n)t\}}\right] \right)^2 \; .$$

Since $\bar{F}_X(u_n) \to 0$ as $n \to \infty$, Assumption 1 implies that the first term dominates and $\lim_{n\to\infty} \mathrm{Var}(\mathbb{T}_n(s,t))$ exists.

Furthermore, noting that for arbitrary $\delta > 0$ and any random variable Y and $c > 0$ we have $\mathbb{1}_{\{|Y| > c\}} \leq |Y|^\delta / c^\delta$, we have

$$\begin{aligned} \mathbb{E}[Z_{n,j}^2(s,t) \mathbb{1}_{\{|Z_{n,j}(s,t)| > \epsilon\}}] &\leq \epsilon^{-\delta} \mathbb{E}[|Z_{n,j}(s,t)|^{2+\delta}] \\ &\leq \frac{K}{(n\bar{F}_X(u_n))^{1+\delta/2}} \mathbb{E}\left[\mathbb{1}_{\{X_j > u_n s, Y_j \leq b(u_n)t\}}\right] \end{aligned}$$

and hence

$$\sum_{j=1}^n \mathbb{E}[Z_{n,j}^2(s,t) \mathbb{1}_{\{|Z_{n,j}(s,t)| > \epsilon\}}] \leq K(n\bar{F}_X(u_n))^{-\delta/2} \left\{ \frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\mathbb{1}_{\{X > u_n s, Y_1 \leq b(u_n)t\}}\right] \right\} \; .$$

Using Assumption 1 and since $\delta > 0$, the expression on the right hand side converges to 0.

Finally, recalling that the vectors (X_j, Y_j), $j \geq 1$, are independent, we have

$$\begin{aligned}
\mathrm{Cov}(\mathbb{T}_n(s,t), \mathbb{T}_n(s',t')) &= \mathrm{Cov}\left(\sum_{j=1}^n Z_{n,j}(s,t), \sum_{i=1}^n Z_{n,i}(s',t')\right) \\
&= \sum_{j=1}^n \sum_{i=1}^n \mathrm{Cov}(Z_{n,j}(s,t), Z_{n,i}(s',t')) \\
&= n\mathrm{Cov}(Z_{n,1}(s,t), Z_{n,1}(s',t')) \\
&= \frac{1}{\bar{F}_X(u_n)} \mathbb{E}[\mathbb{1}_{\{X > u_n s, Y_1 \leq b(u_n)t\}} \mathbb{1}_{\{X > u_n s', Y_1 \leq b(u_n)t'\}}] - o(1) \ ,
\end{aligned}$$

where $o(1)$ stands for

$$\bar{F}_X(u_n) \left(\frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\mathbb{1}_{\{X > u_n s, Y_1 \leq b(u_n)t\}}\right] \right) \left(\frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\mathbb{1}_{\{X > u_n s', Y_1 \leq b(u_n)t'\}}\right] \right) \ .$$

We obtain

$$\lim_{n\to\infty} \mathrm{Cov}(\mathbb{T}_n(s,t), \mathbb{T}_n(s',t')) = T(s \vee s', t \wedge t') \ .$$

∎

We proceed with stochastic equicontinuity. We use the methodology from Section 3.2.2.

Let $(\mathcal{F}, \rho)$ be a function space equipped with a semi-metric ρ. Recall from Definition 3.2.2 that the sequence of processes $\{\bar{Z}_n, n \geq 1\}$ indexed by $(\mathcal{F}, \rho)$ is asymptotically uniformly ρ-equicontinuous if for every $\epsilon > 0$ and an arbitrary sequence $\delta_n \downarrow 0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{f,g:\rho(f,g)<\delta_n} |\bar{Z}_n(f) - \bar{Z}_n(g)| > \epsilon \right) = 0 \ .$$

Let $Z_{n,j}(g)$, $j = 1, \dots, m_n$, be independent stochastic processes indexed by a class $\mathcal{F}$. Define the random semi-metric d_n by

$$d_n^2(f,g) = \sum_{j=1}^n \{Z_{n,j}(f) - Z_{n,j}(g)\}^2 \ .$$

Let $N(\epsilon, \mathcal{F}, d_n)$ be the minimal number of balls needed to cover $\mathcal{F}$. In the present context,

$$\mathcal{F} = \{g_{s,t}, s \geq s_0, t > 0\} \ , \qquad g_{s,t}(x,y) = \mathbb{1}_{\{x>s, y\leq t\}} \ ,$$

the metric ρ is given by

$$\rho(g_{s,t}, g_{s',t'}) = \rho((s,t),(s',t')) = |s^{-\alpha} - (s')^{-\alpha}| + |T(s_0,t) - T(s_0,t')| \ .$$

Here, we have $Z_n(g) = \sum_{i=1}^n Z_{n,j}(g) = \sum_{j=1}^n Z_{n,j}(s,t)$, where $Z_{n,j}(s,t)$ is defined in (4.4.5) and $\bar{Z}_n(g) = Z_n(g) - \mathbb{E}[Z_n(g)]$.

Lemma 4.4.3 *Let Assumptions 1 and 2 hold. Then the sequence $\mathbb{T}_n(\cdot,\cdot)$ is asymptotically uniformly equicontinuous.*

Proof: According to Theorem 3.2.16, we need to verify

$$\lim_{n\to\infty} \sum_{j=1}^n \mathbb{E}\left[\|Z_{n,j}\|_{\mathcal{F}}^2 \mathbb{1}_{\{\|Z_{n,j}\|_{\mathcal{F}}>\epsilon\}}\right] = 0 \ , \tag{4.4.6}$$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{\rho(f,g)\leq\delta} \sum_{j=1}^n \mathbb{E}[\{Z_{n,j}(f) - Z_{n,j}(g)\}^2] = 0 \ , \tag{4.4.7}$$

where $\|Z_{n,j}\|_{\mathcal{F}} = \sup_{g\in\mathcal{F}} |Z_{n,j}(g)| = \sup_{s\geq s_0, t>0} |Z_{n,j}(s,t)|$, and finiteness of the entropy integral, which is obvious since we consider cells in $\mathbb{R}^2$.

We have

$$\mathbb{E}\left[\|Z_{n,j}\|_{\mathcal{F}}^2 \mathbb{1}_{\{\|Z_{n,j}\|_{\mathcal{F}}>\epsilon\}}\right] \leq \epsilon^{-\delta} \mathbb{E}\left[\|Z_{n,j}\|_{\mathcal{F}}^{2+\delta}\right] \leq \epsilon^{-\delta} \mathbb{E}[|Z_{n,j}(s_0,\infty)|^{2+\delta}]$$

and we can conclude (4.4.6) by the same argument as in the proof of Lemma 4.4.2.

As for (4.4.7) we note that for $s_0 < s < s'$ and $t, t' \in \mathbb{R}$, we have

$$\begin{aligned}&\left|\mathbb{1}_{\{Y_j \le b_h(u_n)t, X_j > su_n\}} - \mathbb{1}_{\{Y_j \le b_h(u_n)t', X_j > s'u_n\}}\right| \\ &\qquad \le \mathbb{1}_{\{su_n < X_j \le s'u_n\}} + \mathbb{1}_{\{t \wedge t' < Y_j/b_h(u_n) \le t \vee t'\}} \mathbb{1}_{\{X_j > s_0 u_n\}} \; .\end{aligned}$$

Fix $\epsilon > 0$. The above decomposition yields

$$n\mathbb{E}\left[\{Z_{n,1}(s,t) - Z_{n,1}(s',t')\}^2\right] \le \epsilon + |T(s,\infty) - T(s',\infty)| + |T(s_0,t) - T(s_0,t')| \; ,$$

where ϵ comes from the uniform convergence of $T_n(\cdot,\infty)$ to $T(\cdot,\infty)$ and $T_n(s_0,\cdot)$ to $T(s_0,\cdot)$.

Therefore,

$$\begin{aligned}&\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\rho(f,g)\le\delta} n\mathbb{E}\left[\{Z_{n,1}(s,t) - Z_{n,1}(s',t')\}^2\right] \\ &\le \lim_{\delta\to 0}\sup_{\rho(f,g)\le\delta}\{\epsilon + \rho((s,t),(s',t'))\} = \epsilon \; .\end{aligned}$$

Since ϵ is arbitrary, the proof is concluded. ∎

Chapter 5

Statistical Inference

In this chapter, we discuss statistical inference. In Section 5.1 we introduce some notations and collect some standard results that will be used throughout this chapter. We discuss the convergence of order statistics, second order regular variation and the Hill estimator.

In Section 5.2 we discuss the estimation for i.i.d. vectors under extremal dependence. We propose an alternative nonparametric approach for estimating the conditional tail distribution that is based on the *quasi-spectral* decomposition introduced in Section 2.2.4. We argue that the proposed estimation procedure based on the quasi-spectral representation may lead to an improvement in terms of efficiency or in terms of the conditions required to achieve asymptotic normality, as compared to other nonparametric methods. We discuss estimation of the conditional tail distribution in Section 5.2.1, while expected shortfall is discussed in Section 5.2.2. Simulation studies are conducted in Section 5.2.3, while real data analysis is considered in Section 5.2.4. The material discussed in this section is the original contribution of the author, see [26].

In Section 5.3 we discuss estimation for regularly varying time series with extremal dependence. In Section 5.3.1 we prove the asymptotic normality of extremogram estimators using deterministic levels. In Section 5.3.2 we replace the deterministic levels with random levels and obtain limit theorems for the appropriately modified estimators. We note that the limiting behaviour changes as compared to the case with the deterministic levels. Section 5.3.3 is devoted to the extremogram estimators when the tail index has to be estimated. The particular case of the tail dependence coefficient is discussed in Section 5.3.4. In Section 5.3.5 we present AR(1) and a stationary solution to the stochastic recurrence equation as examples. Simulation studies and data analysis are performed in Sections 5.3.6 and 5.3.7, respectively. The material discussed in this section is based on the author's original contribution, see [30].

In Section 5.4, we use the conditional extreme value model from Section 2.4 to estimate the Marginal Expected Shortfall defined in (2.4.4). In Section 5.4.1, we construct an estimator and state its asymptotic normality. We discuss the estimation of the scaling function, estimation of the tail index and the conditional scaling exponent and estimation of expected shortfall. We also compare our results with [8] in terms of applicability and assumptions. In Section 5.4.2 we give some examples where our theory applies, followed by the case where our approach does not work. Sections 5.4.3 and 5.4.4 deal with simulation studies and data analysis. Proofs are given in Section 5.4.5. The material discussed in this section is based on the author's original contribution, see [27].

5.1 Some notation and technical preliminaries

We start with some notation that is used throughout this chapter. Let $k \to \infty$ be such that $k/n \to 0$. The sequence u_n will be defined by $k = n\bar{F}_X(u_n)$ or $k = n\bar{F}(u_n)$, as appropriate, so that $u_n \to \infty$ and $n\bar{F}_X(u_n) \to \infty$, $n\bar{F}(u_n) \to \infty$.

5.1.1 Convergence of order statistics

For a sample $X_1, \dots, X_n$ from the distribution F_X, let $X_{n:1} \le \dots \le X_{n:n}$ be increasing order statistics. Recall the process $\mathbb{G}_n(\cdot)$ from Section 4.2.1; see the comment after (4.2.5):

$$\mathbb{G}_n(s) = \sqrt{n\bar{F}_X(u_n)} \left\{ \frac{1}{n\bar{F}_X(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n s\}} - \mathbb{E}\left[\frac{1}{n\bar{F}_X(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n s\}} \right] \right\} .$$

In Proposition 4.2.5 we argued that for i.i.d. random variables X_j we have

$$\sqrt{k} \left\{ \frac{X_{n:n-k}}{u_n} - 1 \right\} \xrightarrow{\mathrm{d}} -\alpha^{-1}\mathbb{G}(1) , \tag{5.1.1}$$

jointly with another process $\mathbb{G}_n(\cdot; \psi, C)$ considered there.

Now, in the univariate case the process $\mathbb{M}_n(\cdot; 0)$ defined in (4.3.3) reduces to $\mathbb{G}_n$. Hence, Proposition 4.3.2 and the same technique as the one used to prove Proposition 4.2.5, imply (5.1.1) for dependent random variables X_j, with $\mathbb{G}$ replaced with $\mathbb{M}(\cdot; 0)$. Again, the convergence holds jointly with the processes $\mathbb{M}_n(\cdot; h)$, $\mathbb{M}_n^{\psi}(\cdot; h)$ considered there.

Finally, the same comments apply to the process $\mathbb{T}_n$ considered in Section 4.4.

In summary, for the purpose of this chapter, (5.1.1) holds jointly with different processes used in the construction of our estimators.

5.1.2 Second order regular variation

The following concept is used to control the bias in case of estimation of extremal quantities.

Definition 5.1.1 *Let F_X be a distribution function. We say that $\bar{F}_X$ is second order regularly varying with indices $-\alpha$ and β, in shorthand $\bar{F}_X \in 2RV(-\alpha, \beta)$, if there exists a bounded non increasing function η^* on $[0, \infty)$, regularly varying at infinity* *with index $-\alpha\beta$ for some $\beta \geq 0$, and such that $\lim_{t\to\infty} \eta^*(t) = 0$ and there exists a measurable function η such that for $x > 0$ and constant $c \neq 0$,*

$$\mathbb{P}(X > x) = cx^{-\alpha} \exp\left(\int_1^x \frac{\eta(u)}{u} \,\mathrm{d}u\right) , \tag{5.1.2}$$

$$\exists C > 0 \ , \quad \forall u \geq 0 \ , \quad |\eta(u)| \leq C\eta^*(u) \ . \tag{5.1.3}$$

We note that there are several different definitions that are called *second order regular variation*; see [28, Exercises 3.15-3.17], [12, Section 2.3]. Here, use the one that was introduced in [15] and used in the context of tail empirical processes in both [15] and [22].

Under the second order assumption we have the following bound; cf. [22, Lemma 4.1].

Lemma 5.1.2 *If (5.1.2) and (5.1.3) hold, if η^* is regularly varying at infinity with index $-\alpha\beta$, for some $\beta \leq 0$, then for any $\epsilon > 0$, there exists a constant C such that*

$$\forall t \geq 1 \ , \quad \forall s > 0 \ , \quad \left|\frac{\mathbb{P}(X > st)}{\mathbb{P}(X > t)} - s^{-\alpha}\right| \leq C\eta^*(t) s^{-\alpha-\alpha\beta}(s \vee s^{-1})^\epsilon \ .$$

Lemma 5.1.2 and Potter's bound (see Lemma 2.1.10) imply that for n large enough and a $\delta > 0$ we have

$$\sqrt{n\bar{F}_X(u_n)} \sup_{s>s_0} \left|\frac{\bar{F}_X(su_n)}{\bar{F}_X(u_n)} - s^{-\alpha}\right| \leq \sqrt{n\bar{F}_X(u_n)} u_n^{-\alpha\beta+\delta} \sup_{s>s_0} s^{-\alpha(\beta+1)+\epsilon} \ . \tag{5.1.4}$$

The above bound will be used to control the bias that appears when dealing with tail empirical processes. See Theorem 5.3.1 along with Remark 5.3.2, Assumption 5 of Theorem 5.4.4.

5.1.3 Hill estimator

If X is regularly varying with the tail index α, then the tail index is usually estimated by the Hill estimator:

$$H_{k,n} := \frac{1}{\widehat{\alpha}} = \frac{1}{k} \sum_{j=1}^{n} \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right) \ .$$

Under the general conditions (like i.i.d. or weak dependence) together with the second order regular variation, the Hill estimator of $1/\alpha$ is asymptotically normal with mean zero. Moreover, if the random variables are i.i.d., then the limiting variance is $1/\alpha^2$. Furthermore, in case of weak dependence and nonnegative random variables the limiting variance is

$$\alpha^{-2}\left\{1+2\sum_{j=1}^{\infty}\mathbb{E}\left[(\Theta_j\wedge 1)^{\alpha}\right]\right\} ,$$

where Θ_j is the spectral tail process introduced in Section 2.2; cf. [25].

Remark 5.1.3 For every choice of k, we obtain another estimator for $1/\alpha$. We usually plot the estimates $H_{k,n}$ against k. However, these plots typically are far from being constant, which makes it difficult to use the estimator in practice without further guideline on how to choose the value k.

5.2 Estimation for i.i.d. vectors: extremal dependence

Let (X,Y) be a regularly varying random vector with index $-\alpha$ (see Definition 2.2.1) and the same marginals denoted by $F=F_X$. For simplicity, we assume that the random variables are nonnegative. In this section our goal is to estimate quantities like

$$\mathbb{E}\left[\psi\left(\frac{X}{x},\frac{Y}{x}\right)\mid (X,Y)\in xC\right] , \tag{5.2.1}$$

where $\psi:\mathbb{R}_+^2\to\mathbb{R}$, C is a suitably chosen subset of $\overline{\mathbb{R}}^2\setminus\{\mathbf{0}\}$, $\mathbf{0}=(0,0)$ and x is large. For example, x can be chosen as $x=x_p=F_X^{\leftarrow}(1-p)=:Q_X(p)$, where p is small (The value x_p is called in financial applications the *Value-at-Risk*). Of course, the function and the set have to be selected in such the way that the expectation is finite. Special cases include estimation of the conditional tail distribution

$$\lim_{x\to\infty} \mathbb{P}(Y > yx \mid X > x) \ , \quad y > 0 \ , \tag{5.2.2}$$

the tail dependence coefficient (cf. Section 2.2.2) or the marginal expected shortfall (cf. (2.2.10))

$$\mathbb{E}\left[Y \mid X > Q_X(p)\right] \ .$$

In specific cases estimators of (5.2.1) can be obtained in a parametric or semiparametric way and rely on a particular model chosen. Alternatively, one can consider nonparametric approaches (see [4, Chapter 9] for related theory and methods, as well as an extensive list of references). Specifically, having an i.i.d. sample (X_j, Y_j), $j = 1, \ldots, n$, from (X, Y), estimation of the conditional tail distribution in (5.2.2) can be achieved by

$$\frac{1}{k}\sum_{j=1}^{n} \mathbb{1}_{\{Y_j > yX_{n:n-k}, X_j > X_{n:n-k}\}} \ , \quad y > 0 \ , \tag{5.2.3}$$

where k is a deterministic sequence such that $k \to \infty$, $k/n \to 0$ and $X_{n:1} \leq \cdots \leq X_{n:n}$ are the order statistics from $X_1, \ldots, X_n$. However, in order to provide reliable estimates of the conditional tail distribution one needs an appropriate number of pairs of observations such that the both components exceed the level $X_{n:n-k}$. This usually requires a very large number of observations. In summary, the estimator (5.2.3) may not be particularly useful in practice.

We propose an alternative nonparametric approach for estimating the conditional tail distribution and more generally for expressions like the one in (5.2.1). The idea is based on the *quasi-spectral* decomposition introduced in Section 2.2.4.

As a consequence of Proposition 2.2.5, since we assumed for simplicity that all random variables are nonnegative, the conditional tail distribution can be expressed in terms of Θ_2 as

$$\lim_{x\to\infty} \mathbb{P}(Y > yx \mid X > x) = \mathbb{E}\left[\left(\frac{\Theta_2}{y} \wedge 1\right)^\alpha\right] = \lim_{x\to\infty} \mathbb{E}\left[\left(\frac{Y}{yX} \wedge 1\right)^\alpha \mid X > x\right] \; ; \tag{5.2.4}$$

see (2.2.16). Thus, the estimator (5.2.3) can be replaced with

$$\frac{1}{k}\sum_{j=1}^{n}\left(\frac{Y_j}{yX_j}\right)^\alpha \mathbb{1}_{\{X_j > X_{n:n-k}\}} \; . \tag{5.2.5}$$

We will argue below that the estimator (5.2.5) is more efficient than the one in (5.2.3). Of course, if α is unknown, it needs to be replaced with its estimator, however, we will provide conditions that guarantee that estimation of α does not influence the limiting behaviour of the estimator of the conditional tail distribution. Our theoretical results will be also confirmed by simulation studies. Also, we note that the bivariate case can be easily extended to a general multivariate situation, still requiring only one component to be large.

Furthermore, the quasi-spectral decomposition can be useful in approximating the expected shortfall. It turns out that

$$\lim_{x\to\infty} x^{-1}\mathbb{E}[Y \mid X > x] = \mathbb{E}[V_1]\mathbb{E}[\Theta_2] = \frac{\alpha}{\alpha-1}\lim_{x\to\infty} \mathbb{E}\left[(Y/X) \mid X > x\right]$$

whenever $\alpha > 1$. Using the above identity we can construct two estimators of the expected shortfall. Asymptotic normality of an estimator that is based on the left-hand side of the above expression requires finiteness of the second moment, while an estimator motivated by the quasi-spectral representation on the right-hand side may have finite variance even when $\alpha \in (1, 2)$.

In summary, the proposed estimation procedure based on the quasi-spectral representation may lead to an improvement in terms of efficiency or in terms of the conditions required to achieve asymptotic normality, as compared to other nonparametric methods.

We discuss estimation of the conditional tail distribution in Section 5.2.1, while expected shortfall is discussed in Section 5.2.2. Simulation studies are conducted in Section 5.2.3, while real data analysis appears in Section 5.2.4.

The material discussed in this section is the original contribution of the author, see [26]. In what follows, we apply Theorems 4.2.1 and 4.2.6.

5.2.1 Conditional tail distribution

In Section 4.2 we considered the tail empirical process based on i.i.d. random vectors. The tail empirical function was defined as

$$\widetilde{G}_n(s;\psi,C) = \frac{1}{n\bar{F}(u_n)}\sum_{j=1}^{n}\psi\left(\frac{X_j}{u_n},\frac{Y_j}{u_n}\right)\mathbb{1}_{\{(X_j,Y_j)\in s u_n C\}}\ , \qquad s\ge s_0\ ; \tag{5.2.6}$$

see (4.2.2), where $u_n\to\infty$, $n\bar{F}(u_n)\to\infty$. If we choose $\psi\equiv 1$ and $C=\{(x_1,x_2): x_1>1, x_2>y\}$, $y>0$, then $\widetilde{G}_n(s;\psi,C)$ becomes

$$\widetilde{G}_n^{(1)}(s;y) := \frac{1}{n\bar{F}(u_n)}\sum_{j=1}^{n}\mathbb{1}_{\{X_j>su_n, Y_j>su_n y\}}\ . \tag{5.2.7}$$

Furthermore,

$$G_n^{(1)}(s;y) := \frac{\mathbb{P}(X>su_n, Y>su_n y)}{\mathbb{P}(X>u_n)}\ ,\quad G^{(1)}(s;y) := s^{-\alpha}\int_{(1,\infty]\times(y,\infty]}\boldsymbol{\nu}(\mathrm{d}v_1,\mathrm{d}v_2)\ , \tag{5.2.8}$$

where $\boldsymbol{\nu}$ is the exponent measure of (X,Y) that appears in the definition of the bivariate regular variation. Hence, $G^{(1)}(1;y)=\lim_{n\to\infty}\mathbb{P}(Y>u_n y\mid X>u_n)$ is the limiting conditional tail distribution and $G^{(1)}(1;1)$ is the tail dependence coefficient (since we assumed that X and Y have the same marginal distribution). Theorem 4.2.1 implies that $\sqrt{n\bar{F}(u_n)}\left\{\widetilde{G}_n^{(1)}(s;y)-G_n^{(1)}(s;y)\right\}$ converges to a Gaussian process, say, $\mathbb{G}^{(1)}(s;y)$ with the limiting variance that can be expressed in terms of the quasi-spectral representation as

$$s^{-\alpha}\mathbb{E}\left[\left(\frac{\Theta_2}{y}\wedge 1\right)^{\alpha}\right]\ . \tag{5.2.9}$$

If we choose $\psi(x_1, x_2) = (x_2/(yx_1) \wedge 1)^\alpha$ and $C = \{(x_1, x_2) : x_1 > 1\}$ then $\widetilde{G}_n(s; \psi, C)$ becomes

$$\widetilde{G}_n^{(2)}(s; y) := \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n} \left(\frac{Y_j}{yX_j} \wedge 1 \right)^\alpha \mathbb{1}_{\{X_j > su_n\}} , \tag{5.2.10}$$

$$G_n^{(2)}(s; y) := \frac{1}{\bar{F}(u_n)} \mathbb{E} \left[\left(\frac{Y}{yX} \wedge 1 \right)^\alpha \mathbb{1}_{\{X > su_n\}} \right] , \qquad G^{(2)}(s; y) = \lim_{n \to \infty} G_n^{(2)}(s; y) . \tag{5.2.11}$$

In particular, using (5.2.4),

$$G^{(2)}(1; y) := \mathbb{E} \left[\left(\frac{\Theta_2}{y} \wedge 1 \right)^\alpha \right] = \lim_{n \to \infty} \mathbb{P}(Y > u_n y \mid X > u_n) .$$

Certainly,

$$G^{(1)}(1; y) = G^{(2)}(1; y) =: G(1; y) \tag{5.2.12}$$

Theorem 4.2.1 implies that $\sqrt{n\bar{F}(u_n)} \left\{ \widetilde{G}_n^{(2)}(s; y) - G_n^{(2)}(s; y) \right\}$ converges to a Gaussian process, say, $\mathbb{G}^{(2)}(s; y)$ with the limiting variance

$$s^{-\alpha} \mathbb{E} \left[\left(\frac{\Theta_2}{y} \wedge 1 \right)^{2\alpha} \right] . \tag{5.2.13}$$

which is smaller than the one given in (5.2.9) whenever $y \geq 1$.

Hence, both tail empirical functions in (5.2.7) and (5.2.10) can be used to construct estimators of the limiting conditional tail distribution. Specifically, we can use

$$\widehat{G}_n^{(1)}(1; y) := \frac{1}{k} \sum_{j=1}^{n} \mathbb{1}_{\{Y_j > yX_{n:n-k}, X_j > X_{n:n-k}\}} , \tag{5.2.14}$$

$$\widehat{G}_n^{(2)}(1; y) := \frac{1}{k} \sum_{j=1}^{n} \left(\frac{Y_j}{yX_j} \wedge 1 \right)^\alpha \mathbb{1}_{\{X_j > X_{n:n-k}\}} , \tag{5.2.15}$$

the latter one when α is known. The above discussion indicates that the second estimator can be asymptotically more efficient than the first one.

Unknown α

Let $\widehat{\alpha}$ be an estimator of α. For example, we can choose $\widehat{\alpha}$ to be the Hill estimator considered in Section 5.1.3. We redefine $\widehat{G}_n^{(2)}(1;y)$ from (5.2.15) as

$$\widehat{G}_n^{(2),\widehat{\alpha}}(1;y) = \frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\widehat{\alpha}} \mathbb{1}_{\{X_j > X_{n:n-k}\}} \ . \tag{5.2.16}$$

We have

$$\begin{aligned}
&\sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\widehat{\alpha}} \mathbb{1}_{\{X_j > X_{n:n-k}\}} - G^{(2)}(1;y)\right\} \\
&= \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\widehat{\alpha}} \mathbb{1}_{\{X_j > X_{n:n-k}\}} - \frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_j > X_{n:n-k}\}}\right\} \\
&\quad + \sqrt{k}\left\{\frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_j > X_{n:n-k}\}} - G^{(2)}(1;y)\right\} = U_1(y) + U_2(y) \ .
\end{aligned}$$

We already know from Theorem 4.2.6, that

$$U_2(y) \Rightarrow \mathbb{G}^{(2)}(1;y) + \alpha^{-1}(G^{(2)})'(1;y)\mathbb{G}(1) \ ,$$

cf. (4.2.16), given that the bias condition (4.2.14) holds. Using the first order Taylor expansion for $\alpha \to z^\alpha$, we have

$$U_1(y) = \sqrt{k}\left\{O_P(\widehat{\alpha} - \alpha)\frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\alpha} \log\left(\frac{Y_j}{yX_j} \wedge 1\right) \mathbb{1}_{\{X_j > X_{n:n-k}\}}\right\} .$$

Let $k = o(k_\alpha)$ and $\widehat{\alpha}_{k_\alpha}$ be the Hill estimator based on k_α order statistics. Since X_j, $j \geq 1$, are i.i.d. regularly varying random variables, from Section 5.1.3 we know that $\sqrt{k_\alpha}(\widehat{\alpha}_{k_\alpha} - \alpha)$ converges to a centered normal random variable. Hence, in order to show that $U_1(y)$ is of a smaller order than $U_2(y)$ it suffices to justify that

$$\frac{1}{k}\sum_{j=1}^n \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\alpha} \log\left(\frac{Y_j}{yX_j} \wedge 1\right) \mathbb{1}_{\{X_j > X_{n:n-k}\}}$$

is bounded in probability, uniformly in y. Assume that for $\delta > 0$ we have

$$\mathbb{E}\left[(\Theta_2 \wedge 1)^{\alpha+\delta} |\log(\Theta_2 \wedge 1)|^{1+\delta}\right] < \infty \ , \tag{5.2.17}$$

where recall that Θ_2 is the random variable that appears in the quasi spectral decomposition, as discussed in Section 2.2.4. Then recalling that $k = n\bar{F}(u_n)$ and $X_{n:n-k}/u_n \xrightarrow{\mathrm{P}} 1$,

$$\limsup_{n\to\infty} \mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{n}\left(\frac{Y_j}{yX_j}\wedge 1\right)^{\alpha}\left|\log\left(\frac{Y_j}{yX_j}\wedge 1\right)\right| \mathbb{1}_{\{X_j>X_{n:n-k}\}}\right]$$
$$\leq \mathbb{E}\left[(\Theta_2\wedge 1)^{\alpha}|\log(\Theta_2\wedge 1)|\right] .$$

Hence, $U_1(y)$ is negligible and there is no effect of estimation of α. The reason for this phenomena is that the estimator of α was chosen to have the faster rate of convergence $1/\sqrt{k_\alpha}$ than the tail empirical process, as long as $k = o(k_\alpha)$. Of course, if we choose $k \approx k_\alpha$, then estimation of α will affect the limit. Furthermore, if $k_\alpha = o(k)$, then estimation of α will completely determine the limiting distribution.

The above argument relies on the condition (5.2.17). Intuitively, this condition is fulfilled whenever we have a strong extremal dependence. Consider for example the following simple linear model: $Y = \phi X + \sigma|Z|$, where $\phi \in (0,1)$, $\sigma > 0$, X is Pareto with $\alpha > 0$ and Z is standard normal, independent of X. Then $\Theta_2 = \phi$ and (5.2.17) is clearly satisfied. On the other hand, if X and Y are independent, then $\Theta_2 = 0$ and the condition is not fulfilled. However, in this case there is no reason to estimate the conditional tail expectation at the first place.

Summary of the results for the tail dependence coefficient

We summarize the above discussion in the following corollaries, gathering the conditions needed in Theorems 4.2.1, 4.2.6, applied to the present context.

The first corollary deals with estimators $\widetilde{G}_n^{(1)}(s;y)$, $\widetilde{G}_n^{(2)}(s;y)$ defined in (5.2.7) and (5.2.10); see also (5.2.8), (5.2.11). It uses Theorem 4.2.1 only.

Corollary 5.2.1 *Assume that (X_j, Y_j), $j = 1,\dots,n$, are i.i.d. regularly varying nonnegative random vectors with index $-\alpha$ and the same continuous marginals. Let u_n be a deterministic sequence such that $u_n \to \infty$, $n\bar{F}(u_n) \to \infty$.*

1. *Let $s_0 \in (0,1)$. Assume that for each $y > 0$,*

$$\lim_{n\to\infty} \sqrt{n\bar{F}(u_n)} \sup_{s\geq s_0} |G_n^{(1)}(s;y) - G^{(1)}(s;y)| = 0 \ . \tag{5.2.18}$$

Then for

$$\widetilde{G}_n^{(1)}(s;y) := \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n} \mathbb{1}_{\{X_j > su_n, Y_j > su_n y\}}$$

we have

$$\sqrt{n\bar{F}(u_n)} \left\{ \widetilde{G}_n^{(1)}(s;y) - G^{(1)}(s;y) \right\} \xrightarrow{\mathrm{d}} N\left(0, s^{-\alpha}\mathbb{E}\left[\left(\frac{\Theta_2}{y} \wedge 1\right)^{\alpha}\right]\right) \ .$$

2. *Let $s_0 \in (0,1)$. Assume that for each $y > 0$,*

$$\lim_{n\to\infty} \sqrt{n\bar{F}(u_n)} \sup_{s\geq s_0} |G_n^{(2)}(s;y) - G^{(2)}(s;y)| = 0 \ . \tag{5.2.19}$$

Then for

$$\widetilde{G}_n^{(2)}(s;y) := \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{n} \left(\frac{Y_j}{yX_j} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_j > su_n\}} \ ,$$

we have

$$\sqrt{n\bar{F}(u_n)} \left\{ \widetilde{G}_n^{(2)}(s;y) - G^{(2)}(s;y) \right\} \xrightarrow{\mathrm{d}} N\left(0, s^{-\alpha}\mathbb{E}\left[\left(\frac{\Theta_2}{y} \wedge 1\right)^{2\alpha}\right]\right) \ .$$

Remark 5.2.2 We note that (5.2.18) and (5.2.19) hold for some values of u_n (and hence of k) due to (5.2.8) and (5.2.11). The choice of k is not discussed here.

The next two results rely additionally on Theorem 4.2.6.

Corollary 5.2.3 *Assume that (X_j, Y_j), $j = 1, \ldots, n$, are i.i.d. regularly varying non-negative random vectors with index $-\alpha$ and the same continuous marginals. Assume moreover that*

$$\lim_{n\to\infty} \frac{\mathrm{d}}{\mathrm{d}s} G_n(s) = -\alpha s^{-\alpha-1} \ ,$$

uniformly in the neighbourhood of 1, where $G_n(s) = P(X_0 > su_n)/P(X_0 > u_n)$. Let u_n be a deterministic sequence such that $u_n \to \infty$, $n\bar{F}(u_n) \to \infty$. Set $k = n\bar{F}(u_n)$.

1. *Assume that for each $y > 0$, (5.2.18) holds. Assume that $(G^{(1)})'(1;y) = (\mathrm{d}/\mathrm{d}s)G^{(1)}(s;y)\,|_{s=1}$ is finite. Then for*

$$\widehat{G}_n^{(1)}(1;y) := \frac{1}{k}\sum_{j=1}^{n} \mathbb{1}_{\{Y_j > yX_{n:n-k}, X_j > X_{n:n-k}\}} ,$$

we have

$$\sqrt{k}\left\{\widehat{G}_n^{(1)}(1;y) - G^{(1)}(1;y)\right\} \stackrel{\mathrm{d}}{\to} \mathbb{G}(1;y) - (G^{(1)})'(1;y)\mathbb{G}(1) ,$$

where $\mathbb{G}(1;y)$ is a centered normal random variable with variance

$$\boldsymbol{\nu}((1,\infty)\times(y,\infty)) = \mathbb{E}\left[\left(\frac{\Theta_2}{y}\wedge 1\right)^{\alpha}\right] ,$$

and $\mathbb{G}(1)$ is a standard normal random variable.

Corollary 5.2.4 *Assume that (X_j, Y_j), $j = 1,\dots,n$, are i.i.d. regularly varying non-negative random vectors with index $-\alpha$ and the same continuous marginals. Assume moreover that*

$$\lim_{n\to\infty} \frac{\mathrm{d}}{\mathrm{d}s} G_n(s) = -\alpha s^{-\alpha-1} ,$$

uniformly in the neighbourhood of 1, where $G_n(s) = P(X_0 > su_n)/P(X_0 > u_n)$. Let u_n be a deterministic sequence such that $u_n \to \infty$, $n\bar{F}(u_n) \to \infty$. Set $k = n\bar{F}(u_n)$.

1. *Assume that for each $y > 0$, (5.2.19) holds. Assume that $(G^{(2)})'(1;y) = (\mathrm{d}/\mathrm{d}s)G^{(2)}(s;y)\,|_{s=1}$ is finite. Then for*

$$\widehat{G}_n^{(2)}(1;y) := \frac{1}{k}\sum_{j=1}^{n} \left(\frac{Y_j}{yX_j}\wedge 1\right)^{\alpha} \mathbb{1}_{\{X_j > X_{n:n-k}\}} ,$$

we have

$$\sqrt{k}\left\{\widehat{G}_n^{(2)}(1;y) - G^{(2)}(1;y)\right\} \stackrel{\mathrm{d}}{\to} \mathbb{G}^*(1;y) - (G^{(2)})'(1;y)\mathbb{G}(1) ,$$

where $\mathbb{G}^*(1;y)$ is a centered normal random variable with variance

$$\int_{(1,\infty)\times(y,\infty)} (v_2/(yv_1)\wedge 1)^{2\alpha}\boldsymbol{\nu}(\mathrm{d}v_1,\mathrm{d}v_2) = \mathbb{E}\left[\left(\frac{\Theta_2}{y}\wedge 1\right)^{2\alpha}\right] ,$$

and $\mathbb{G}(1)$ is a standard normal random variable.

2. *Moreover, if $\widehat{\alpha}_{k_\alpha}$ is the Hill estimator of α and $k = o(k_\alpha)$, then*

$$\sqrt{k}\left\{\widehat{G}_n^{(2),\widehat{\alpha}}(1;y) - G^{(2)}(1;y)\right\} \stackrel{\mathrm{d}}{\to} \mathbb{G}^*(1;y) - (G^{(2)})'(1;y)\mathbb{G}(1) .$$

5.2.2 Conditional tail expectation

Here, we briefly discuss the estimation, without precise limiting statements, as in Section 5.2.1. Let $\alpha > 1$. If we choose $\psi(x_1, x_2) = x_2$ and $C = \{(x_1, x_2) : x_1 > 1\}$ then $\widetilde{G}_n(s;\psi, C)$ in (4.2.2) becomes

$$\widetilde{G}_n^{(3)}(s) := \frac{1}{n\bar{F}(u_n)}\sum_{j=1}^{n}\frac{Y_j}{u_n}\mathbb{1}_{\{X_j > su_n\}} . \tag{5.2.20}$$

Also,

$$G_n^{(3)}(s) := \frac{1}{\bar{F}(u_n)}\mathbb{E}\left[\frac{Y}{u_n}\mathbb{1}_{\{X_1 > su_n\}}\right] , G^{(3)}(s) := s^{1-\alpha}\int_{(1,\infty]\times(0,\infty]} v_2\boldsymbol{\nu}(\mathrm{d}v_1, \mathrm{d}v_2) .$$

We note that if additionally $\alpha > 2$ then the limiting variance of $\sqrt{n\bar{F}(u_n)}\widetilde{G}_n^{(3)}(s)$ can be represented as

$$s^{2-\alpha}\frac{\alpha}{\alpha-2}\mathbb{E}[\Theta_2^2] . \tag{5.2.21}$$

If we choose $\psi(x_1, x_2) = \frac{\alpha}{\alpha-1}\frac{x_2}{x_1}$ and $C = \{(x_1, x_2) : x_1 > 1\}$ then

$$\widetilde{G}_n^{(4)}(s) := \frac{1}{n\bar{F}(u_n)}\frac{\alpha}{\alpha-1}\sum_{j=1}^{n}\frac{Y_j}{X_j}\mathbb{1}_{\{X_j > su_n\}} . \tag{5.2.22}$$

Also,

$$G^{(4)}(s) := \lim_{n\to\infty}\mathbb{E}[\widetilde{G}_n^{(4)}(s)] = s^{-\alpha}\frac{\alpha}{\alpha-1}\int_{(1,\infty]\times(0,\infty]}\frac{v_2}{v_1}\boldsymbol{\nu}(\mathrm{d}v_1, \mathrm{d}v_2) .$$

In particular, by (2.2.18)

$$G^{(4)}(1) = \frac{\alpha}{\alpha-1}\lim_{n\to\infty}\mathbb{E}\left[\frac{Y}{X} \mid X > u_n\right] = \lim_{n\to\infty}\mathbb{E}\left[\frac{Y}{u_n} \mid X > u_n\right] . \tag{5.2.23}$$

We have furthermore

$$\mathrm{Var}(\mathbb{G}^{(4)}(s)) = s^{-\alpha}\left(\frac{\alpha}{\alpha-1}\right)^2 \int_{(0,\infty)} v_2^2 \int_1^{\infty} \frac{1}{v_1^2}\boldsymbol{\nu}(\mathrm{d}v_1, \mathrm{d}v_2) \tag{5.2.24}$$

$$\leq s^{-\alpha}\left(\frac{\alpha}{\alpha-1}\right)^2 \int_{(0,\infty)} v_2^2 \int_1^{\infty} \boldsymbol{\nu}(\mathrm{d}v_1, \mathrm{d}v_2)\ . \tag{5.2.25}$$

The integral in (5.2.25) is finite whenever $\alpha > 2$. However, the integral in (5.2.24) may exists even when $\alpha < 2$ (take trivially the situation of $Y = X$ or $Y = \phi X + \sigma|Z|$, where $\phi > 0$, X is regularly varying with index $-\alpha$ and support contained in (ϵ, ∞), $\epsilon > 0$, independent of a standard normal random variable Z.)

The limiting variance $\mathrm{Var}(\mathbb{G}^{(4)}(s))$ can be written as

$$s^{-\alpha}\left(\frac{\alpha}{\alpha-1}\right)^2 \mathbb{E}[\Theta_2^2]\ . \tag{5.2.26}$$

We note that for $s = 1$ the limiting variance in (5.2.26) is smaller than the one in (5.2.21). Furthermore, the effect of estimating α is negligible if we use an estimator of α with a faster rate of convergence, as described in Section 5.2.1.

5.2.3 Implementation. Simulation studies

We perform simulation studies to illustrate our theoretical results. We consider estimation of the tail dependence coefficient

$$\lambda_{\mathrm{TDC}} := \lim_{x\to\infty} \mathbb{P}(Y > x \mid X > x)\ ,$$

cf. (2.2.9) bearing in mind that X and Y have the same distribution. We use the estimators $\widehat{G}_n^{(1)}(1;1)$, $\widehat{G}_n^{(2)}(1;1)$, $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$ defined in (5.2.14), (5.2.15), (5.2.16). At the first step we plot estimates computed for different numbers k of order statistics. Next, we conduct Monte Carlo estimation for particular choices of k (5%, 10%, 20%, 30% and 40% of observations). Number of Monte Carlo iterations is chosen to be 1000.

Our simulations indicate that the quasi-spectral method is less variable and more robust (in terms of the choice of k) than the standard empirical method, even if the parameter α has to be estimated.

A toy example: simple linear model

We simulate 1000 observations from the model $Y = \phi X + \sigma|Z|$, where $\phi \in (0,1)$, $\sigma > 0$, X is Pareto with $\alpha > 0$ and Z is standard normal, independent of X. In this case the tail dependence coefficient is ϕ^α, which is obvious since Y is approximately ϕX when X is large (the random effect of Z is negligible).

Figure 5.1 shows the estimated values using the three estimators, computed for different values of k, where k is the number of order statistics being used. On the x-axes actual values of order statistics $X_{n:1}, \ldots, X_{n:n}$ are plotted in the increasing order. Hence, the estimators computed at the left-end of each plot use a large number of order statistics, while at the right-end they use few order statistics, unlike the usual Hill plot. The first property (not surprisingly) is that the empirical estimator $\widehat{G}_n^{(1)}(1;1)$ is very sensitive with respect to the number of order statistics k, and is completely useless when plotted against large values of order statistics. The estimators motivated by the quasi-spectral representation are more "stable", even if the parameter α has to be estimated.

Figures 5.2 and 5.3 show Monte Carlo estimates of TDC using $\widehat{G}_n^{(1)}(1;1)$, $\widehat{G}_n^{(2)}(1;1)$ (Figure 5.2) and $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$ (Figure 5.3), where the estimators are computed based on $k = 5\%, 10\%, 20\%, 30\%, 40\%$ upper order statistics. The parameter α in $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$ is estimated using the Hill estimator based on $k_\alpha = 5\%, 10\%, 20\%, 40\%$ of upper order statistics.

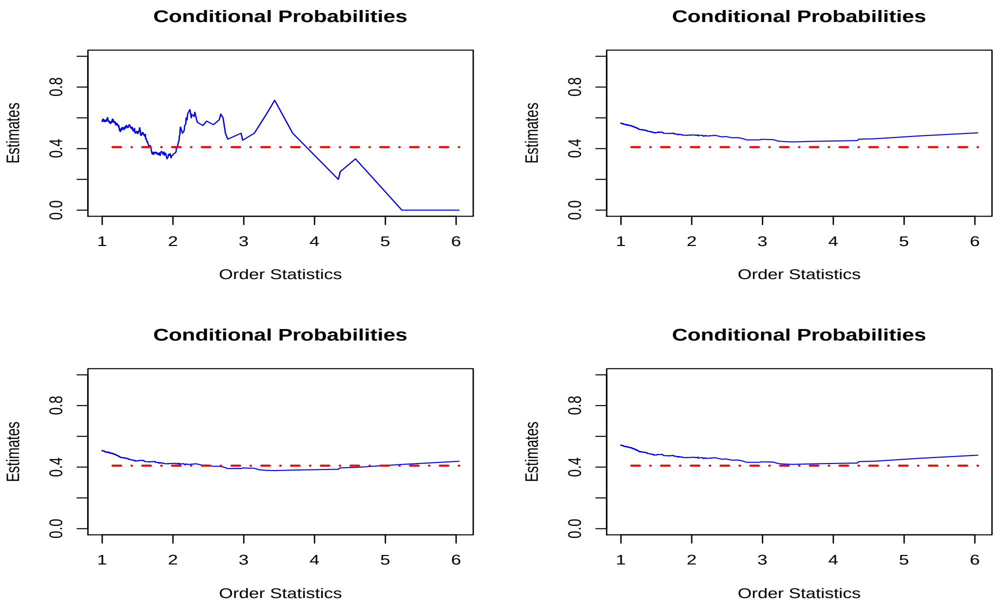

Figure 5.1: Estimation of TDC for the model $Y = \phi X + \sigma|Z|$ with $\phi = 0.8$, $\alpha = 4$, $\sigma = 0.1$. The dash-dotted line shows the true value ϕ^α. Top panel, left: estimator $\widehat{G}_n^{(1)}(1;1)$; top line, right: estimator $\widehat{G}_n^{(2)}(1;1)$; bottom panel: estimators $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$, where α is estimated using the Hill estimator based on 10% (left picture) and 20% (right picture) of order statistics.

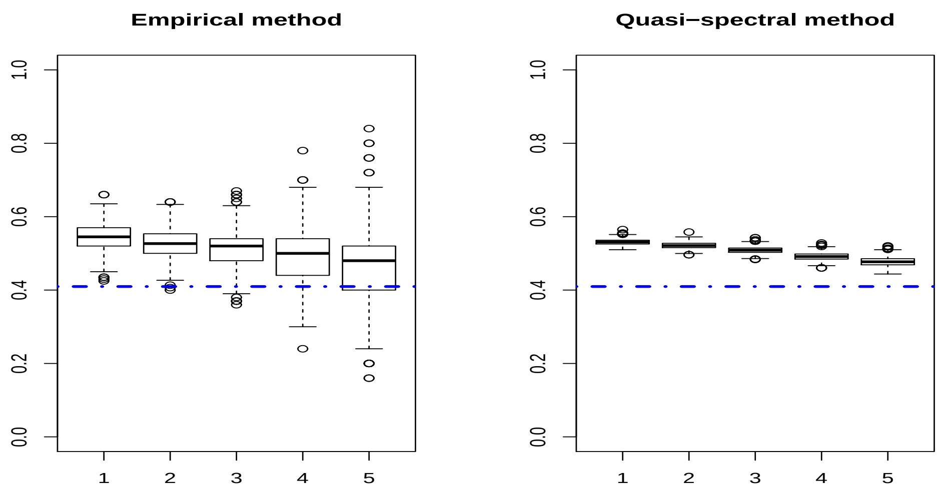

Figure 5.2: Monte Carlo estimation of TDC for the model $Y = \phi X + \sigma|Z|$ with $\phi = 0.8$, $\alpha = 4$, $\sigma = 0.1$. The dash-dotted line shows the true value ϕ^{α}. Left panel: estimator $\widehat{G}_n^{(1)}(1;1)$; right panel: estimator $\widehat{G}_n^{(2)}(1;1)$. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

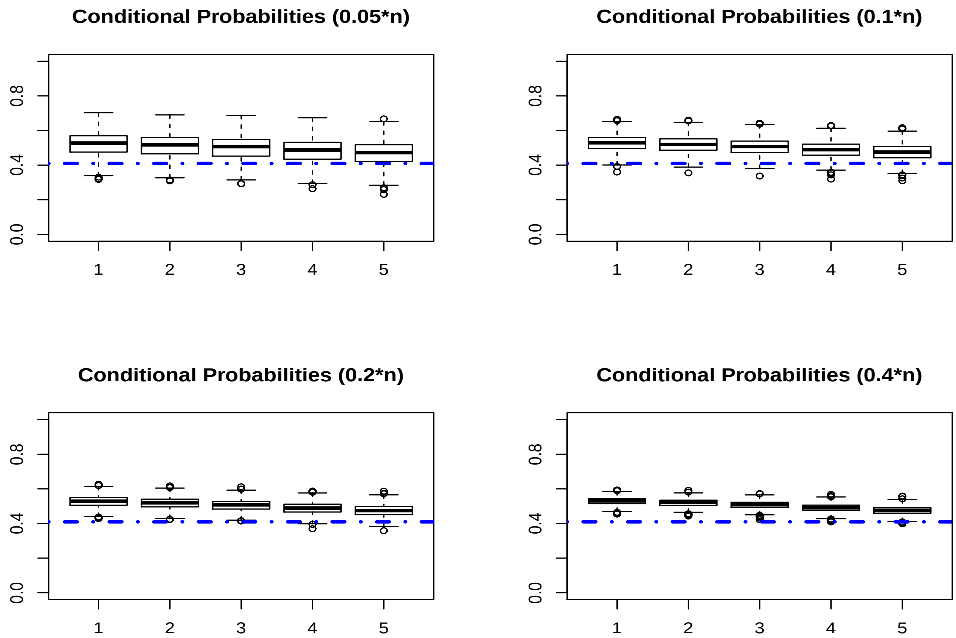

Figure 5.3: Monte Carlo estimation of TDC for the model $Y = \phi X + \sigma|Z|$ with $\phi = 0.8$, $\alpha = 4$, $\sigma = 0.1$ using different estimates of α. The dash-dotted line shows the true value ϕ^α. Estimators $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$ computed for $\widehat{\alpha}$ obtained by the Hill estimator based on 5% (top left), 10% (top right), 20% (bottom left) and 40% (bottom right) order statistics. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

Bivariate t

In Figure 5.5,we simulate 1000 observations from the bivariate t-distribution, that is $(X, Y) = \sqrt{W}(|Z_1|, |Z_2|)$, where α/W is chi-square with $\alpha = 4$ degrees of freedom and (Z_1, Z_2) are standard normal with correlation $\phi = 0.9$. In this case the tail dependence coefficient is 0.63, see [13]. The findings are very similar as for the toy model.

In Figures 5.5, 5.6 and 5.7 we show Monte Carlo estimates of TDC using $\widehat{G}_n^{(1)}(1; 1)$, $\widehat{G}_n^{(2)}(1; 1)$ for different α and ϕ. It seems that the smaller α and higher ϕ are, the better performance of the estimators are. In addition, the quasi-spectral estimators have not only smaller variation but also less bias.

In Figure 5.8, we also show Monte Carlo estimates of $\widehat{G}_n^{(2),\widehat{\alpha}}(1; 1)$ for various levels of order statistics. It is clear that the estimator is not very sensitive to the choice of order statistics.

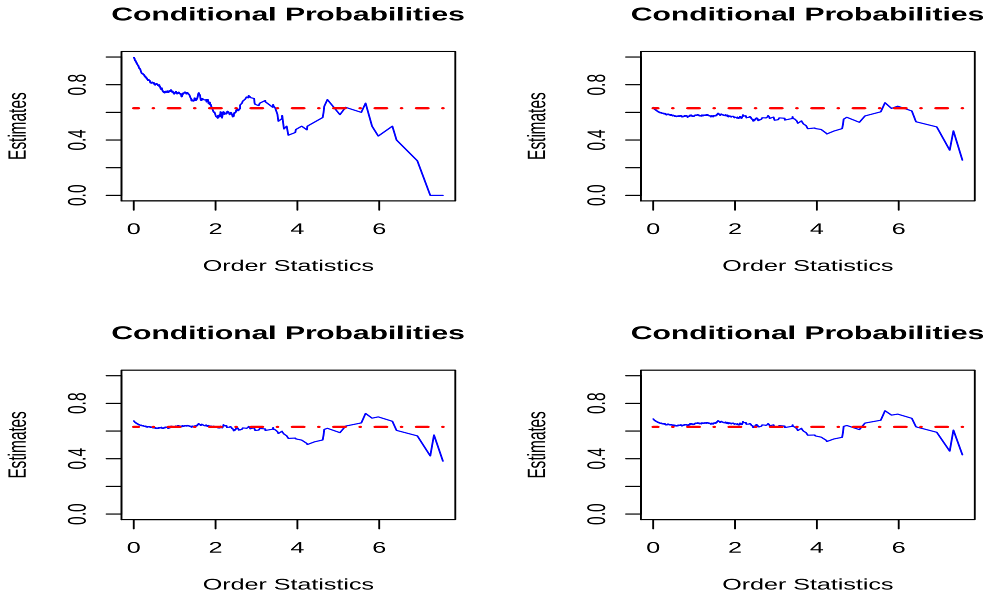

Figure 5.4: Estimation of DTC for the bivariate t with $\alpha = 4$ and $\phi = 0.9$. Top panel, left: estimator $\widehat{G}_n^{(1)}(1;1)$; top panel, right: estimator $\widehat{G}_n^{(2)}(1;1)$; bottom panel: estimators $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$, where α is estimated using the Hill estimator based on 10% (left picture) and 20% (right picture) of order statistics.

Summary of the simulation study: Our simulations indicate that the quasi-spectral method could be less variable and more robust (in terms of the choice of k) than the standard empirical method, even if the parameter α has to be estimated.

5.2.4 Data analysis

We estimate the tail dependence coefficient for the absolute log-returns of S&P500 and NASDAQ composite indices from January 2, 2013 until June 24, 2014.[1] The scatter plot in Figure 5.9 indicates strong dependence in the upper tail. This is confirmed by the estimation of the tail dependence coefficient. Again, in Figure 5.10 the quasi-spectral method is less variable than the empirical one and robust with

[1]Source: https://ca.finance.yahoo.com

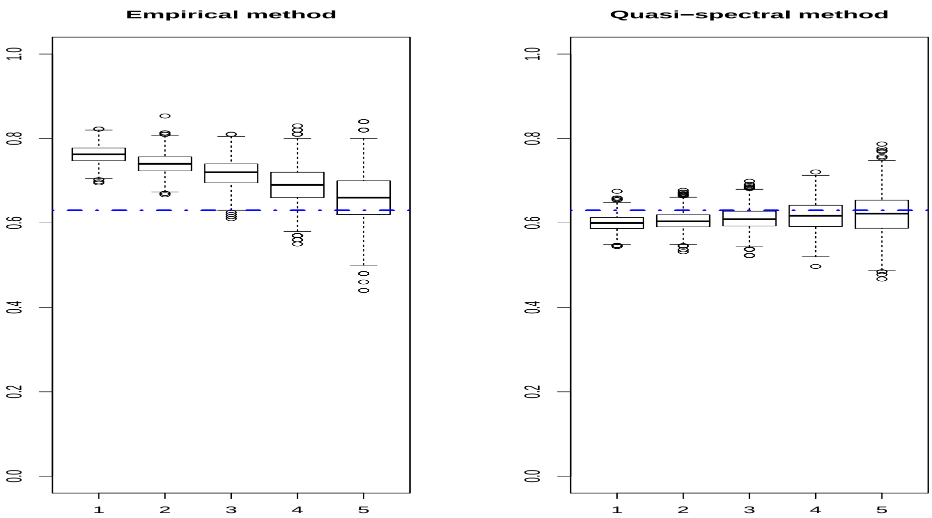

Figure 5.5: Estimation of TDC for the bivariate t with $\alpha = 4$ and $\phi = 0.9$. Left panel: estimator $\widehat{G}_n^{(1)}(1;1)$; right panel: estimator $\widehat{G}_n^{(2)}(1;1)$. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

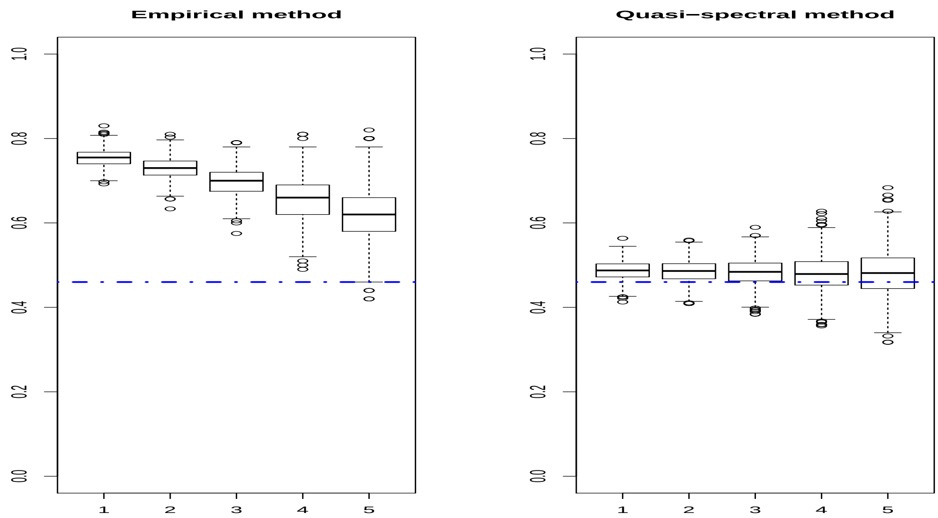

Figure 5.6: Estimation of TDC for the bivariate t with $\alpha = 10$ and $\phi = 0.9$. Left panel: estimator $\widehat{G}_n^{(1)}(1;1)$; right panel: estimator $\widehat{G}_n^{(2)}(1;1)$. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

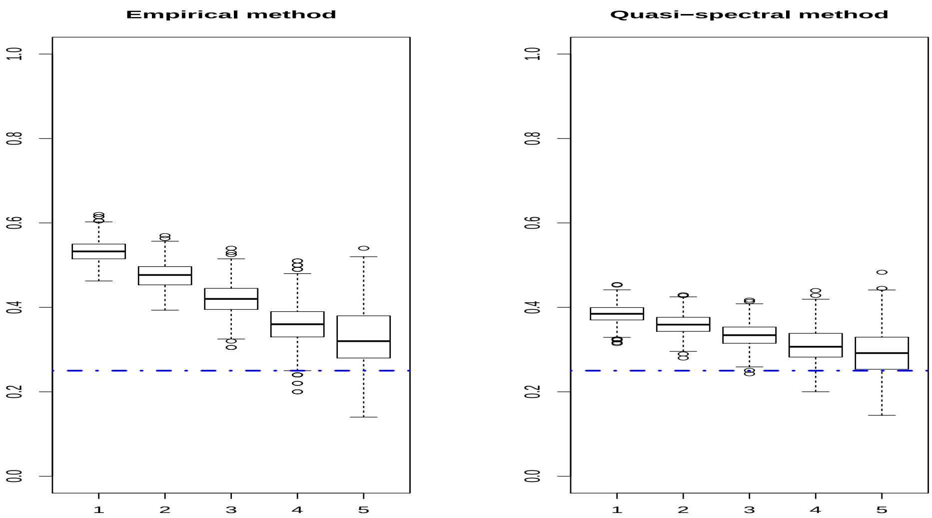

Figure 5.7: Estimation of TDC for the bivariate t with $\alpha = 4$ and $\phi = 0.5$. Left panel: estimator $\widehat{G}_n^{(1)}(1;1)$; right panel: estimator $\widehat{G}_n^{(2)}(1;1)$. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

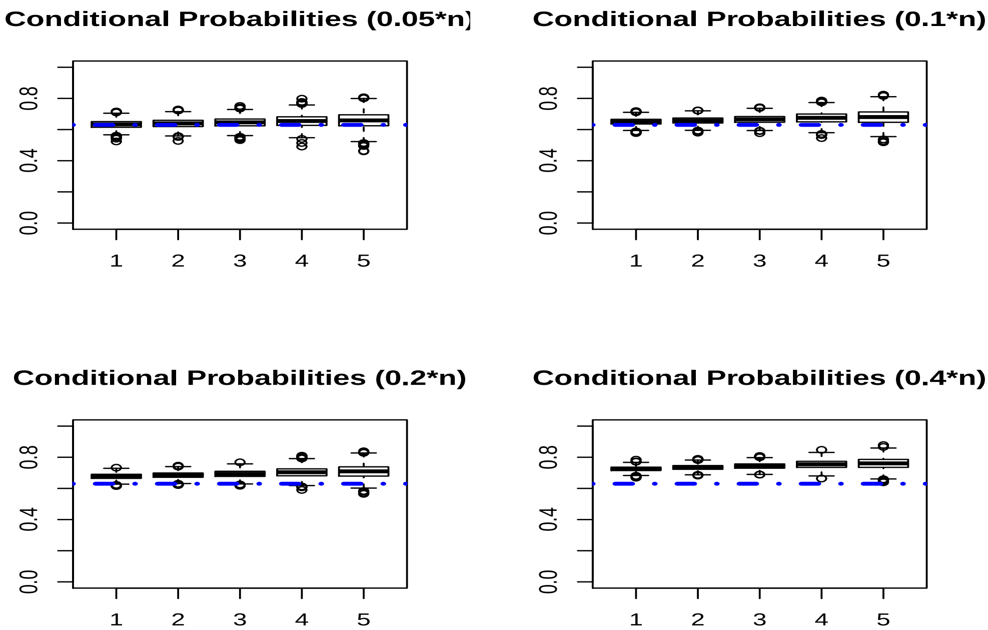

Figure 5.8: Estimation of TDC for the bivariate t. Estimators $\widehat{G}_n^{(2),\widehat{\alpha}}(1;1)$ computed for $\widehat{\alpha}$ obtained by the Hill estimator based on 5% (top left), 10% (top right), 20% (bottom left) and 40% (bottom right) order statistics. Each figure shows the boxplots for estimated values of the conditional probability computed for five different values of k. The first boxplot is computed based on 40% of observations, the second one based on 30% of observations, and the remaining ones based on 20%, 10% and 5%.

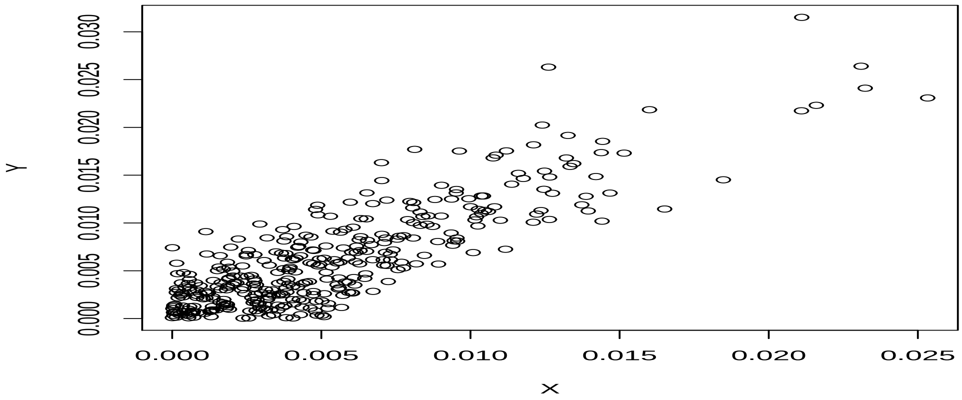

Figure 5.9: Scatter plot for S&P vs. NASDAQ

respect to the number k of the order statistics and estimation of α.

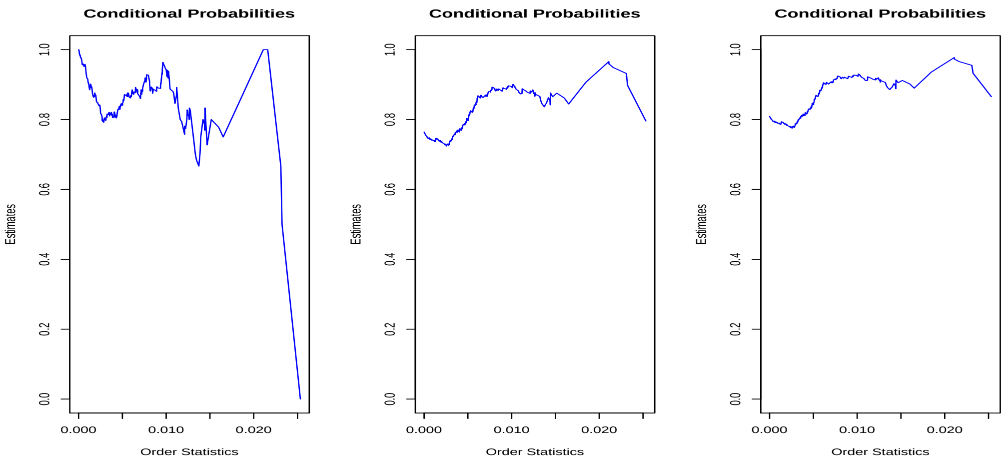

Figure 5.10: Estimation of TDC for S&P and NASDAQ. Left plot: empirical method; middle plot: quasi-spectral method with $k_\alpha = 0.1n$; right plot: quasi-spectral method with $k_\alpha = 0.2n$

5.3 Estimation for regularly varying time series

The set-up is as in Sections 2.3 and 4.3. We consider a multivariate regularly varying time series $\{\boldsymbol{X}_j\}$ with values in $\mathbb{R}^d$; $\|\cdot\|$ is a vector norm on $\mathbb{R}^d$, $d \geq 1$; F is the distribution function of $\|\boldsymbol{X}_0\|$. As usual, we consider the sequence $u_n \to \infty$ such that $n\bar{F}(u_n) \to \infty$.

Take $A = B = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| > 1\}$. Let $\{\boldsymbol{Y}_j, j \in \mathbb{N}\}$ and $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ be the corresponding tail and spectral tail processes. Then the extremogram (cf. Section 2.5 and Eq. (2.5.1)) becomes

$$\lim_{x\to\infty} \mathbb{P}(\boldsymbol{X}_h \in xB \mid \boldsymbol{X}_0 \in xA) = \mathbb{E}\left[\|\boldsymbol{\Theta}_h\|^\alpha \wedge 1\right] .$$

Thus, for the specific sets A and B the extremogram can be represented as the expectation of the spectral tail process, similarly to Section 5.2. The choice of the sets A and B may seem to be restrictive, but suffices in most applications. Also, one can easily extend our considerations to the case of two different norms. For our choice of the sets we will write $\rho(h)$ to denote the corresponding extremogram at lag h.

Extremogram estimation using the spectral tail process. If $\{\boldsymbol{\Theta}_j, j \in \mathbb{N}\}$ is the spectral tail process, then for "nice" functions $f : \mathbb{R}^d \to \mathbb{R}$ (see Section 2.3 for details) we have

$$\mathbb{E}\left[f(\boldsymbol{\Theta}_h)\right] = \lim_{x\to\infty} \mathbb{E}\left[f\left(\frac{\boldsymbol{X}_h}{\|\boldsymbol{X}_0\|}\right) \mid \|\boldsymbol{X}_0\| > x\right] .$$

Hence

$$\rho(h) = \mathbb{E}\left[\|\boldsymbol{\Theta}_h\|^\alpha \wedge 1\right] = \lim_{x\to\infty} \mathbb{E}\left[\left(\frac{\|\boldsymbol{X}_h\|}{\|\boldsymbol{X}_0\|} \wedge 1\right)^\alpha \mid \|\boldsymbol{X}_0\| > x\right] .$$

Thus, for the sets $A = B = \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| > 1\}$ we have two methods of estimating the extremogram $\rho(h)$:

$$\widehat{\rho}_n^{\mathrm{emp}}(h) = \frac{\sum_{j=1}^{n-h} \mathbb{1}_{\{\|\boldsymbol{X}_j\|>u_n\}}\mathbb{1}_{\{\|\boldsymbol{X}_{j+h}\|>u_n\}}}{\sum_{j=1}^{n} \mathbb{1}_{\{\|\boldsymbol{X}_j\|>u_n\}}} , \tag{5.3.1}$$

$$\widehat{\rho}_n^{\mathrm{sp}}(h) = \frac{\sum_{j=1}^{n-h} \left(\frac{\|\boldsymbol{X}_{j+h}\|}{\|\boldsymbol{X}_j\|} \wedge 1\right)^\alpha \mathbb{1}_{\{\|\boldsymbol{X}_j\|>u_n\}}}{\sum_{j=1}^{n} \mathbb{1}_{\{\|\boldsymbol{X}_j\|>u_n\}}} . \tag{5.3.2}$$

The first estimator was introduced in [11]. Of course, below we will replace u_n with random levels and α with its estimator.

In this section we proceed as follows. In Section 5.3.1 we discuss extremogram estimation using (5.3.1) and (5.3.2) (see Theorem 5.3.1). The latter theorem is followed by the estimation using the random levels, where u_n is replaced with the order statistics (see Theorem 5.3.3). In Theorem 5.3.5 we extend the latter result to the case of estimated α. In Section 5.3.4 we apply our theory to the tail dependence coefficient. For two time series models, $AR(1)$ and the solution to the stochastic recurrence equation, we are able to evaluate the limiting variance explicitly. We argue that the spectral method yields smaller variability; see Section 5.3.5. Our theoretical findings are supported by simulation studies. We conclude with financial data analysis.

5.3.1 Extremogram estimation: deterministic levels

Recall the notation from Section 4.3. In this section we consider the estimators $\widehat{\rho}_n^{\mathrm{emp}}(h)$ and $\widehat{\rho}_n^{\mathrm{sp}}(h)$ (cf. (5.3.1), (5.3.2)) defined with the help of the deterministic sequence u_n.

Theorem 5.3.1 *Let $\{\boldsymbol{X}_j\}$ be a regularly varying sequence with values in $\mathbb{R}^d$ such that the conditions of Proposition 4.3.2 are satisfied. Assume moreover that the following conditions hold:*

$$\sqrt{n\bar{F}(u_n)} \lim_{n\to\infty} \sup_{s\ge s_0} \{M_n(s;h) - M(s;h)\} = 0 \;, \tag{5.3.3}$$

$$\sqrt{n\bar{F}(u_n)} \lim_{n\to\infty} \sup_{s\ge s_0} \left\{M_n^{\psi}(s;h) - M^{\psi}(s;h)\right\} = 0 \;. \tag{5.3.4}$$

Then

$$\sqrt{n\bar{F}(u_n)}\,\{\widehat{\rho}_n^{\mathrm{emp}}(h) - \rho(h)\} \xrightarrow{\mathrm{d}} \mathbb{M}(1;h) - \rho(h)\mathbb{M}(1;0) \;,$$
$$\sqrt{n\bar{F}(u_n)}\,\{\widehat{\rho}_n^{\mathrm{sp}}(h) - \rho(h)\} \xrightarrow{\mathrm{d}} \mathbb{M}^{\,\mathrm{sp}}(1;h) - \rho(h)\mathbb{M}(1;0) \;.$$

Remark 5.3.2 We note that (5.3.3)-(5.3.4) are ad-hoc conditions that guarantee that the bias is negligible. Because convergence in (4.3.5)-(4.3.6) is uniform in $s \ge s_0$, the conditions are certainly satisfied for *some* sequences u_n. In particular, if $d = 1$ and $h = 0$ in (5.3.3), then the latter can be recognized as the classical no-bias condition that can be handled using the second order regular variation; see Section 5.1.2.

Proof: By Proposition 4.3.2 we have joint convergence

$$\left(\mathbb{M}_n(\cdot;0), \mathbb{M}_n(\cdot;h), \mathbb{M}_n^{\psi}(\cdot;h)\right) \Rightarrow \left(\mathbb{M}(\cdot;0), \mathbb{M}(\cdot,h), \mathbb{M}^{\psi}(\cdot;h)\right) .$$

Recall the definition of $\widehat{\rho}_n^{\mathrm{emp}}(h)$ given in (5.3.1) and of the tail empirical function given in (4.3.1). We have

$$\sqrt{n\bar{F}(u_n)}\,\{\widehat{\rho}_n^{\mathrm{emp}}(h) - \rho(h)\} = \sqrt{n\bar{F}(u_n)} \left\{ \frac{\widetilde{M}_n(1;h)}{\widetilde{M}_n(1;0)} - \rho(h) \right\}$$

$$= \sqrt{n\bar{F}(u_n)} \left\{ \widetilde{M}_n(1;h) - \rho(h) \right\} + \widetilde{M}_n(1;h) \sqrt{n\bar{F}(u_n)} \left\{ \frac{1}{\widetilde{M}_n(1;0)} - 1 \right\} .$$

The delta method and the Slutsky theorem yield that the second term converges in distribution to $-\rho(h)\mathbb{M}(1;0)$. Furthermore, under the condition (5.3.3) the process

$$\sqrt{n\bar{F}(u_n)} \left\{ \widetilde{M}_n(\cdot;h) - M(\cdot;h) \right\}$$

also converges weakly to $\mathbb{M}(\cdot;h)$. In summary, using the joint convergence,

$$\sqrt{n\bar{F}(u_n)} \left\{ \widehat{\rho}_n^{\,\mathrm{emp}}(h) - \rho(h) \right\}$$

converges in distribution to $\mathbb{M}(1;h) - \rho(h)\mathbb{M}(1;0)$, as desired.

The same approach yields the asymptotics for $\widehat{\rho}_n^{\,\mathrm{sp}}(h)$. ∎

5.3.2 Extremogram estimation: random levels

In this section we replace the deterministic sequence u_n in (5.3.1)-(5.3.2) with order statistics. Set $V_j = \|\boldsymbol{X}_j\|$ and let $V_{n:n} \geq V_{n:n-n} \geq \ldots \geq V_{n:1}$ be the corresponding order statistics. If we choose k to be an integer such that $k = n\bar{F}(u_n)$ then

$$\sqrt{k} \left\{ \frac{V_{n:n-k}}{u_n} - 1 \right\} \xrightarrow{\mathrm{d}} \alpha^{-1}\mathbb{M}(1;0) . \tag{5.3.5}$$

This follows from the convergence of the process $\mathbb{M}_n(\cdot;0)$, together with Vervaat's Lemma. The convergence (5.3.5) holds jointly with $\mathbb{M}_n^{\psi}(\cdot;h)$ which in turn follows from the Skorokhod representation. We refer to Section 5.1.1 for the detailed explanation.

The above considerations suggest the following estimators:

$$\widehat{\rho}_{n,k}^{\,\mathrm{emp}}(h) = \frac{1}{k} \sum_{j=1}^{n-h} \mathbb{1}_{\{V_j > V_{n:n-k}\}} \mathbb{1}_{\{V_{j+h} > V_{n:n-k}\}} , \tag{5.3.6}$$

$$\widehat{\rho}_{n,k}^{\,\mathrm{sp}}(h) = \frac{1}{k}\sum_{j=1}^{n-h}\left(\frac{V_{j+h}}{V_j}\wedge 1\right)^{\alpha}\mathbb{1}_{\{V_j>V_{n:n-k}\}}\ . \tag{5.3.7}$$

Theorem 5.3.3 *Let $\{\boldsymbol{X}_j\}$ be a regularly varying sequence with values in $\mathbb{R}^d$ such that the conditions of Proposition 4.3.2 are satisfied. Assume moreover that (5.3.3)-(5.3.4) hold. Then*

$$\sqrt{k}\left\{\widehat{\rho}_{n,k}^{\,\mathrm{emp}}(h)-\rho(h)\right\}\stackrel{\mathrm{d}}{\to}\mathbb{M}(1;h)-\rho(h)\mathbb{M}(1;0)\ ,$$
$$\sqrt{k}\left\{\widehat{\rho}_{n,k}^{\,\mathrm{sp}}(h)-\rho(h)\right\}\stackrel{\mathrm{d}}{\to}\mathbb{M}^{\,\mathrm{sp}}(1;h)-\rho(h)\mathbb{M}(1;0)\ .$$

Remark 5.3.4 We note that the limiting distributions in Theorems 5.3.1 and 5.3.3 are the same. Furthermore, in the proof we will skip some technical steps; the arguments below can be made precise following [29], [22] or [25].

Proof: Note that

$$\widehat{\rho}_{n,k}^{\,\mathrm{emp}}(h)=\widetilde{M}_n(V_{n:n-k}/u_n;h)$$

and recall that $M(1;h)=\rho(h)$. Thus

$$\begin{aligned}&\sqrt{k}\left\{\widehat{\rho}_{n,k}^{\,\mathrm{emp}}(h)-\rho(h)\right\}=\sqrt{k}\left\{\widetilde{M}_n(V_{n:n-k}/u_n;h)-\rho(h)\right\}\\&=\mathbb{M}_n(V_{n:n-k}/u_n;h)+\sqrt{k}\left\{M(V_{n:n-k}/u_n;h)-M(1;h)\right\}\\&+\sqrt{k}\left\{M_n(V_{n:n-k}/u_n;h)-M(V_{n:n-k}/u_n;h)\right\}:=I_1+I_2+I_3\ .\end{aligned}$$

By (5.3.5), $V_{n:n-k}/u_n\stackrel{\mathrm{P}}{\to}1$. Since the convergence in Proposition 4.3.2 is uniform on compact sets of $(0,\infty)$, we have $I_1\stackrel{\mathrm{d}}{\to}\mathbb{M}(1;h)$. The last term I_3 is negligible by (5.3.3).

As for I_2 we note that $M(s;h)$ defined in (4.3.5) can be written as

$$M(s;h)=\boldsymbol{\nu}_{\mathbf{0},\boldsymbol{h}}((s,\infty)\times\mathbb{R}^{d(h-1)}\times(s,\infty))\ .$$

The homogeneity of the limiting measure implies that $M(s;h)=s^{-\alpha}M(1;h)$. The delta method and (5.3.5) imply that

$$I_2=\sqrt{k}\left\{(V_{n:n-k}/u_n)^{-\alpha}-1\right\}M(1;h)\stackrel{\mathrm{d}}{\to}-\mathbb{M}(1;0)M(1;h)=-\rho(h)\mathbb{M}(1;0)\ ,$$

as desired.

A similar argument applies to the spectral estimator $\widehat{\rho}^{\,\mathrm{sp}}_{n,k}(h)$. ∎

5.3.3 Extremogram estimation: estimated α

The argument here follows the same lines as in [17]; see also Section 5.2.1. Let $\widehat{\alpha}$ be an estimator of α. We redefine $\widehat{\rho}^{\,\mathrm{sp}}_{n,k}(h)$ from (5.3.7) as

$$\widehat{\rho}^{\,\mathrm{sp},\widehat{\alpha}}_{n,k}(h) = \frac{1}{k}\sum_{j=1}^{n-h}\left(\frac{V_{j+h}}{V_j}\wedge 1\right)^{\widehat{\alpha}} \mathbb{1}_{\{V_j > V_{n:n-k}\}} \, . \tag{5.3.8}$$

A Taylor expansion of the function $\alpha \to z^{\alpha}$ gives

$$z^{\widehat{\alpha}} - z^{\alpha} = z^{\alpha}\log(z)(\widehat{\alpha}-\alpha) + \frac{1}{2} z^{\alpha+\lambda(\widehat{\alpha}-\alpha)}(\log z)^2(\widehat{\alpha}-\alpha)^2 \, ,$$

where λ is a random number between 0 and 1. The functions $z \to z^{\alpha}\log(z)$ and $z \to z^{\widehat{\alpha}}\log^2(z)$ are bounded for $z \in (0,1]$ and for $\widehat{\alpha}$ in a neighbourhood of α. Thus

$$\begin{aligned}
&\sqrt{k}\left\{\widehat{\rho}^{\,\mathrm{sp}}_{n,k}(h) - \widehat{\rho}^{\,\mathrm{sp},\widehat{\alpha}}_{n,k}(h)\right\} = \\
&= \sqrt{k}(\widehat{\alpha}-\alpha)\frac{1}{k}\sum_{j=1}^{n-h}\left(\frac{V_{j+h}}{V_j}\wedge 1\right)^{\alpha}\log\left(\frac{V_{j+h}}{V_j}\wedge 1\right)\mathbb{1}_{\{V_j > V_{n:n-k}\}} \\
&+ \sqrt{k}(\widehat{\alpha}-\alpha)^2 O_P\left(\frac{1}{k}\sum_{j=1}^{n-h}\mathbb{1}_{\{V_j > V_{n:n-k}\}}\right) \\
&=: \sqrt{k}(\widehat{\alpha}-\alpha)J_1 + O_P\left(\sqrt{k}(\widehat{\alpha}-\alpha)^2\right) \, .
\end{aligned} \tag{5.3.9}$$

Assume that for $\delta > 0$ we have

$$\mathbb{E}\left[(\|\boldsymbol{\Theta}_h\|\wedge 1)^{\alpha+\delta}|\log(\|\boldsymbol{\Theta}_h\|\wedge 1)|^{1+\delta}\right] < \infty \, . \tag{5.3.10}$$

Then recalling that $k = n\bar{F}(u_n)$ and $s_0 \in (0,1)$,

$$\limsup_{n\to\infty}\frac{1}{n\bar{F}(u_n)}\mathbb{E}\left[\sup_{s\ge s_0}\sum_{j=1}^{n-h}\left(\frac{V_{j+h}}{V_j}\wedge 1\right)^{\alpha}\log\left(\frac{V_{j+h}}{V_j}\wedge 1\right)\mathbb{1}_{\{V_j > su_n\}}\right]$$

$$\leq s_0^{-\alpha}\mathbb{E}\left[(\|\boldsymbol{\Theta}_h\| \wedge 1)^{\alpha}|\log(\|\boldsymbol{\Theta}_h\| \wedge 1)|\right] \ .$$

Since for sufficiently large n, $V_{n:n-k}/u_n$ lies in a neighbourhood of 1, we conclude that J_1 is bounded in probability.

As in case of (5.2.17), we have to justify the finiteness of the expression in (5.3.10). As before, this condition prevents extremal independence. For example, it is clearly fulfilled if the time series is AR(1).

Assume now that $k_\alpha \to \infty$, $k_\alpha/n \to 0$ and for a nondegenerate random variable Δ we have

$$\sqrt{k_\alpha}\,(\widehat{\alpha} - \alpha) \stackrel{\mathrm{d}}{\to} \Delta \ .$$

Depending on the interplay between k and k_α we conclude:

Theorem 5.3.5 *Let $\{\boldsymbol{X}_j\}$ be a regularly varying sequence with values in $\mathbb{R}^d$ such that the conditions of Proposition 4.3.2 are satisfied. Assume moreover that (5.3.4) and (5.3.10) hold and $\sqrt{k}/k_\alpha \to 0$. Denote $r = \lim_{n\to\infty} k/k_\alpha \in [0,\infty)$. Then:*

$$\begin{aligned}&\sqrt{k}\left\{\widehat{\rho}^{\,\mathrm{sp}}_{n,k}(h) - \rho(h)\right\}\\ &\stackrel{\mathrm{d}}{\to} \mathbb{M}^{\,\mathrm{sp}}(1;h) - \rho(h)\mathbb{M}(1;0) + \sqrt{r}\Delta\mathbb{E}\left[(\|\boldsymbol{\Theta}\|_h \wedge 1)^{\alpha}\log(\|\boldsymbol{\Theta}\|_h \wedge 1)\right] \ .\end{aligned}$$

Remark 5.3.6 The condition $\sqrt{k}/k_\alpha \to 0$ guarantees that there is no contribution from the second term in (5.3.9). If $\sqrt{k}/k_\alpha \to 0$, then there is no effect of estimation of α.

In general, the joint distribution of $\mathbb{M}^{\,\mathrm{sp}}(1;h) - \rho(h)\mathbb{M}(1;0)$ and Δ is impossible to obtain. Typically, the tail index is estimated using the Hill estimator. Then, under the conditions of Proposition 4.3.2, Δ is a normal random variable that can be expressed as an integral functional of the tail empirical process $\mathbb{M}_n(\cdot;0)$; see [25].

5.3.4 Application to the tail dependence coefficient

Note that $M(1;h) = \rho(h)$ and for the specific choice of

$$\psi(\boldsymbol{x}_0, \boldsymbol{x}_h) = \min\{(\|\boldsymbol{x}_h\|/\|\boldsymbol{x}_0\|)^\alpha, 1\} ,$$

we also have $M^\psi(1;h) = \rho(h)$. For this choice of ψ, we write $\widetilde{M}_n^{\mathrm{sp}}$, M_n^{sp}, $\mathbb{M}_n^{\mathrm{sp}}$ instead of using the ψ-notation.

We note that

$$\widehat{\rho}_n^{\mathrm{emp}}(h) = \frac{\widetilde{M}_n(1;h)}{\widetilde{M}_n(1;0)} , \quad \widehat{\rho}_n^{\mathrm{sp}}(h) = \frac{\widetilde{M}_n^{\mathrm{sp}}(1;h)}{\widetilde{M}_n(1;0)} ,$$

hence to obtain limiting distribution for both estimators we need to study weak convergence of tail empirical processes defined in (4.3.3)-(4.3.4).

In this section we shall assume that $\{X_j, j \in \mathbb{N}\}$ is a univariate regularly sequence of nonnegative random variables. Hence, we have $F = F_X$, where F_X is the distribution function of X_0. The tail dependence coefficient at lag h (see Section 2.2) is defined as

$$\lambda(h) = \lambda_{\mathrm{TDC}}(h) = \lim_{x\to\infty} \mathbb{P}(X_h > x \mid X_0 > x) .$$

Thus, $\lambda(h) = \rho_{AB}(h)$ for $A = B = (1, \infty)$.

Empirical estimator

The empirical estimator (5.3.6) becomes

$$\widehat{\rho}_{n,k}^{\,\mathrm{emp}}(h) = \frac{1}{k}\sum_{j=1}^{n-h} \mathbb{1}_{\{X_j > X_{n:n-k}, X_{j+h} > X_{n:n-k}\}} . \tag{5.3.11}$$

Theorem 5.3.1 implies the limiting distribution of the estimator is

$$\mathbb{M}(1;h) - \lambda(h)\mathbb{M}(1;0) ,$$

where $\mathbb{M}(\cdot;h)$ is a Gaussian process with the covariance

$$C^{(h)}(s,t) = c_0^{(h)}(s,t) + \sum_{j=1}^{\infty}\{c_j^{(h)}(s,t) + c_j^{(h)}(t,s)\} ,$$

where

$$c_0^{(h)}(s,t) = \lim_{n\to\infty} \frac{\mathbb{P}(X_0 > (s\vee t)u_n, X_h > (s\vee t)u_n)}{\bar{F}_X(u_n)} ,$$

$$c_j^{(h)}(s,t) = \lim_{n\to\infty} \frac{\mathbb{P}(X_0 > su_n, X_h > u_n t, X_j > su_n, X_{j+h} > tu_n)}{\bar{F}_X(u_n)} , \quad j \geq 1 .$$

Using the associated tail process $\{Y_j\}$ or the spectral tail process $\{\Theta_j\}$ (w.r.t. the norm $\|\boldsymbol{x}\| = x$), the coefficients $c_j^{(h)}(1,1)$ can be represented as:

$$c_0^{(h)}(1,1) = \mathbb{P}(Y_h > 1 \mid Y_0 > 1) = \mathbb{E}\left[\Theta_h^\alpha \wedge 1\right]$$

and for $j \geq 1$:

$$\begin{aligned} c_j^{(h)}(1,1) &= \mathbb{P}(Y_h > 1, Y_j > 1, Y_{j+h} > 1 \mid Y_0 > 1) \\ &= \mathbb{E}\left[\{\Theta_h \wedge \Theta_j \wedge \Theta_{j+h}\}^\alpha \wedge 1\right] . \end{aligned}$$

This yields

$$\operatorname{Var}(\mathbb{M}(1;h)) = C^{(h)}(1,1) = \mathbb{E}\left[\Theta_h^\alpha \wedge 1\right] + 2\sum_{j=1}^{\infty}\mathbb{E}[(\Theta_h \wedge \Theta_j \wedge \Theta_{j+h})^\alpha \wedge 1] , \tag{5.3.12}$$

$$\operatorname{Var}(\mathbb{M}(1;0)) = C^{(0)}(1,1) = 1 + 2\sum_{j=1}^{\infty}\mathbb{E}[\Theta_j^\alpha \wedge 1] . \tag{5.3.13}$$

Furthermore

$$\operatorname{Cov}(\mathbb{M}(s;h), \mathbb{M}(t;0)) = e_0(s,t) + \sum_{j=1}^{\infty}\{e_j(s,t) + f_j(s,t)\} ,$$

where

$$e_0(s,t) = \lim_{n\to\infty} \frac{\mathbb{P}(X_0 > (s\vee t)u_n, X_h > su_n)}{\bar{F}_X(u_n)} ,$$

$$e_j(s,t) = \lim_{n\to\infty} \frac{\mathbb{P}(X_0 > su_n, X_h > su_n, X_j > tu_n)}{\bar{F}_X(u_n)} ,$$
$$f_j(s,t) = \lim_{n\to\infty} \frac{\mathbb{P}(X_j > su_n, X_{j+h} > su_n, X_0 > tu_n)}{\bar{F}_X(u_n)}$$

so that

$$\begin{aligned}&\operatorname{Cov}(\mathbb{M}(1;h), \mathbb{M}(1;0)) \\ &= \mathbb{E}\left[\Theta_h^\alpha \wedge 1\right] + \sum_{j=1}^{\infty} \mathbb{E}[(\Theta_h \wedge \Theta_j)^\alpha \wedge 1] + \sum_{j=1}^{\infty} \mathbb{E}[(\Theta_j \wedge \Theta_{j+h})^\alpha \wedge 1] .\end{aligned} \tag{5.3.14}$$

Combination of (5.3.12) and (5.3.14) (with $h \geq 1$ and $h = 0$) yields

$$\begin{aligned}&\operatorname{Var}\left(\mathbb{M}(1;h) - \lambda(h)\mathbb{M}(1;0)\right) = \\ &= \mathbb{E}\left[\Theta_h^\alpha \wedge 1\right] + 2\sum_{j=1}^{\infty} \mathbb{E}[(\Theta_h \wedge \Theta_j \wedge \Theta_{j+h})^\alpha \wedge 1] \\ &+ \lambda^2(h)\left\{1 + 2\sum_{j=1}^{\infty} \mathbb{E}[\Theta_j^\alpha \wedge 1]\right\} \\ &- 2\lambda(h)\left\{\mathbb{E}\left[\Theta_h^\alpha \wedge 1\right] + \sum_{j=1}^{\infty} \mathbb{E}[(\Theta_h \wedge \Theta_j)^\alpha \wedge 1] + \sum_{j=1}^{\infty} \mathbb{E}[(\Theta_h \wedge \Theta_j \wedge \Theta_{j+h})^\alpha \wedge 1]\right\} .\end{aligned} \tag{5.3.15}$$

Spectral Estimator

Here the spectral estimator (5.3.7) becomes

$$\widehat{\rho}^{\,\mathrm{sp}}_{n,k}(h) = \frac{1}{k}\sum_{j=1}^{n-h} \left(\frac{X_{j+h}}{X_j} \wedge 1\right)^\alpha \mathbb{1}_{\{X_j > X_{n:n-k}\}} . \tag{5.3.16}$$

The limiting distribution has the form

$$\mathbb{M}^{\,\mathrm{sp}}(1;h) - \lambda(h)\mathbb{M}(1;0) ,$$

where $\mathbb{M}^{\,\mathrm{sp}}(\cdot;h)$ is a Gaussian process with the covariance

$$D^{(h)}(s,t) = d_0^{(h)}(s,t) + \sum_{j=1}^{\infty}\{d_j^{(h)}(s,t) + d_j^{(h)}(t,s)\} ,$$

with

$$d_0^{(h)}(s,t) = \lim_{n\to\infty} \frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\left(\frac{X_h}{X_0} \wedge 1\right)^{2\alpha} \mathbb{1}_{\{X_0 > (s\vee t)u_n\}}\right] = (s\vee t)^{-\alpha}\mathbb{E}\left[(\Theta_h \wedge 1)^{2\alpha}\right] ,$$

and for $j \geq 1$

$$\begin{aligned} d_j^{(h)}(s,t) &= \lim_{n\to\infty} \frac{1}{\bar{F}_X(u_n)} \mathbb{E}\left[\left(\frac{X_h}{X_0} \wedge 1\right)^{\alpha} \left(\frac{X_{j+h}}{X_j} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_0 > su_n\}} \mathbb{1}_{\{X_j > tu_n\}}\right] = \\ &= s^{-\alpha}\mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}(\Theta_{j+h}\Theta_j^{-1} \wedge 1)^{\alpha}\mathbb{1}_{\{Y_0\Theta_j > t/s\}}\right] \\ &= s^{-\alpha}\mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}(\Theta_{j+h}\Theta_j^{-1} \wedge 1)^{\alpha}(\Theta_j s/t \wedge 1)^{\alpha}\right] . \end{aligned}$$

Furthermore

$$\mathrm{Cov}(\mathbb{M}^{\,\mathrm{sp}}(s;h), \mathbb{M}(t;0)) = g_0(s,t) + \sum_{j=1}^{\infty}\{g_j(s,t) + h_j(s,t)\} ,$$

where

$$\begin{aligned} g_0(s,t) &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\frac{X_h}{X_0} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_0 > (s\vee t)u_n\}}\right]}{\bar{F}_X(u_n)} = (s\vee t)^{-\alpha}\mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}\right] , \\ g_j(s,t) &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\frac{X_h}{X_0} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_0 > su_n\}}\mathbb{1}_{\{X_j > tu_n\}}\right]}{\bar{F}_X(u_n)} , \\ h_j(s,t) &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\frac{X_{j+h}}{X_j} \wedge 1\right)^{\alpha} \mathbb{1}_{\{X_j > su_n\}}\mathbb{1}_{\{X_0 > tu_n\}}\right]}{\bar{F}_X(u_n)} , \end{aligned}$$

so that

$$\begin{aligned} g_j(1,1) &= \mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}\mathbb{1}_{\{Y_0\Theta_j > 1\}}\right] = \mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}(\Theta_j \wedge 1)^{\alpha}\right] , \\ h_j(1,1) &= \mathbb{E}\left[(\Theta_{j+h}\Theta_j^{-1} \wedge 1)^{\alpha}\mathbb{1}_{\{Y_0\Theta_j > 1\}}\right] = \mathbb{E}\left[(\Theta_{j+h}\Theta_j^{-1} \wedge 1)^{\alpha}(\Theta_j \wedge 1)^{\alpha}\right] . \end{aligned}$$

Recalling (5.3.14) (for $h = 0$) and combining calculations from this section we obtain

$$\begin{aligned} &\mathrm{Var}\left(\mathbb{M}^{\,\mathrm{sp}}(1;h) - \lambda(h)\mathbb{M}(1;0)\right) = \qquad (5.3.17) \\ &= \mathbb{E}\left[(\Theta_h \wedge 1)^{2\alpha}\right] + 2\sum_{j=1}^{\infty}\mathbb{E}\left[(\Theta_h \wedge 1)^{\alpha}(\Theta_{j+h}\Theta_j^{-1} \wedge 1)^{\alpha}(\Theta_j \wedge 1)^{\alpha}\right] \end{aligned}$$

$$+\lambda^2(h)\left\{1+2\sum_{j=1}^{\infty}\mathbb{E}[\Theta_j^\alpha\wedge 1]\right\}$$
$$-2\lambda(h)\left\{\mathbb{E}\left[\Theta_h^\alpha\wedge 1\right]+\sum_{j=1}^{\infty}\mathbb{E}\left[(\Theta_h\wedge 1)^\alpha(\Theta_j\wedge 1)^\alpha\right]+\sum_{j=1}^{\infty}\mathbb{E}\left[(\Theta_{j+h}\Theta_j^{-1}\wedge 1)^\alpha(\Theta_j\wedge 1)^\alpha\right]\right\}.$$

5.3.5 Examples

AR(1)

Consider AR(1) process $X_j=\rho X_{j-1}+\varepsilon_j$, where $\rho\in(0,1)$ and $\{\varepsilon_j\}$ are i.i.d., non-negative and regularly varying random variables. For this process we can calculate all the expressions of interest explicitly.

We have $\lambda(h)=\rho^{h\alpha}$, $\Theta_j=\rho^j$, $j\geq 0$. Thus

$$\begin{aligned}
&\mathrm{Var}\left(\mathbb{M}(1;h)-\lambda(h)\mathbb{M}(1;0)\right)=\\
&=\left[\rho^{h\alpha}+2\sum_{j=1}^{\infty}\rho^{(j+h)\alpha}\right]+\rho^{2h\alpha}\left\{1+2\sum_{j=1}^{\infty}\rho^{j\alpha}\right\}\\
&-2\rho^{h\alpha}\left\{\rho^{h\alpha}+\sum_{j=1}^{h}\rho^{h\alpha}+\sum_{j=h+1}^{\infty}\rho^{j\alpha}+\sum_{j=1}^{\infty}\rho^{(j+h)\alpha}\right\}\\
&=\left(\rho^{h\alpha}+\rho^{2h\alpha}\right)\left\{1+\frac{2}{1-\rho^\alpha}\right\}-2\rho^{2h\alpha}\left\{1+h+\frac{2}{1-\rho^\alpha}\right\}.
\end{aligned}\tag{5.3.18}$$

Furthermore, for the spectral estimator we have:

$$\begin{aligned}
&\mathrm{Var}\left(\mathbb{M}^{\,\mathrm{sp}}(1;h)-\lambda(h)\mathbb{M}(1;0)\right)=\\
&=\left[\rho^{2h\alpha}+2\sum_{j=1}^{\infty}\rho^{(j+2h)\alpha}\right]+\rho^{2h\alpha}\left\{1+2\sum_{j=1}^{\infty}\rho^{j\alpha}\right\}\\
&-2\rho^{h\alpha}\left\{\rho^{h\alpha}+\rho^{h\alpha}\sum_{j=1}^{\infty}\rho^{j\alpha}+\rho^{h\alpha}\sum_{j=1}^{\infty}\rho^{j\alpha}\right\}\\
&=2\rho^{2h\alpha}\left\{1+\frac{2}{1-\rho^\alpha}\right\}-2\rho^{2h\alpha}\left\{1+\frac{2}{1-\rho^\alpha}\right\}=0\,.
\end{aligned}$$

That is, the only contribution to the limiting variance comes from estimation of the tail index. Set $k=k_\alpha$. Let $\widehat{\alpha}_k$ be the corresponding Hill estimator. From Section 5.1.3

we know that the limiting distribution for the Hill estimator of $1/\alpha$ is asymptotically normal with mean zero and variance

$$\alpha^{-2}\left\{1+2\sum_{j=1}^{\infty}\mathbb{E}\left[(\Theta_j\wedge 1)^{\alpha}\right]\right\}=\alpha^{-2}\left\{1+\frac{2}{1-\rho^{\alpha}}\right\} .$$

The delta method and Theorem 5.3.5 yield the asymptotic variance of the spectral estimator with the estimated α to be

$$\alpha^4\left\{1+\frac{2}{1-\rho^{\alpha}}\right\}^{-2}\times(\rho^{h\alpha}h\log(\rho))^2 .$$

Heuristic for AR(1): We note that when X_j is large, then X_{j+h}/X_j behaves like ρ^h. Thus, the estimator in (5.3.7) behaves like $(1/k))\rho^{h\alpha}\sum_{j=1}^{n}\mathbb{1}_{\{X_j>X_{n:n-k}\}}=\rho^{h\alpha}$, Hence, the estimator *becomes deterministic*. The only randomness comes from estimation of α. Using the first order Taylor expansion, the asymptotic distribution of the spectral estimator is the same as that of

$$\sqrt{k}(\widehat{\alpha}-\alpha)\rho^{h\alpha}\log(\rho) .$$

This is a short *proof* in case of AR(1).

Stochastic Recurrence Equation (SRE)

Assume that (A_j, B_j), $j\in\mathbb{N}$, are i.i.d. random vectors of nonnegative random variables and define

$$X_{j+1}=A_{j+1}X_j+B_{j+1},\quad j\geq 0 .$$

The stationary solution exists whenever $-\infty\leq\mathbb{E}[\log A_1]<0$ and $\mathbb{E}[\log^+ B_1]$; see [2]. If there exists $\alpha>0$ such that $\mathbb{E}[A_1^{\alpha}]=1$, $\mathbb{E}[A_1^{\alpha}\log^+ A_1]<\infty$ and $\mathbb{E}[B_1^{\alpha}]<\infty$, then the stationary solution is regularly varying by [21] and [19]. Moreover, the sequence $\{X_j\}$ is regularly varying; see [2].

In particular, ARCH(1) is defined as

$$X_j=\sigma_j Z_j , \tag{5.3.19}$$

$$\sigma_j^2 = \beta + \lambda X_{j-1}^2 \,, \tag{5.3.20}$$

where Z_j are i.i.d. standard normal. The sequence $\{X_j^2\}$ can be written in terms of SRE with

$$A_j = \lambda Z_j^2 \,, \quad B_j = \beta Z_j^2 \,.$$

The stationary solution exists when $\lambda < 1$. Then $\Theta_j = \prod_{i=1}^j A_i$. Set $h = 1$. Note that $\Theta_{j+1}/\Theta_j = A_{j+1}$ is independent of Θ_i, $i \le j$. Furthermore, $\lambda(1) = \mathbb{E}[(A_1 \wedge 1)^\alpha]$. Using formula (5.3.17) we have

$$\begin{aligned}
&\mathrm{Var}\,(\mathbb{M}^{\,\mathrm{sp}}(1;1) - \lambda(1)\mathbb{M}(1;0)) = \\
&= \mathbb{E}\left[(A_1 \wedge 1)^{2\alpha}\right] + 2\mathbb{E}\left[(A_1 \wedge 1)^\alpha\right] \sum_{j=1}^\infty \mathbb{E}\left[(\Theta_1 \wedge 1)^\alpha (\Theta_j \wedge 1)^\alpha\right] \\
&+ \lambda^2(1) \left\{ 1 + 2 \sum_{j=1}^\infty \mathbb{E}[\Theta_j^\alpha \wedge 1] \right\} \\
&- 2\lambda(1) \left\{ \mathbb{E}\left[A_1^\alpha \wedge 1\right] + \sum_{j=1}^\infty \mathbb{E}\left[(A_1 \wedge 1)^\alpha (\Theta_j \wedge 1)^\alpha\right] + \mathbb{E}\left[(A_1 \wedge 1)^\alpha\right] \sum_{j=1}^\infty \mathbb{E}\left[(\Theta_j \wedge 1)^\alpha\right] \right\} . \\
&= \mathbb{E}\left[(A_1 \wedge 1)^{2\alpha}\right] + 2\lambda(1) \left\{ \mathbb{E}\left[(A_1 \wedge 1)^{2\alpha}\right] + \sum_{j=2}^\infty \mathbb{E}\left[(\Theta_1 \wedge 1)^\alpha (\Theta_j \wedge 1)^\alpha\right] \right\} \\
&+ \lambda^2(1) \left\{ 1 + 2 \sum_{j=1}^\infty \mathbb{E}[\Theta_j^\alpha \wedge 1] \right\} \\
&- 2\lambda(1) \left\{ \lambda(1) + \mathbb{E}\left[(A_1 \wedge 1)^{2\alpha}\right] + \sum_{j=2}^\infty \mathbb{E}\left[(A_1 \wedge 1)^\alpha (\Theta_j \wedge 1)^\alpha\right] + \lambda(1) \sum_{j=1}^\infty \mathbb{E}\left[(\Theta_j \wedge 1)^\alpha\right] \right\} . \\
&= \mathbb{E}\left[(A_1 \wedge 1)^{2\alpha}\right] - \mathbb{E}^2\left[(A_1 \wedge 1)^{2\alpha}\right] = \mathrm{Var}((A_1 \wedge 1)^\alpha) \,.
\end{aligned} \tag{5.3.21}$$

Since $\Theta_1 = A_1$, we can estimate the last quantity by calculating the sample variance of $U_{j,n}$, where

$$U_{j,k} = ((X_{j+1}/X_j)^\alpha \wedge 1)\mathbb{1}_{\{X_j > X_{n:n-k}\}} \,, \quad j = 1, \ldots, n-1 \,. \tag{5.3.22}$$

Of course, α has to be estimated in practice.

<u>Heuristic for SRE:</u> When X_j is large, then random variables (X_{j+1}/X_j), $j \geq 1$, are approximately $\Theta_j = A_j$ and hence independent. Thus, the only contribution to the limiting variance is $\mathrm{Var}(\Theta_1^\alpha \wedge 1) = \mathrm{Var}(A_1^\alpha \wedge 1)$. The case of $h \geq 2$ is more involved, since the random variables (X_{j+h}/X_j), $j \geq 1$, (when X_j is large) are h-dependent.

A similar discussion is valid for Markov chains that have multiplicative structure of the tail process.

5.3.6 Implementation. Simulation studies

We perform numerical studies on estimation of the tail dependence coefficient. We consider the ARCH(1) process defined in (5.3.19)-(5.3.20) with $\beta = 0.1$, $\lambda = 0.7$ and Z_1 being standard normal. According to Table 8.4.8 in [18] this set of parameters yields the tail index of X_1^2 to be 1.59. The exact value of the tail dependence coefficient at lag 1 is $\mathrm{E}[(A_1 \wedge 1)^\alpha]$, where $A_1 = \lambda Z_1^2$. This value is estimated by the Monte Carlo simulation based on 1000 replicates.

First, we perform the Monte Carlo experiment (with 1000 repetitions), where we simulate 1000 observations from the ARCH(1) process and estimate the tail dependence coefficient using the empirical estimator (5.3.11) and the spectral estimator (5.3.16). Both estimators are based on 5%, 10%, 15% and 20% of the largest observations. The tail index in (5.3.16) is estimated using the Hill estimator with the corresponding number of order statistics. The results are depicted in Figure 5.11. Along with the boxplot, we summarize the results by calculating the interquartile ranges and median absolute deviations of the estimates as shown in Tables 5.1 and 5.2, respectively. We see reduction in the variability, similar to Section 5.2.

Finally, the full applicability of the spectral method is illustrated on Figure 5.12, where we depict the Hill plot for the spectral estimator along with the confidence

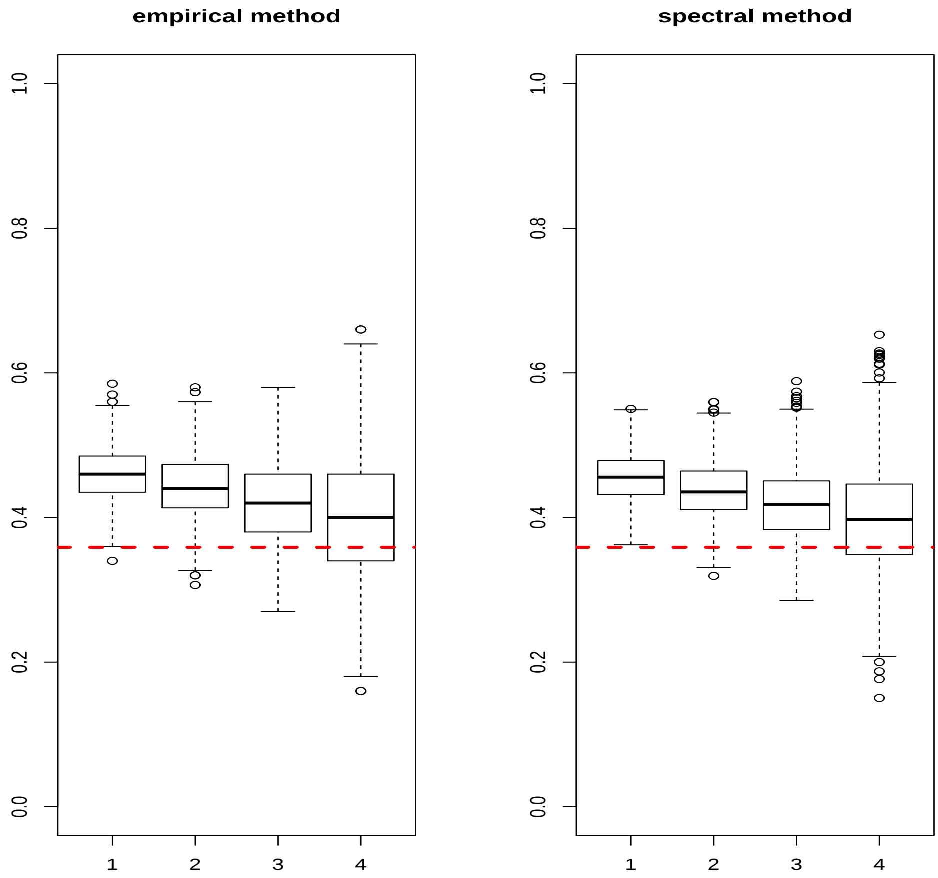

Figure 5.11: Boxplots for estimators for the tail dependence coefficient: empirical estimator (5.3.11) - (left panel) and spectral estimator (5.3.16) - (right panel). From the left to the right: 20%, 15%, 10% and 5% of order statistics used.

k	Empirical	Spectral	Improvement
20%	0.05	0.047	6.20%
15%	0.06	0.053	11.05%
10%	0.08	0.067	15.85%
5%	0.12	0.097	19.01%

Table 5.1: Values of the interquartile range for the estimates of the tail dependence coefficient.

k	Empirical	Spectral	Improvement
20%	0.025	0.023	5.78%
15%	0.033	0.027	17.42%
10%	0.040	0.034	15.34%
5%	0.06	0.048	19.10%

Table 5.2: Values of the median absolute deviation for the estimates of the tail dependence coefficient.

interval obtained by the method described in (5.3.22), again with α replaced with its Hill estimator with the corresponding number or order statistics. We also plot the empirical estimator, however its confidence interval requires some resampling scheme.

5.3.7 Data analysis

We analyse daily stock price for Exxon Mobil. from January 4, 2010 to November 3, 2015.[2] We calculate squared log-returns and estimate the tail dependence coefficients for lags 1, 2 and 3. Interestingly, lags 2 and 3 have stronger tail dependence; see thick blue lines on Figure 5.13. The Hill plots for lags 2 and 3 stabilize around the value of 0.2.

Do the data follow GARCH(1,1) model? It is very often assumed in the econometrics literature that financial time series follow a GARCH(1,1) model. We fitted a GARCH(1,1) model $X_j = \mu + \sigma_j Z_j$, $\sigma_j^2 = \omega + \alpha_1 X_{j-1}^2 + \beta_1 \sigma_{j-1}^2$ to the log-returns of the Exxon Mobil data. The estimated parameters are $\mu = -3.601310e-04$, $\omega = 3.651305e-06$, $\alpha_1 = 7.858571e-02$ and $\beta_1 = 8.960275e-01$. Hence, $\alpha_1+\beta_1 < 1$ yielding stationarity. The Jarque-Bera test rejects normality of the residuals. We fit-

[2]Source: https://ca.finance.yahoo.com

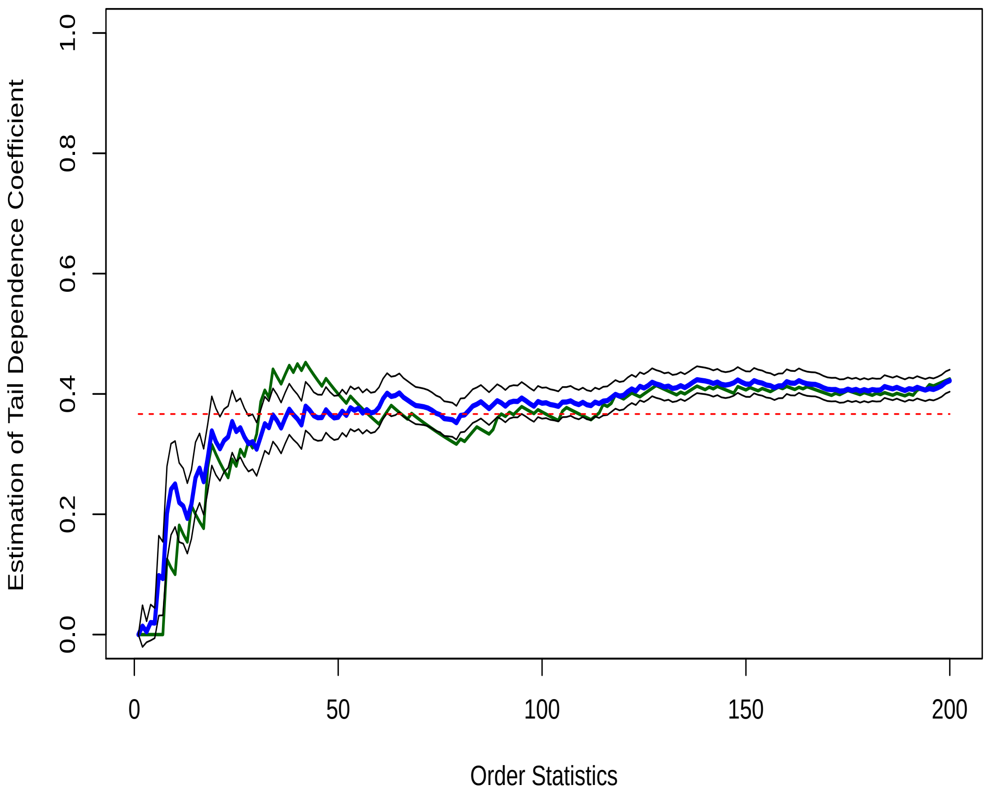

Figure 5.12: Estimation of the Tail Dependence Coefficient: thick solid blue line - the spectral estimator; thin solid black lines - confidence interval; thick solid green line - the empirical estimator.

ted t-distribution with 6 degrees of freedom.

Then, we simulated 100 GARCH(1,1) processes with the estimated parameters, of length 1468 (the same as the Exxon Mobil series) and produced the Hill plots for the tail dependence coefficients of the squares of the simulated GARCH at lags 1, 2, 3 (see Figure 5.13). One can argue that GARCH(1,1) methodology overestimates the extremal dependence for the log-returns at lag 1, but captures properly the extremal dependence at lags 2 and 3.

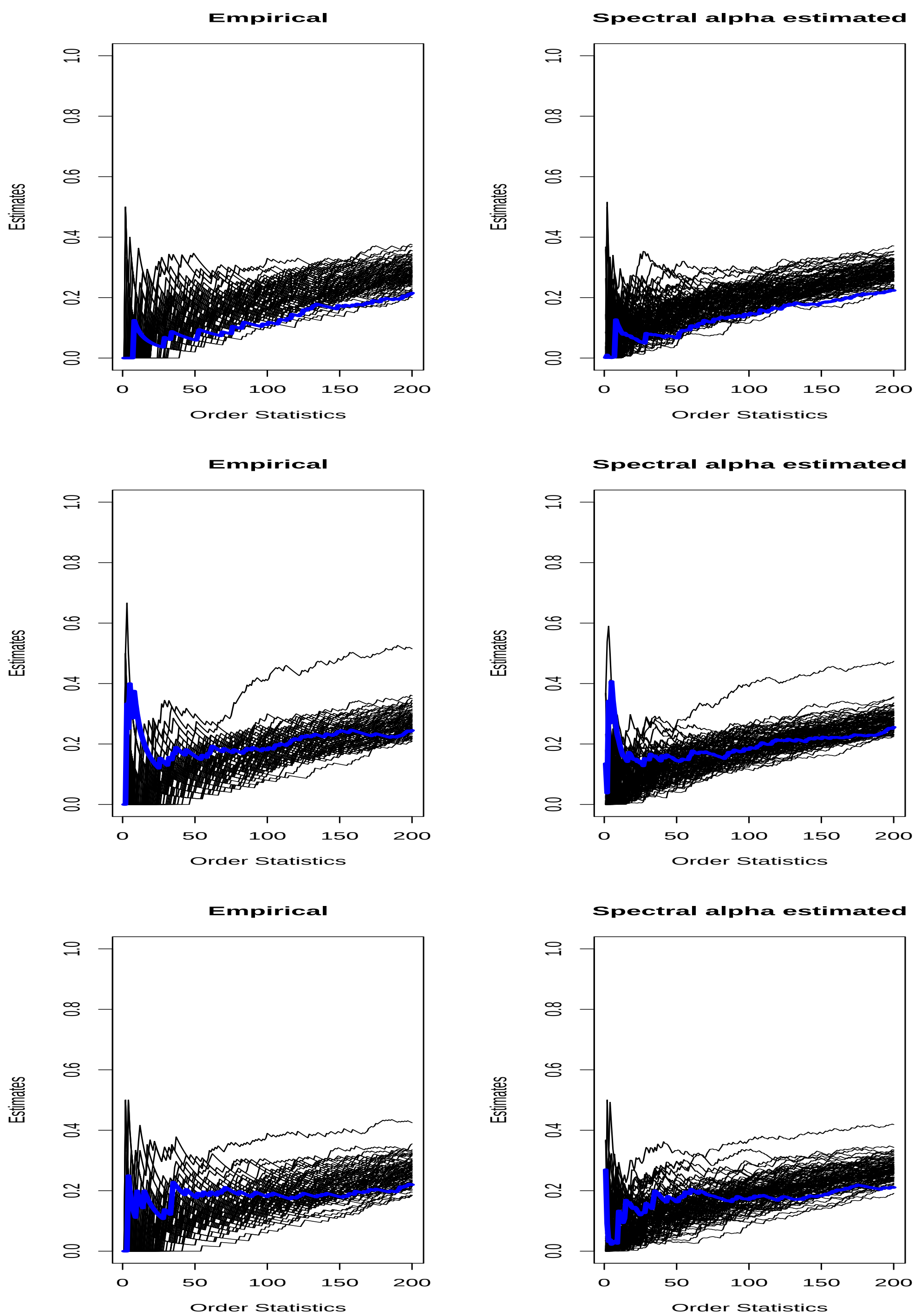

Figure 5.13: Estimation of the Tail Dependence Coefficient for GARCH(1,1) model; lag 1 - top panel, lag 2 - middle panel, lag 3 - bottom panel. Black thin lines - estimates for 100 realizations. Blue thick line - nonparametric estimate using the empirical (left panel) and spectral estimator (right panel).

5.4 Estimation of the Marginal Expected Shortfall under extremal independence

Recall the conditional extreme value model from Section 2.4. Our goal is to estimate the Marginal Expected Shortfall defined in (2.4.4):

$$\theta(p) := \mathbb{E}[Y \mid X > Q_X(p)] \, . \tag{5.4.1}$$

We proceed as follows.

In Section 5.4.1 we construct the estimators of $\theta(p)$; see equations (5.4.7) and (5.4.8). We prove the consistency and asymptotic normality of the estimator of the scaling function b (see Proposition 5.4.1). Then, we discuss estimation of ϕ, the index of regular variation of b. The main result is stated in Theorem 5.4.4, where we prove the asymptotic normality and consistency of the estimator of $\theta(p)$. The section is concluded with several comments (comparison with [8], discussion of assumptions, etc.). Section 5.4.2 contains a number of examples which indicate that our approach is, in general, more favorable than the one from [8]. We continue with a simulation study and data analysis. We conclude with proofs. The material of this section is based on author's original contribution, see [27].

5.4.1 Asymptotic normality of MES estimators

Assumption 1 leads to the following estimation procedure for $\theta(p)$ defined in (5.4.1). Since the tail distribution $\bar{F}_X$ of X is regularly varying with the index $-\alpha$, the function $Q_X(p) = F_X^{\leftarrow}(1-p)$ is regularly varying at 0 with index $-1/\alpha$. Let $k = k_n \to \infty$ be a sequence of integers such that $k/n \to 0$ and let $p = p_n$ be such that $p_n \to 0$ as $n \to \infty$. Assume that

$$\lim_{n\to\infty} \frac{k_n}{np_n} = r \in (0, \infty) \, .$$

Hence, by Lemma 2.1.3 along with Remark 2.1.4 applied with $f = Q_X$, when $n \to \infty$,

$$\frac{Q_X(p_n)}{Q_X(k_n/n)} \sim \left(\frac{k_n}{np_n}\right)^{1/\alpha} \sim r^{1/\alpha} . \tag{5.4.2}$$

Recall that $k_n = n\bar{F}_X(u_n)$. If F_X is continuous and strictly increasing, then $u_n = F_X^{\leftarrow}(1 - k_n/n) = Q_X(k_n/n)$.

Noting that

$$\theta(p_n) = \mathbb{E}[Y \mid X > Q_X(p_n)] , \quad \theta(k_n/n) = \mathbb{E}[Y \mid X > Q_X(k_n/n)] = \mathbb{E}[Y \mid X > u_n] ,$$

Lemma 2.4.3 implies that

$$\lim_{n\to\infty} \frac{\theta(p_n)}{b\left(Q_X(p_n)\right)} = \lim_{n\to\infty} \frac{\theta(k_n/n)}{b\left(u_n\right)} = \aleph_{\mathrm{CTE}}(1) \in (0, \infty) . \tag{5.4.3}$$

W.l.o.g. we shall assume that $\aleph_{\mathrm{CTE}}(1) = 1$. The latter expression, regular variation of the function b and Lemma 2.1.3 applied with $f = b$ imply that

$$\theta(p_n) \sim b\left(Q_X(p_n)\right) \sim b(Q_X(k_n/n)) \left(\frac{k_n}{np_n}\right)^{\phi/\alpha} = b(u_n) \left(\frac{k_n}{np_n}\right)^{\phi/\alpha} . \tag{5.4.4}$$

Combining (5.4.3)-(5.4.4) we get

$$\theta(p_n) \sim \theta(k_n/n) \left(\frac{k_n}{np_n}\right)^{\phi/\alpha} \sim \theta(k_n/n) r^{\phi/\alpha} , \qquad n \to \infty . \tag{5.4.5}$$

From now on, we will omit the subscript n in k_n and p_n.

Assume that we have an i.i.d. sample (X_j, Y_j) from (X, Y). We can estimate $Q_X(k/n)$ by $X_{n:n-k}$, where $X_{n:n} \geq X_{n:n-1} \geq \cdots \geq X_{n:1}$ are the order statistics from $X_1, \ldots, X_n$. Hence $\theta(k/n)$ can be estimated by

$$\widehat{\theta}(k/n) = \frac{1}{k} \sum_{j=1}^{n} Y_j \mathbb{1}_{\{X_j > X_{n:n-k}\}} . \tag{5.4.6}$$

This leads to the following estimators $\theta(p)$

$$\widetilde{\theta}_n(p) = \widehat{\theta}(k/n) \left(\frac{k}{np}\right)^{\phi/\alpha} , \tag{5.4.7}$$

when α and ϕ are known and

$$\widehat{\theta}_n(p) = \widehat{\theta}(k/n) \left(\frac{k}{np}\right)^{\widehat{\phi}/\widehat{\alpha}} , \tag{5.4.8}$$

when α and ϕ have to be estimated.

The goal is to prove consistency and central limit theorem for these estimators. We proceed as follows:

- In Proposition 5.4.1 we state consistency and the central limit theorem for $\widehat{\theta}(k/n)$. The proof uses the techniques based on tail empirical processes developed in Section 4.4. In due course we introduce several assumptions.
- Then we discuss how to estimate α and ϕ.
- In Theorem 5.4.4 we finally state the central limit theorem for the estimators.
- Next, we discuss the assumptions and compare our approach to [8].

Estimation of the scaling function

Assume that we have an i.i.d. sample (X_j, Y_j), $j = 1, \ldots, n$, from the distribution of (X, Y). Recall that k is a sequence of integers such that $k \to \infty$, $k/n \to 0$. Let u_n be the sequence defined by $u_n = Q_X(k/n)$, that is k may be chosen as $k = n\bar{F}_X(u_n)$. We note that $u_n \to \infty$ and $n\bar{F}_X(u_n) \to \infty$. Also, from Section 5.1.1 we have

$$\frac{X_{n:n-k}}{u_n} \xrightarrow{\mathrm{P}} 1 .$$

Hence, $\widehat{\theta}(k/n)$ defined in (5.4.6) can be viewed as an estimate of $\theta(k/n)$, which in turn can be approximated by $\aleph_{\mathrm{CTE}}(1) \times b(u_n)$; see (2.4.5). To state the limit theorem for $\widehat{\theta}(k/n)$ we need additional moment, continuity and bias assumptions. The conditions are stated in the language of u_n but can be re-phrased using k. To proceed, recall that (cf. (4.4.1))

$$T_n(s,t) = \frac{\mathbb{P}(X > u_n s \ , \ Y \le b(u_n)t)}{\mathbb{P}(X > u_n)} \ , \quad T(s,t) = \lim_{n\to\infty} T_n(s,t) \ .$$

For **consistency** we need:

Assumption 3 *Let $s_0 \in (0,1)$. The function $T(\cdot,\cdot)$ is continuous on $[s_0,\infty) \times \mathbb{R}_+$.*

Note that it appeared already as Assumption 2, before Proposition 4.4.1.

Assumption 4 *For $q = 1, 2$ there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} \mathbb{E}\left[\frac{Y^{q+\delta}}{b^{q+\delta}(u_n)} \mid X > u_n\right] < \infty \ .$$

In fact for consistency one needs the latter assumption with $q = 1$ only, while $q = 2$ is needed for the asymptotic normality.

For **asymptotic normality** we also need the following no-bias conditions. To introduce the first one, recall the second order assumption, cf. Definition 5.1.1 and inequality (5.1.4).

Assumption 5 *We assume that $\bar{F}_X \in 2RV(-\alpha, \beta)$ and for some $\delta > 0$,*

$$\lim_{n\to\infty} \sqrt{n\bar{F}_X(u_n)}\, u_n^{-\alpha\beta+\delta} = 0 \ . \tag{5.4.9}$$

Assumption 6 *We assume that*

$$\lim_{n\to\infty} \sqrt{n\bar{F}_X(u_n)} \sup_{s\ge s_0} \left(\int_s^\infty \int_0^\infty yT_n(\mathrm{d}x,\mathrm{d}y) - \int_s^\infty \int_0^\infty yT(\mathrm{d}x,\mathrm{d}y)\right) = 0 \ .$$

We are ready to state the following result. The proof is given in Section 5.4.5.

Proposition 5.4.1 *Let Assumptions 1, 3, 4 (with $q = 1$) hold. Then*

$$\frac{\widehat{\theta}(k/n)}{b(u_n)} \xrightarrow{p} \aleph_{\mathrm{CTE}}(1) \ .$$

If moreover Assumptions 4 with $q = 2$, 5 and 6 hold then

$$\sqrt{k}\left(\frac{\widehat{\theta}(k/n)}{b(u_n)} - \aleph_{\mathrm{CTE}}(1)\right) \xrightarrow{\mathrm{d}} Z_1 + Z_2 \ ,$$

where

$$Z_1 = \int_1^\infty \int_0^\infty y\mathbb{T}(dx, dy)\;, \qquad Z_2 = \alpha^{-1}\mathbb{T}(1,\infty)\int_0^\infty yT(1, \mathrm{d}y)\;,$$

and $\mathbb{T}$ *is a Gaussian process defined on* $[s_0, \infty) \times (0, \infty)$ *with the covariance function* $T(s \vee s', t \wedge t')$.

Remark 5.4.2 The limiting variable is the sum of two dependent normal random variables with mean zero. The formulas for $\mathrm{Var}(Z_1), \mathrm{Var}(Z_2)$ as well as $\mathrm{Cov}(Z_1, Z_2)$ involve the unknown function T and hence are not particularly useful. The limiting variance can be in principle estimated by resampling techniques, but this is not addressed here.

Remark 5.4.3 Again, note that Assumptions 5 and 6 hold for some values of u_n (or, equivalently, k).

Estimation of the tail index and the conditional scaling exponent

The tail index α of X is estimated by the standard Hill estimator; see Section 5.1.3:

$$\frac{1}{\widehat{\alpha}} = \frac{1}{k}\sum_{j=1}^{n} \log\left(\frac{X_{n:n-j+1}}{X_{n:n-k}}\right)\;.$$

Under the conditions stated in Section 5.1.3, the estimator is asymptotically normal with mean zero and variance $1/\alpha^2$. To estimate ϕ we note that in [24] the authors showed that under Assumption 1 and additional technical conditions, the random variable XY is regularly varying with index $\zeta := \alpha/(1+\phi)$. Thus, we can estimate ζ by $\widehat{\zeta}$, the Hill estimator based on the order statistics of X_jY_j, $j = 1, \ldots, n$. Again, regular variation of the product implies consistency of $\widehat{\zeta}$. If moreover the second order condition holds for XY then we also have

$$\sqrt{k}\left\{\frac{1}{\widehat{\zeta}} - \frac{1}{\zeta}\right\} \xrightarrow{\mathrm{d}} N\left(0, \frac{1}{\zeta^2}\right)\;.$$

However, the dependence structure of $(\widehat{\alpha}, \widehat{\zeta})$ is rather complicated. As such one can conjecture that the limiting distribution of

$$\widehat{\phi} = \widehat{\alpha}/\widehat{\zeta} - 1 ,$$

say, Δ, is normal but the question about its variance remains open.

Nevertheless, this suggests the following estimation procedure for ϕ.

- Estimate the tail index α of X using e.g. the Hill estimator $\widehat{\alpha}$;
- Let $\zeta = \alpha/(1+\phi)$. Estimate ζ using the Hill estimator $\widehat{\zeta}$ based on XY;
- Estimate ϕ by $\widehat{\phi} = \widehat{\alpha}/\widehat{\zeta} - 1$.

Estimation of the expected shortfall

We combine the above results into the following theorem. To state this we need another "no-bias" assumption.

Assumption 7 *We have*

$$\lim_{n\to\infty} \sqrt{k} \left\{ \frac{\aleph_{\mathrm{CTE}}(1) \times b(u_n)}{\theta(p)} \left(\frac{k}{np}\right)^{\phi/\alpha} - 1 \right\} = 0 .$$

We note that this assumption is needed for the asymptotic normality only, since the expression in the bracket converges to zero under CEV and the moment assumptions. This follows from (5.4.3) and (5.4.5). Hence, Assumption 7 holds for *some* sequence k, but the valid range of k has to be justified for each model.

Theorem 5.4.4 *Let Assumptions 1, 3, 4 (with $q = 1$) hold. Then*

$$\frac{\widetilde{\theta}_n(p)}{\theta(p)} \xrightarrow{p} 1 .$$

If moreover Assumptions 4 with $q = 2$, 5, 6 and 7 hold, then

$$\sqrt{k}\left\{\frac{\widetilde{\theta}_n(p)}{\theta(p)} - 1\right\} \xrightarrow{\mathrm{d}} (\aleph_{\mathrm{CTE}}(1))^{-1}(Z_1 + Z_2) \tag{5.4.10}$$

and

$$\sqrt{k}\left\{\frac{\widehat{\theta}_n(p)}{\theta(p)} - 1\right\} \xrightarrow{\mathrm{d}} (\aleph_{\mathrm{CTE}}(1))^{-1}(Z_1 + Z_2) + \Delta\ ,$$

where Δ is a distributional limit of $\sqrt{k}\log\left(\frac{k}{np}\right)\left(\widehat{\beta} - \beta\right)$, $\beta = \phi/\alpha$, $\widehat{\beta} = \widehat{\phi}/\widehat{\alpha}$ and $\widehat{\phi}, \widehat{\alpha}$ are estimators of ϕ, α, respectively.

Proof: We have

$$\sqrt{k}\left\{\frac{\widetilde{\theta}_n(p)}{\theta(p)} - 1\right\} = \sqrt{k}(\aleph_{\mathrm{CTE}}(1))^{-1}\left\{\frac{\widehat{\theta}(k/n)}{b(u_n)} - \aleph_{\mathrm{CTE}}(1)\right\}$$
$$+ \frac{\widehat{\theta}(k/n)}{\aleph_{\mathrm{CTE}}(1) \times b(u_n)}\sqrt{k}\left\{\frac{\aleph_{\mathrm{CTE}}(1) \times b(u_n)}{\theta(p)}\left(\frac{k}{np}\right)^{\phi/\alpha} - 1\right\}.$$

The first expression converges in distribution to $(\aleph_{\mathrm{CTE}}(1))^{-1}(Z_1 + Z_2)$ by Proposition 5.4.1, while the second one converges in probability to zero by Assumption 7 and since

$$\widehat{\theta}(k/n)/(\aleph_{\mathrm{CTE}}(1) \times b(u_n)) \xrightarrow{\mathrm{P}} 1$$

again by Proposition 5.4.1.

Furthermore, set $\beta = \phi/\alpha$ and $\widehat{\beta}$ is an estimator of β. Then

$$\sqrt{k}\left\{\frac{\widehat{\theta}_n(p)}{\theta(p)} - 1\right\} = \sqrt{k}\left\{\frac{\widehat{\theta}(k/n)}{\theta(p)}\left(\frac{k}{np}\right)^{\widehat{\beta}} - 1\right\}$$
$$= \sqrt{k}\left\{\frac{\widehat{\theta}(k/n)}{\theta(p)}\left(\frac{k}{np}\right)^{\beta} - 1\right\} + \frac{\widehat{\theta}(k/n)}{\theta(k/n)}\sqrt{k}\frac{\theta(k/n)}{\theta(p)}\left\{\left(\frac{k}{np}\right)^{\widehat{\beta}} - \left(\frac{k}{np}\right)^{\beta}\right\}.$$

We have already proven that the first part converges to $(\aleph_{\mathrm{CTE}}(1))^{-1}(Z_1 + Z_2)$ by (5.4.10). Since $\widehat{\theta}(k/n)/\theta(k/n) \xrightarrow{\mathrm{P}} 1$ we can ignore it in further consideration of the second part. We apply the second order Taylor expansion to the such modified second part to obtain:

$$\underbrace{\sqrt{k}\log\left(\frac{k}{np}\right)(\widehat{\beta}-\beta)}_{\stackrel{\mathrm{d}}{\to}\Delta}\frac{\theta(k/n)}{\theta(p)}\left(\frac{k}{np}\right)^{\beta}+\underbrace{\frac{1}{2}\sqrt{k}\log^2\left(\frac{k}{np}\right)(\widehat{\beta}-\beta)^2}_{=o_P(1)}\frac{\theta(k/n)}{\theta(p)}\left(\frac{k}{np}\right)^{\beta_n},$$

where $\beta_n=\beta+\lambda(\widehat{\beta}-\beta)$ and λ is a (random) value between 0 and 1. By (5.4.5) the first part converges in distribution to Δ, while the second term converges in probability to zero.

∎

Assumptions: discussion

We discuss Assumptions 3, 4, 5, 6 and 7.

- Assumption 3 is a relatively mild continuity condition. In the trivial case of X and Y being independent, we have $T(s,t)=s^{-\alpha}\mathbb{P}(Y\le t)$ and it clearly suffices that the distribution function F_Y of Y is continuous.

  For more complicated models, the assumption will be verified directly.

- Assumption 4 implies, by Lemma 2.4.3, that the limiting conditional expectation is well-defined. It is also needed for validity of the central limit theorem. If X and Y are independent, then it suffices that $\mathbb{E}[Y^{2+\delta}]<\infty$. If $X=Y$ then, bearing in mind that X is regularly varying with index α, one needs $\alpha>2$.

  Otherwise, we will verify this condition for each model separately.

- Assumption 5 follows from the classical second order regular variation as indicated in Section 5.1.

- Assumptions 6 is the ad-hoc condition to deal with the bias in the bivariate tail empirical process. It is used only at the beginning of the proof of Proposition 5.4.1 and can be replaced by either

$$\lim_{n\to\infty} \sqrt{n\bar{F}_X(u_n)} \sup_{s>s_0,t>0} |T_n(s,t) - T(s,t)| = 0$$

  or uniform convergence of $(\mathrm{d}/\mathrm{d}s)T_n(s,t)$. We believe that Assumption 6 is easier to verify than the latter two. We verify it for specific models.

- Assumption 7 is another ad-hoc assumptions to deal with a different source of bias. Again, we will verify it for some specific models in Section 5.4.2.

Marginal transformation

The Marginal Expected Shortfall is invariant under a transformation of X. As such, we should obtain the same estimator for the original situation and for the transformed one. The next example shows that this is the case.

Example 5.4.5 Take $f(x) = x^\gamma$, $\gamma \in (0,\infty)$. If $b(u)$ is regularly varying with index ϕ, then $b \circ f^{\leftarrow}$ is regularly varying with index $\widetilde{\phi} = \alpha/\gamma$, while the marginal distribution of $\widetilde{X}$ is regularly varying with index α/γ. It is clear that the estimator (5.4.8) remains the same for both models.

Comparison with [8]

Let us focus first on the bivariate regularly varying random vector (X,Y). To illustrate the difference in terms of applicability between bivariate regular variation, the conditional extreme value assumption and the approach from [8], let us consider the following simple toy model. Let X be regularly varying random variable with index $-\alpha$, $\alpha > 1$. For $\phi \in (0,1)$, define $Y = X^\phi$. Then (X,Y) is bivariate regularly varying with the exponent measure concentrated on the horizontal axis. Hence (X,Y) is extremally independent and the limit in (2.2.11) vanishes. The transformed model $(X, Y^{1/\phi})$ is obviously extremally dependent and $R \not\equiv 0$, where R is defined in (2.2.12). Hence, [8] is applicable whenever the bivariate regularly varying vector with (possibly) extremal independence can be transformed to extremally dependent vector. For example, in [8] the authors consider a vector $(X, Y) = (|U|^{2/5}, |V|)$, where (U, V) is standard Cauchy distribution on $\mathbb{R}^2$. We note that (X, Y) is bivariate regularly varying with extremal independence, but it can be transformed to $(|U|, |V|)$ which is extremally dependent.

However, [8] is not applicable when $R \equiv 0$ and we will provide several examples in Section 5.4.2 below.

When we drop the bivariate regular variation assumption, we have the following difference in terms of applicability:

- Our approach: (X, Y) fulfills CEV assumption (Assumption 1). This implies that X is regularly varying with index $\alpha > 0$.
- [8]: (X, Y) fulfills (2.2.12) and Y is regularly varying with index $\gamma > 1$. There is no assumption on the marginal behaviour of X.

In terms of technical assumptions imposed in the main theorems, the main difference is that for the asymptotic normality we impose Assumption 7.

5.4.2 Examples

In what follows we give some examples when our theory applies, followed by the case when our approach does not work. The first example is the generic model for CEV (See Example 2.4.5) and we provide all details for Assumptions 1, 3-7. The second one is a continuation of Example 2.4.6. Next, we verify the conditions for consistency in case of Example 2.4.7. In all these cases we argue that the approach of [8] is not applicable.

This is followed by an example that can be handled by both [8] and our approach, however, one can argue that the former approach is favourable. It is a continuation of Example 2.4.8.

Example 5.4.6 Let (X, Y) be as in Example 2.4.5, that is X, V are independent regularly varying random variables with index $\alpha > 1$ and for $\phi \in (0, 1)$, $Y = X^{\phi} V$. The pair (X, Y) is bivariate regularly varying with index α and the exponent measure concentrated on both axes and hence extremally independent. Also, $R \equiv 0$ and the approach from [8] is not applicable.

We have already checked Assumption 1 with the limiting measure

$$\boldsymbol{\mu}((s, \infty] \times [0, t]) = s^{-\alpha} \left(1 - \int_1^\infty \mathbb{P}(V > z^{-\phi} t / s^{\phi}) \alpha z^{-\alpha-1} \mathrm{d}z\right) .$$

From the above formula we conclude also that Assumption 3 is fulfilled whenever the distribution function of V is continuous.

As for Assumption 4, we have by the regular variation of X,

$$\begin{aligned}\lim_{u \to \infty} \mathbb{E}\left[\left(\frac{Y}{u^{\phi}}\right)^{2+\delta} \mid X > u\right] &= \lim_{u \to \infty} \mathbb{E}\left[\left(\frac{X^{\phi} V}{u^{\phi}}\right)^{2+\delta} \mid X > u\right] \\ &= \mathbb{E}[V^{2+\delta}] \int_1^\infty z^{\phi(2+\delta)} \alpha z^{-\alpha-1} \mathrm{d}z \; . \end{aligned} \tag{5.4.11}$$

The above integral is finite if $\alpha > \phi(2 + \delta)$ (and hence we also recover the necessary condition in Lemma 2.4.3).

Finally, for Assumptions 6 and 7, let us assume in this example that X is Pareto, that is $\mathbb{P}(X > x) = x^{-\alpha}$, $x > 1$. We do not make any additional assumption on V, except of the aforementioned regular variation with $\alpha > 1$. Then, for any $x > 1$,

$$\mathbb{E}[Y \mid X > u] = \mathbb{E}[V]\mathbb{E}[X^{\phi} \mid X > u] = \mathbb{E}[V] \frac{\alpha}{\alpha - \phi} u^{\phi} = \frac{\alpha^2}{(\alpha - 1)(\alpha - \phi)} u^{\phi} .$$

From this we can conclude

$$\aleph_{\mathrm{CTE}}(1) = \lim_{u \to \infty} \mathbb{E}\left[\frac{Y}{u^{\phi}} \mid X > u\right] = \frac{\alpha^2}{(\alpha - 1)(\alpha - \phi)} .$$

Using the conditional expectation formula we observe that the expressions in Assumptions 6 and 7 vanish.

Now, we discuss Assumption 6 in the general case of regularly varying X. Recall that Assumption 6 states that

$$\lim_{n\to\infty}\sqrt{n\bar{F}_X(u_n)}\sup_{s\geq s_0}\left(\int_s^\infty\int_0^\infty yT_n(\mathrm{d}x,\mathrm{d}y)-\int_s^\infty\int_0^\infty yT(\mathrm{d}x,\mathrm{d}y)\right)=0\ .$$

Integration by parts yields

$$\int_s^\infty\int_0^\infty yT_n(\mathrm{d}x,\mathrm{d}y)=\frac{1}{\bar{F}_X(u_n)}\mathbb{E}\left[\frac{Y}{b(u_n)}\mathbb{1}_{\{X>u_ns\}}\right]=\mathbb{E}[V]\frac{1}{\bar{F}_X(u_n)}\mathbb{E}\left[\frac{X^\phi}{u_n^\phi}\mathbb{1}_{\{X>u_ns\}}\right]$$
$$=\mathbb{E}[V]\frac{1}{\bar{F}_X(u_n)}\int_{u_ns}^\infty(x/u_n)^\phi F_X(\mathrm{d}x)$$
$$=\mathbb{E}[V]\left\{s^\phi\frac{\bar{F}_X(u_ns)}{\bar{F}_X(u_n)}+\phi s^\phi\frac{1}{\bar{F}_X(u_n)}\int_1^\infty v^{\phi-1}\bar{F}_X(vsu_n)\mathrm{d}v\right\}\ .$$

The latter expression converges to

$$\mathbb{E}[V]\left\{s^{\phi-\alpha}+s^{\phi-\alpha}\frac{\phi}{\alpha-\phi}\right\}\ .$$

Hence, Assumption 6 is fulfilled if

$$\lim_{n\to\infty}\sqrt{n\bar{F}_X(u_n)}\sup_{s>s_0}\left|s^\phi\frac{\bar{F}_X(u_ns)}{\bar{F}_X(u_n)}-s^{\phi-\alpha}\right|=0$$

and

$$\lim_{n\to\infty}\sqrt{n\bar{F}_X(u_n)}\sup_{s>s_0}s^\phi\int_1^\infty v^{\phi-1}\left|\frac{\bar{F}_X(vsu_n)}{\bar{F}_X(u_n)}-(vs)^{-\alpha}\right|\mathrm{d}v=0\ .$$

The first condition is implied by Assumption 5, while the second one we bound using Lemma 5.1.2 by

$$\sqrt{n\bar{F}_X(u_n)}u_n^{-\alpha\beta+\delta}\sup_{s>s_0}s^{\phi-\alpha(\beta+1)+\epsilon}\int_1^\infty v^{-1-(\alpha(\beta+1)+\epsilon-\phi)}\mathrm{d}v\ .$$

The integral and the above $\sup_s$ are finite. In summary, if Assumption 5 holds then Assumption 6 is fulfilled as well.

We summarize this example as follows:

- Assumption 1 is fulfilled with $b(u) = u^\phi$;
- Assumption 3 holds whenever the distribution function of V is continuous;
- Assumption 4 is fulfilled with $q = 1$ whenever $\phi < \alpha$.

These assumptions suffice for consistency of the estimator. As for the central limit theorem:

- Assumption 4 is fulfilled with $q = 2$ whenever $2\phi < \alpha$;
- Assumption 5 imposes second order regular variation on X and it implies that Assumption 6 holds as well.
- Assumption 7 is valid when X is Pareto.

Example 5.4.7 We consider the situation from Example 2.4.6. Recall that (U, V) is a bivariate regularly varying random vector with the exponent measure $\boldsymbol{\nu}$ that is not concentrated on the axes. Hence both U and V are regularly varying with index $-\alpha$ and tail equivalent. As such, we can write

$$\frac{\mathbb{P}\left(u^{-1}(U, V) \in \cdot\right)}{\mathbb{P}(U > u)} \xrightarrow{\mathrm{v}} \boldsymbol{\nu}$$

on $[0, \infty]^2 \setminus \{\mathbf{0}\}$, as $u \to \infty$. Let (U_i, V_i), $i = 1, 2$, be independent copies from (U, V). For $\phi, \theta \in (0, 1)$ define

$$(X, Y) = B(U_1, V_1^\phi) + (1 - B)(U_2^\theta, V_2) ,$$

where $\mathbb{P}(B = 0) = \mathbb{P}(B = 1) = 1/2$ and B is independent of U_i, V_i, $i = 1, 2$. Both X and Y are regularly varying with index $-\alpha$ due to Breiman lemma and in fact $\mathbb{P}(X > u) \sim 0.5\mathbb{P}(U > u)$ as $u \to \infty$. Furthermore, (X, Y) is bivariate regularly varying, with extremal independence (even though (U, V) may have extremal dependence.) This implies that $R \equiv 0$ and the [8] approach is not applicable.

We have already verified that Assumption 1 holds.

For Assumption 3 we need to assure that the measure $\boldsymbol{\nu}$ that appears in the definition of regular variation of (X, Y) does not have atoms. This holds when the joint distribution of (U, V) is continuous.

We consider Assumption 4. We have

$$\begin{aligned}
&\lim_{u\to\infty}\frac{1}{\bar{F}_X(u)}\mathbb{E}\left[\left(\frac{Y}{u^{\phi}}\right)^{1+\delta}\mathbb{1}_{\{X>u\}}\right] = \\
&= \lim_{u\to\infty}\frac{1}{\bar{F}_U(u)}\mathbb{E}\left[\left(\frac{V_1^{\phi}}{u^{\phi}}\right)^{1+\delta}\mathbb{1}_{\{U_1>u\}}\right] + \lim_{u\to\infty}\frac{1}{\bar{F}_U(u)}\mathbb{E}\left[\left(\frac{V_2}{u^{\phi}}\right)^{1+\delta}\mathbb{1}_{\{U_2^{\theta}>u\}}\right] \\
&= \int_1^{\infty}\int_0^{\infty} y^{\phi(1+\delta)}\boldsymbol{\nu}(\mathrm{d}x,\mathrm{d}y) + \lim_{u\to\infty}\frac{1}{\bar{F}_U(u^{1/\theta})}\mathbb{E}\left[\left(\frac{V}{u^{1/\theta}}\right)^{1+\delta}\mathbb{1}_{\{U>u^{1/\theta}\}}\right]\frac{\bar{F}_U(u^{1/\theta})}{\bar{F}_U(u)}\frac{u^{(1+\delta)/\theta}}{u^{(1+\delta)\phi}} \\
&= \int_1^{\infty}\int_0^{\infty} y^{\phi(1+\delta)}\boldsymbol{\nu}(\mathrm{d}x,\mathrm{d}y) + \int_1^{\infty}\int_0^{\infty} y^{(1+\delta)}\boldsymbol{\nu}(\mathrm{d}x,\mathrm{d}y)\lim_{u\to\infty}\frac{\bar{F}_U(u^{1/\theta})}{\bar{F}_U(u)}\frac{u^{(1+\delta)/\theta}}{u^{(1+\delta)\phi}}
\end{aligned}$$

if the assumptions of Corollary 2.4.4 are fulfilled for the regularly varying pair (U, V). As noted there, the finiteness of the second integral implies $\alpha > 1$. However, we need to guarantee that

$$\lim_{u\to\infty}\frac{\bar{F}_U(u^{1/\theta})}{\bar{F}_U(u)}\frac{u^{(1+\delta)/\theta}}{u^{(1+\delta)\phi}}$$

is finite. For this, it suffices that

$$(\alpha - 1)/\theta > \alpha - \phi$$

which is in fact an additional restriction that has to be imposed.

We summarize this example as follows:

- Assumption 1 is fulfilled with $b(u) = u^{\phi}$;
- Assumption 3 is fulfilled when the joint distribution of (U, V) is continuous;
- Assumption 4 is fulfilled with $q = 1$ whenever $(\alpha - 1)/\theta > \alpha - \phi$.

These assumptions suffice for consistency of the estimator.

Example 5.4.8 As in Example 2.4.7 consider a pair (X, Y) with the following distribution function:

$$F(x,y) = \frac{(1-x^{-\alpha})(1-y^{-\alpha})}{1+(xy)^{-\alpha}} .$$

We note that this example is generated from Ali-Mikhail-Haq copula with $\theta = -1$. The marginal distributions of X and Y are clearly Pareto with index α. Arguing as in [10], we have $R \equiv 0$, Assumption 1 is fulfilled with $b(u) = 1$ and the limiting measure is the product measure

$$\boldsymbol{\mu}((s,\infty] \times [0,t]) = s^{-\alpha}(1-t^{-\alpha}) .$$

In particular, $\aleph_{\mathrm{CTE}}(1) = \mathbb{E}[Y] = \frac{\alpha}{\alpha-1}$ whenever $\alpha > 1$. Assumption 3 is clearly fulfilled while Assumption 4 holds with $q = 1$ whenever $\alpha > 1$. Hence, for consistency we only need $\alpha > 1$.

Example 5.4.9 Recall the situation from Example 2.4.8. Assume that X is a regularly varying random variable with index $-\alpha$, $\alpha > 1$. Let U be a nonnegative random variable, independent of X such that $\mathbb{E}[U^{\alpha/\phi+\delta}] < \infty$ for some $\delta > 0$. For $\phi \in (0,1)$ define $Y = X^{\phi}U$. By Breiman's lemma, Y is also regularly varying with index $-\alpha/\phi$. The pair (X, Y) is bivariate regularly varying with index $-\alpha$, with the exponent measure concentrated on the horizontal axes, and hence extremally independent. The main difference between the current situation and Example 2.4.5 is that the extremal independence here is linked to different tail behaviour of X and Y.

Assumption 1 is satisfied with the scaling function given by $b(u) = u^{\phi}$. On the other hand,

$$R(x,y) = \mathbb{E}\left[x \vee yU^{\alpha/\phi}\right] .$$

That is, the function R is not identical to zero and the model can be handled by the approach from [8]. One can argue that this is the way to go, since one does not need to estimate the parameter ϕ, as in case of CEV methodology.

5.4.3 Simulations

In this section we conduct simulation studies to illustrate the performance of our estimation procedure.

Example 5.4.6 Continued

The first experiment involves the situation from Example 5.4.6. We simulate $n = 1000$ observation from the model with $\alpha = 4$ and $\phi = 1/2$. First, we estimate α, ϕ and the ratio ϕ/α using the appropriate Hill estimators. The results are depicted on Figure 5.14. We notice that although the estimation of α and ϕ is not very good, the results for ϕ/α are much better (and this is what really matters for the performance of the estimator of the expected shortfall).

Next, we deal with estimation of the expected shortfall (5.4.1) with $p = 0.01$ using the estimators (5.4.7)-(5.4.8). In our example, $Q_X(p) = p^{-1/\alpha}$ and we can evaluate that

$$\theta(p) = \frac{\alpha^2}{(\alpha-1)(\alpha-\phi)} p^{-\phi/\alpha} = 1.52381 p^{-0.125} \ .$$

Figure 5.15 shows the Hill plot for the estimators of the expected shortfall. For comparison purposes, we illustrate on Figure 5.16 how the method of [8] works in case of model of Example 5.4.6 (more precisely, how it doesn't work). As mentioned in the Example, the approach from [8] is not theoretically justified in this case. aaaa

Next, we perform a Monte Carlo study for the model described above. We estimate the Expected Shortfall using $= 5\%, 10\%, 15\%, 20\%$ order statistics. The boxplots for the estimates obtained in $N = 1000$ simulations are shown on Figure 5.17. Again, we show that the [8] method is not applicable.

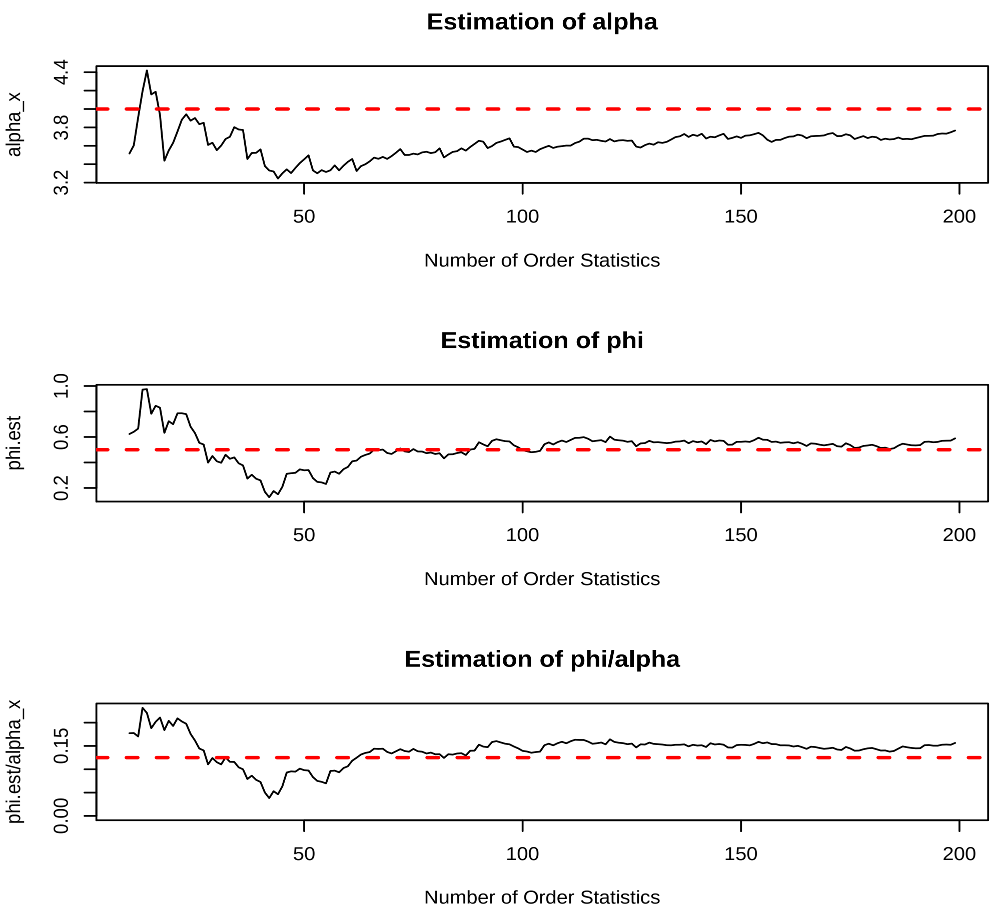

Figure 5.14: Hill plots for α (top panel); ϕ (middle panel); ϕ/α (bottom panel)

Bivariate regular variation

In this section we consider the model $X = |ZU|$, $Y = |ZV|$, where Z is Pareto with $\alpha = 4$, $U = \sin(\Psi)$, $V = \cos(\Psi)$, Ψ is uniform on $[0, 2\pi]$ and is independent of Z. In this case $\phi = 1$, both X, Y are regularly varying with index α and the vector (X, Y) is regularly varying with extremal dependence. Also, the [8] method is applicable.

Figure 5.18 shows boxplots for the estimates of the Expected Shortfall based on $N = 1000$ simulations. We note that the method from that [8] provides better

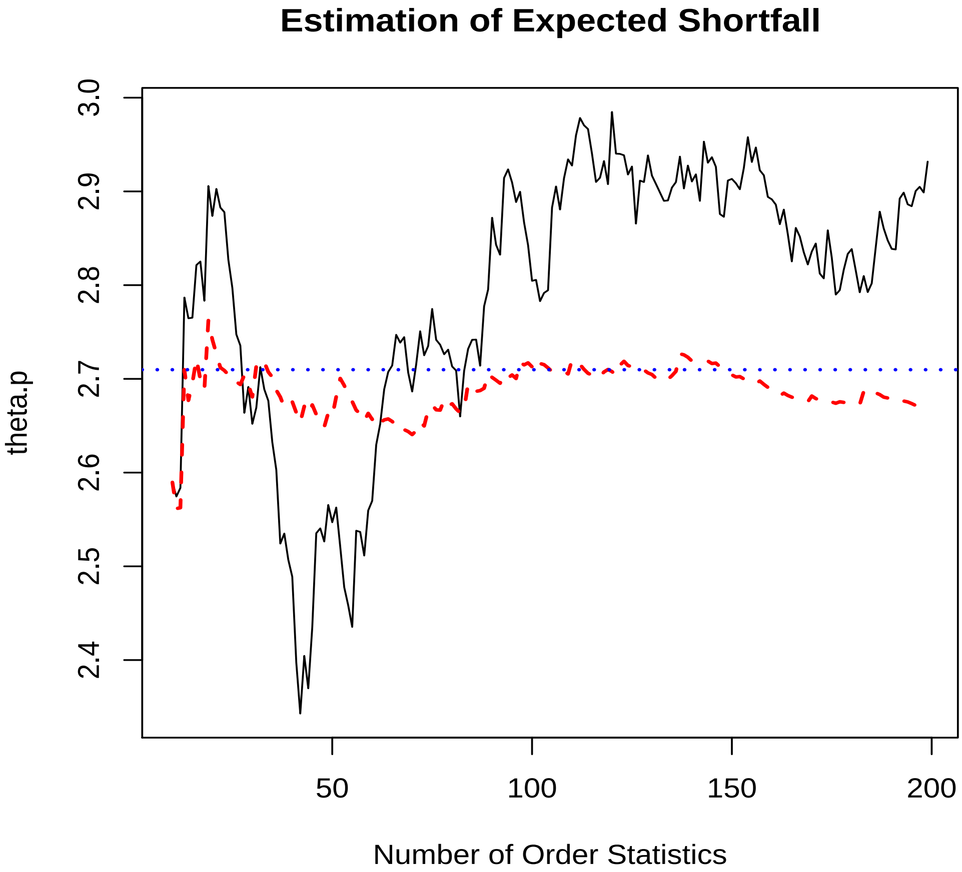

Figure 5.15: Plots for the Expected Shortfall. Solid line - estimation using (5.4.8); dashed line - (5.4.7); horizontal line - the true value of $\theta(p)$.

estimates (less variability). This is not surprising since it requires estimation of the index of regular variation of Y only, instead of α and ϕ.

5.4.4 Data Analysis

We analyse stock prices for Exxon Mobil Corporation and Goldcorp Inc.[3] The time period is January 2, 2005 to Nov 4, 2015. For the absolute log-returns (X and Y in

[3] Source: https://ca.finance.yahoo.com

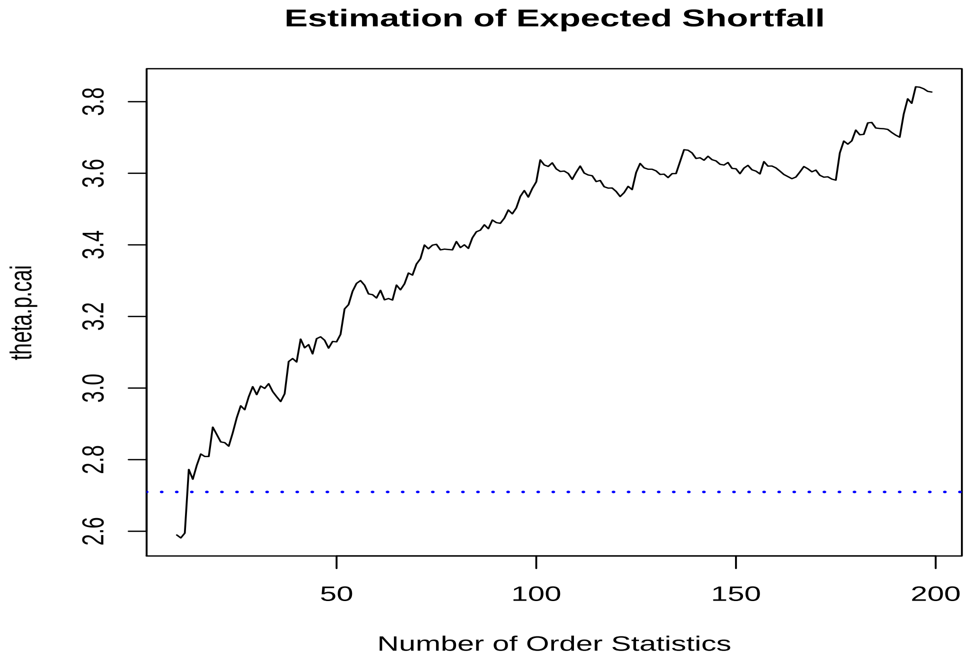

Figure 5.16: Plots for the Expected Shortfall using [8] method.

what follows) we note that the Hill plot indicates the tail indices to be around 4, while different techniques to "verify" the extremal dependence, like the tail dependence coefficient, comparison of the estimated tail indices of X and $\min(X, Y)$, as well as X and XY do not give a clear answer whether the extremal dependence can be assumed.[4] Furthermore, the standard correlation coefficient is very low (around 0.1). In other words, it is not clear for us if the extremal dependence holds and hence if [8] method is applicable.

We proceed with the method of the present paper. Figure 5.19 shows the Hill plot for the estimation of the scaling exponent. For the order statistics $k = 50, \ldots, 300$ the plot stabilizes below 1, indicating extremal independence. We apply our procedure to estimate the expected shortfall at the level $p = 0.001$. We note that the method

[4]The plots are not included in the thesis.

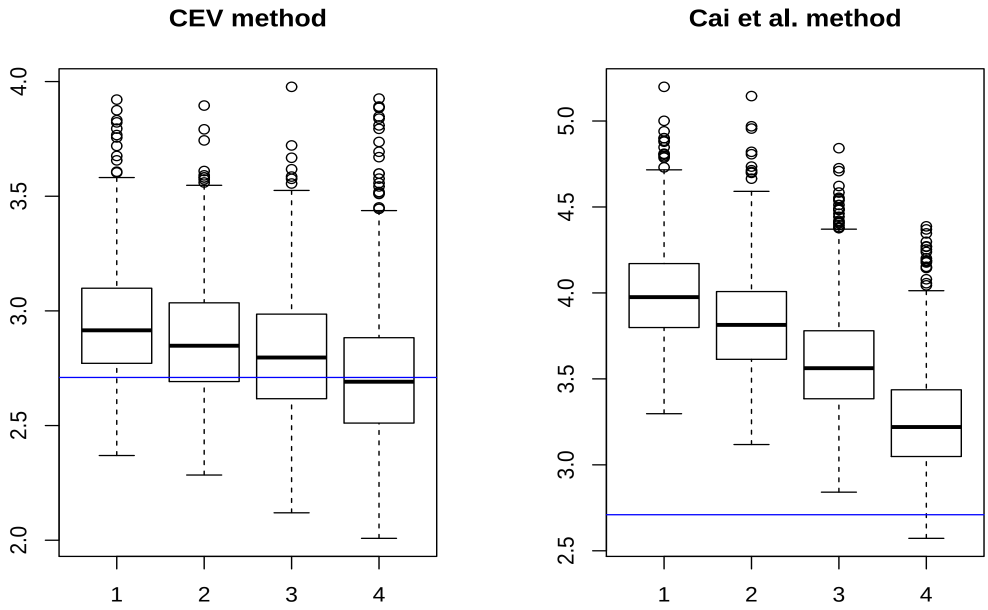

Figure 5.17: Boxplots for the Expected Shortfall. The boxplots correspond to $k = 20\%, 15\%, 10\%, 5\%$ of order statistics being used.

of this paper provides a reasonable estimate to be in the range (0.03,0.035) (based on the order statistics $k = 50, \ldots, 300$), while the [8] method does not provide any conclusive answer, since the plot seems to have a positive trend.

5.4.5 Proofs

Proof of Proposition 5.4.1

We use notation from Section 4.4. Let $\xi_n = \frac{X_{n:n-k}}{u_n}$. We have

$$\int_{\xi_n}^{\infty} \int_0^{\infty} y \mathbb{T}_n(\mathrm{d}x, \mathrm{d}y) = \sqrt{k} \int_{\xi_n}^{\infty} \int_0^{\infty} y(\widetilde{T}_n(\mathrm{d}x, \mathrm{d}y) - T_n(\mathrm{d}x, \mathrm{d}y))$$
$$= \sqrt{k} \sum_{j=1}^{n} \frac{Y_j}{kb(u_n)} \mathbb{1}_{\{X_j > X_{n:n-k}\}} - \sqrt{k} \int_{\xi_n}^{\infty} \int_0^{\infty} y T_n(\mathrm{d}x, \mathrm{d}y)$$

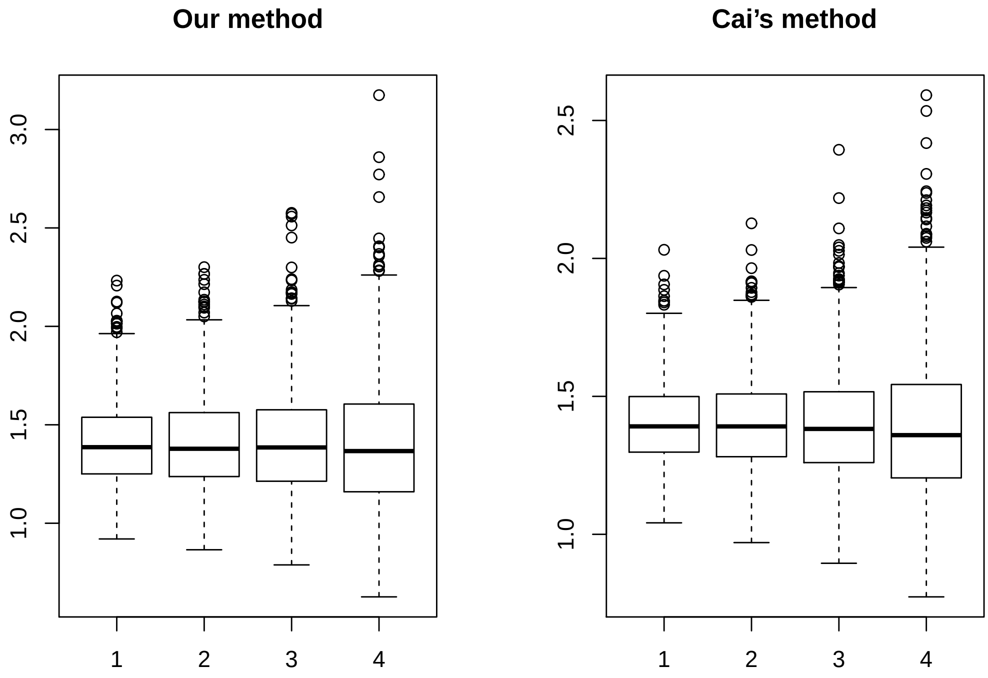

Figure 5.18: Boxplots for the Expected Shortfall - bivariate regular variation.

$$= \sqrt{k}\left(\frac{\widehat{\theta}(k/n)}{b(u_n)} - \aleph_{\mathrm{CTE}}(1)\right) + \sqrt{k}\left(\aleph_{\mathrm{CTE}}(1) - \int_{\xi_n}^{\infty}\int_0^{\infty} yT_n(\mathrm{d}x, \mathrm{d}y)\right) .$$

Hence,

$$\left(\frac{\widehat{\theta}(k/n)}{b(u_n)} - \aleph_{\mathrm{CTE}}(1)\right)$$
$$= \frac{1}{\sqrt{k}}\int_{\xi_n}^{\infty}\int_0^{\infty} y\mathbb{T}_n(\mathrm{d}x, \mathrm{d}y) + \left(\int_{\xi_n}^{\infty}\int_0^{\infty} yT_n(\mathrm{d}x, \mathrm{d}y) - \aleph_{\mathrm{CTE}}(1)\right) := J_1 + J_2$$

and

$$\sqrt{k}\left(\frac{\widehat{\theta}(k/n)}{b(u_n)} - \aleph_{\mathrm{CTE}}(1)\right) = \sqrt{k}J_1 + \sqrt{k}J_2 := I_1 + I_2 .$$

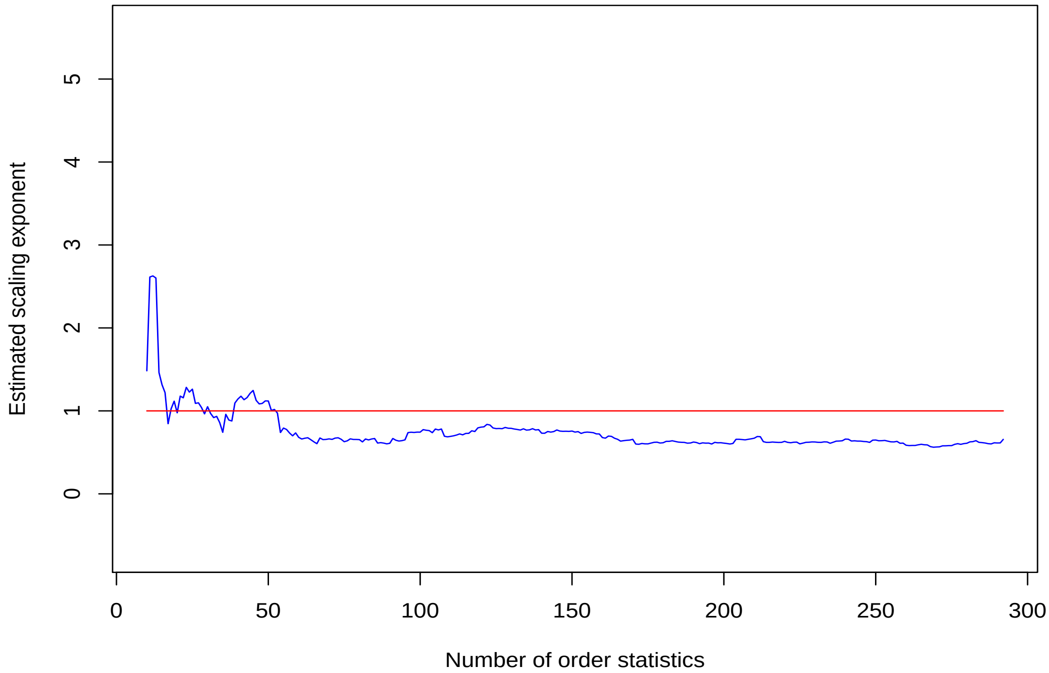

Figure 5.19: Conditional scaling exponent.

We deal with J_1 and I_1 simultaneously. First, we will show that

$$\widetilde{I}_1 := \int_{x=1}^{\infty}\int_{y=0}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y) \xrightarrow{\mathrm{d}} \int_{x=1}^{\infty}\int_{y=0}^{\infty} y\mathbb{T}(\mathrm{d}x,\mathrm{d}y)\ , \quad \widetilde{J}_1 = \frac{1}{\sqrt{k}}\widetilde{I}_1 \xrightarrow{\mathrm{P}} 0\ , \tag{5.4.12}$$

and this will be followed by verifying that

$$|I_1 - \widetilde{I}_1| = \left|\int_{x=\xi_n}^{1}\int_{y=0}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right| \xrightarrow{\mathrm{P}} 0\ , \quad |J_1 - \widetilde{J}_1| \xrightarrow{\mathrm{P}} 0\ . \tag{5.4.13}$$

As a consequence, $I_1 \xrightarrow{\mathrm{d}} Z_1$ and $J_1 = o_P(1)$.

Fix $A > 0$ and decompose

$$\widetilde{I}_1 = \int_{x=1}^{A}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y) + \int_{x=1}^{A}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)$$

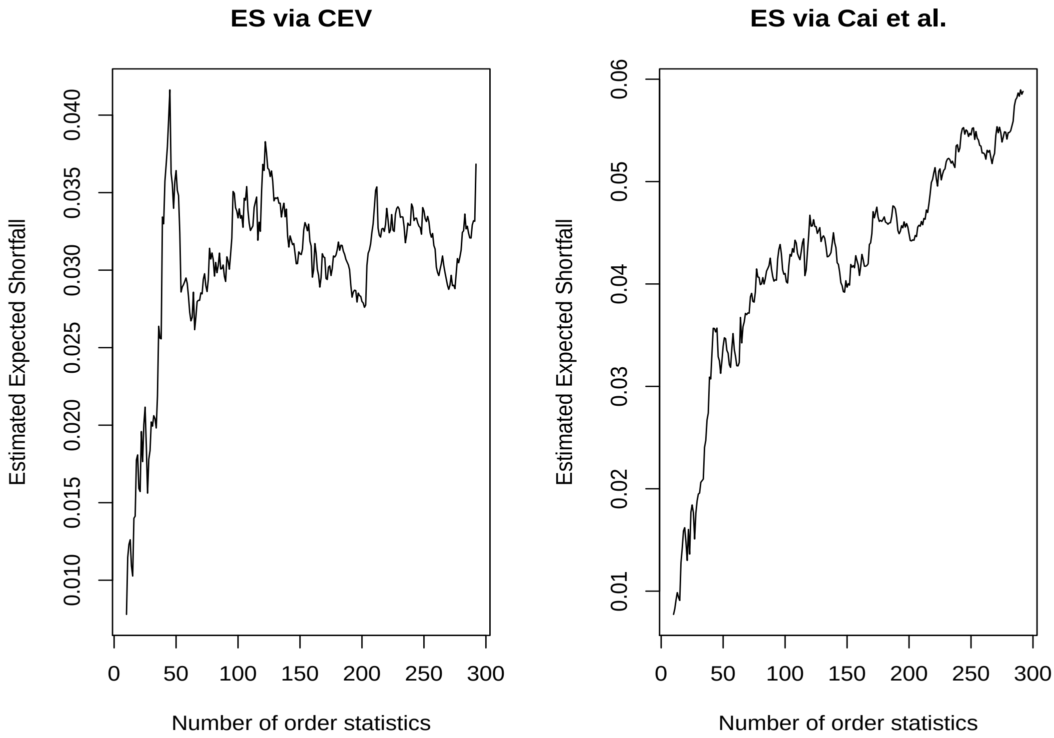

Figure 5.20: Expected Shortfall.

$$+\int_{x=A}^{\infty}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)+\int_{x=A}^{\infty}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y):=\widetilde{I}_{11}+\widetilde{I}_{12}+\widetilde{I}_{13}+\widetilde{I}_{14}\ .$$

Define $\widetilde{J}_{1i}=(1/\sqrt{k})\widetilde{I}_{1i}$, $i=1,\ldots,4$. Since $\mathbb{T}_n$ converges weakly by Proposition 4.4.1, it converges uniformly on compact sets. Hence

$$\widetilde{I}_{11}\xrightarrow{\mathrm{d}}\int_{x=1}^{A}\int_{y=0}^{A} y\mathbb{T}(\mathrm{d}x,\mathrm{d}y)\ ,\quad \widetilde{J}_{11}\xrightarrow{\mathrm{p}}0\ .$$

To conclude (5.4.12) we need to show that

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widetilde{J}_{12}+\widetilde{J}_{14}|>\varepsilon)=0\ ,\quad \lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widetilde{I}_{12}+\widetilde{I}_{14}|>\varepsilon)=0\ ,\qquad (5.4.14)$$

and

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widetilde{J}_{13}|>\varepsilon)=0\ ,\quad \lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\widetilde{I}_{13}|>\varepsilon)=0\ .\qquad (5.4.15)$$

For (5.4.14), it suffices to verify that

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[\frac{1}{\sqrt{k}}\left|\int_{x=1}^{\infty}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right|\right]=0\,,$$
$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathrm{Var}\left(\int_{x=1}^{\infty}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right)=0\,.$$

The first part of (5.4.14) is bounded by

$$\begin{aligned}
&\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{E}\left(\frac{1}{\sqrt{k}}\int_{x=1}^{\infty}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right)\\
&\le \lim_{A\to\infty}\limsup_{n\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y}{b(u_n)}\mathbb{1}_{\{X>u_n\}}\mathbb{1}_{\{Y>b(u_n)A\}}\right)\\
&\le \lim_{A\to\infty}\limsup_{n\to\infty}\frac{n}{k}\frac{1}{A^\delta}\mathbb{E}\left(\frac{Y^{1+\delta}}{b^{1+\delta}(u_n)}\mathbb{1}_{\{X>u_n\}}\right)\,.
\end{aligned}$$

and hence vanishes by Assumption 4 with $q=1$ only. For the second part of (5.4.14) we need Assumption 4 with $q=2$:

$$\begin{aligned}
&\mathrm{Var}\left(\sqrt{k}\int_{x=1}^{\infty}\int_{y=A}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right)=\frac{n}{k}\mathrm{Var}\left(\frac{Y}{b(u_n)}\mathbb{1}_{\{X>u_n\}}\mathbb{1}_{\{Y>b(u_n)A\}}\right)\\
&\le \lim_{A\to\infty}\limsup_{n\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y^2}{b^2(u_n)}\mathbb{1}_{\{X>u_n\}}\mathbb{1}_{\{Y>b(u_n)A\}}\right)\\
&\le \lim_{A\to\infty}\limsup_{n\to\infty}\frac{n}{k}\frac{1}{A^\delta}\mathbb{E}\left(\frac{Y^{2+\delta}}{b^{2+\delta}(u_n)}\mathbb{1}_{\{X>u_n\}}\right)\\
&= \lim_{A\to\infty}\limsup_{n\to\infty}\frac{n}{k}\frac{1}{A^\delta}\bar{F}_X(u_n)\mathbb{E}\left(\frac{Y^{2+\delta}}{b^{2+\delta}(u_n)}\mid X>u_n\right)\\
&= \lim_{A\to\infty}\limsup_{n\to\infty}\frac{1}{A^\delta}\mathbb{E}\left(\frac{Y^{2+\delta}}{b^{2+\delta}(u_n)}\mid X>u_n\right)=0\,.
\end{aligned}$$

For (5.4.15), we argue that

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[\frac{1}{\sqrt{k}}\int_{x=A}^{\infty}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right]=0$$

and

$$\lim_{A\to\infty}\limsup_{n\to\infty}\mathrm{Var}\left(\sqrt{k}\int_{x=A}^{\infty}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right)=0\,.$$

Note that any random variables U, V such that $\mathbb{E}[|U|] + \mathbb{E}[V^2] < \infty$ we have $\mathbb{E}[|U - \mathbb{E}[U]|] \leq 2\mathbb{E}[|U|]$, $\mathrm{Var}(|V - \mathbb{E}[V]|) \leq \mathrm{Var}(V)$. We apply these bounds in the computations that follow.

Recalling that, by Lemma 2.4.3, Assumption 4 used with $q = 1$ implies $\phi - \alpha < 0$, we have

$$\lim_{A\to\infty}\lim_{n\to\infty} \mathbb{E}\left[\frac{1}{\sqrt{k}}\left|\int_{x=A}^{\infty}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right|\right) \leq 2\lim_{A\to\infty}\lim_{n\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y}{b(u_n)}\mathbb{1}_{\{X>u_nA\}}\mathbb{1}_{\{Y<b(u_n)A\}}\right)$$
$$\leq 2\lim_{A\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y}{b(u_n)}\mathbb{1}_{\{X>u_nA\}}\right) \leq \aleph \lim_{A\to\infty} A^{\phi-\alpha} = 0\ .$$

Likewise, using Assumption 4 with $q = 2$ and Lemma 2.4.3 (so that $2\phi - \alpha < 0$) we have

$$\lim_{A\to\infty}\lim_{n\to\infty} \mathrm{Var}\left(\sqrt{k}\int_{x=A}^{\infty}\int_{y=0}^{A} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right) \leq \lim_{A\to\infty}\lim_{n\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y^2}{b^2(u_n)}\mathbb{1}_{\{X>u_nA\}}\mathbb{1}_{\{Y<b(u_n)A\}}\right)$$
$$\leq \lim_{A\to\infty}\frac{n}{k}\mathbb{E}\left(\frac{Y^2}{b^2(u_n)}\mathbb{1}_{\{X>u_nA\}}\right) \leq \aleph \lim_{A\to\infty} A^{2\phi-\alpha} = 0\ .$$

Hence, we have shown (5.4.12).

Now, for arbitrary $\epsilon, \epsilon' > 0$,

$$\begin{aligned}
&\mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon) \\
&= \mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon, |\xi_n - 1| > \epsilon') + \mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon, |\xi_n - 1| < \epsilon') \\
&\leq \mathbb{P}(|\xi_n - 1| > \epsilon') + \mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon, |\xi_n - 1| < \epsilon')\ .
\end{aligned}$$

We know that $\xi_n \xrightarrow{\mathrm{P}} 1$, hence $\mathbb{P}(|\xi_n - 1| > \epsilon') \to 0$. Also

$$\mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon, |\xi_n - 1| < \epsilon') \leq \mathbb{P}\left(\left|\int_{1-\epsilon'}^{1+\epsilon'}\int_{0}^{\infty} y\mathbb{T}_n(\mathrm{d}x,\mathrm{d}y)\right| > \epsilon\right)\ .$$

We obtain

$$\mathbb{P}(|I_1 - \widetilde{I}_1| > \epsilon, |\xi_n - 1| < \epsilon') \leq \epsilon^{-2}\mathrm{Var}\left(\sqrt{k}\frac{1}{k}\sum_{j=1}^{n}\frac{Y_j}{b(u_n)}\mathbb{1}_{\{u_n(1-\epsilon')<X_j<u_n(1+\epsilon')\}}\right)$$

$$\le \epsilon^{-2}\frac{n}{k}\mathbb{E}\left[\frac{Y^2}{b^2(u_n)}\mathbb{1}_{\{u_n(1-\epsilon')<X<u_n(1+\epsilon')\}}\right]$$
$$\to \epsilon^{-2}\left\{(1-\epsilon')^{2\phi-\alpha}-(1+\epsilon')^{2\phi-\alpha}\right\}\int_1^\infty\int_0^\infty y^2 T(\mathrm{d}x,\mathrm{d}y)\ ,$$

as $n\to\infty$ by Assumption 4 and Lemma 2.4.3. Letting $\epsilon'\to 0$, the limit is 0. The similar calculation is applied to $\widetilde{J}_1$ using Assumption 4 with $q=1$ only. The identity (5.4.13) is proven. This concludes the proof of convergence for I_1 and J_1.

We proceed with J_2 and I_2. Convergence of order statistics and the uniform convergence of T_n to T imply that

$$\int_{\xi_n}^\infty\int_0^\infty yT_n(\mathrm{d}x,\mathrm{d}y)=\int_1^\infty\int_0^\infty yT(\mathrm{d}x,\mathrm{d}y)=\aleph_{\mathrm{CTE}}(1)\ .$$

Hence, J_2 converges to 0 in probability. Furthermore, Assumption 6 yields that it suffices to consider the integral, where T_n is replaced with T. The delta method and convergence of order statistics imply that

$$I_2\xrightarrow{\mathrm{d}}\alpha^{-1}\mathbb{T}(1,\infty)\int_0^\infty yT(1,\mathrm{d}y)=Z_2$$

jointly with Z_1.

Bibliography

[1] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[2] Bojan Basrak, Richard A. Davis, and Thomas Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99(1):95–115, 2002.

[3] Bojan Basrak and Johan Segers. Regularly varying multivariate time series. *Stochastic Process. Appl.*, 119(4):1055–1080, 2009.

[4] Jan Beirlant, Yuri Goegebeur, Johann Segers, and Josef Teugels. *Statistics for Extremes: Theory and Applications*. Wiley, 2004.

[5] Patrick Billingsley. *Convergence of probability measures*. New York, Wiley, 1968.

[6] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

[7] Nicholas H. Bingham, Charles M. Goldie, and Jozef L. Teugels. *Regular variation*. Cambridge University Press, Cambridge, 1989.

[8] Juan-Juan Cai, John J.H.J Einmahl, Laurens de Haan, and Chen Zhou. Estimation of the marginal expected shortfall: the mean when a related variable

is extreme. *Journal of the Royal Statistical Society. Series B. Methodological*, 77(2):417–442, 2014.

[9] Daren B. H. Cline. Regular variation of order 1 nonlinear AR-ARCH models. *Stochastic Processes and their Applications*, 117(7):840—861, 2007.

[10] Bikramjit Das and Sidney I. Resnick. Conditioning on an extreme component: model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.

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

[12] Laurens de Haan and Ana Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006.

[13] Stefano Demarta and Alexander McNeil. The t copula and related copulas. Unpublished manuscript, 2004.

[14] Randal Douc, Eric Moulines, and David S. Stoffer. *Nonlinear time series*. Chapman & Hall/CRC Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL, 2014.

[15] Holger Drees. Optimal rates of convergence for estimates of the extreme value index. *The Annals of Statistics*, 26(1):434–448, 1998.

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

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

[18] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997.

[19] Charles M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1(1):126–166, 1991.

[20] Janet E. Heffernan and Sidney I. Resnick. Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571, 2007.

[21] Harry Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.

[22] Rafał Kulik and Philippe Soulier. The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121(1):109 – 134, 2011.

[23] Rafał Kulik and Philippe Soulier. Heavy tailed time series with extremal independence. *Extremes*, 18(2):273 – 299, 2015.

[24] Rafał Kulik and Philippe Soulier. Heavy tailed time series with extremal independence: statistical inference. Preprint, 2016.

[25] Rafał Kulik, Philippe Soulier, and Olivier Wintenberger. The tail empirical process of regularly varying functions of geometrically ergodic Markov chains. To appear in Stochastic Processes and their Applications, 2017.

[26] Rafał Kulik and Zhigang Tong. Multivariate tail estimation: Conditioning on an extreme event. http://arxiv.org/abs/1502.07189, 2015.

[27] Rafał Kulik and Zhigang Tong. Estimation of the expected shortfall given an extreme component under conditional extreme value model. Revised for Extremes, 2016.

[28] Sidney I. Resnick. *Heavy-Tail Phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007.

[29] Holger Rootzén. Weak convergence of the tail empirical process for dependent sequences. *Stoch. Proc. Appl.*, 119(2):468–490, 2009.

[30] Zhigang Tong and Rafał Kulik. Extremogram estimation for regularly varying time series. Submitted to Scandinavian Journal of Statistics, 2016.

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

[32] Ward Whitt. Some useful functions for functional limit theorems. *Mathematics of Operations Research*, 5(1):67–85, 1980.

Index