On Some Connections between Light Tails, Regular Variation and Extremes

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

vorgelegt von
Anja Janßen
aus Hamburg

Göttingen 2010

Referent: Prof. Dr. Martin Schlather
Korreferent: PD Dr. Ulf-Rainer Fiebig
Tag der mündlichen Prüfung: 3. November 2010

Acknowledgement

At this place I would like to thank several people who supported my writing of this PhD thesis.

First of all, I would like to thank my supervisor Prof. Dr. Martin Schlather for his neverending encouragement and for providing me with new ideas while also giving me the opportunity to find and pursue my own interests. Remarkably, even in busy times he always found the time to listen and help.

Furthermore, I would like to thank PD Dr. Ulf-Rainer Fiebig for frequent discussions which started as valuable help regarding the intricacies which emerged during the writing of Chapter 3 and have continued to be a dear source of support and advice.

Additionally, I would like to thank Junior-Prof. Dr. Zakhar Kabluchko and Prof. Dr. Holger Drees for helpful discussions regarding Chapters 2 and 3 and Prof. Johan Segers, Ph.D., and Assistant Prof. Bojan Basrak, Ph.D., for their kind correspondence regarding Chapter 4.

I am also grateful to the Centre for Statistics and the federal state of Lower Saxony for supporting me in form of a Lichtenberg Scholarship.

Last, but certainly not least, I would like to thank my family who always encouraged me to pursuit my aims and supported me wherever they could. My final thanks go to my colleagues Andree and Michael for their kindness, help and friendship.

Contents

Notation

$\mathcal{D}_\infty(G)$	Max domain of attraction of G, page 11
$\overline{F}$	$1 - F$, tail distribution function
$\stackrel{d}{=}, \stackrel{d}{\to}$	Equality in distribution, convergence in distribution
$\stackrel{v}{\to}$	Vague convergence, page 36
$\stackrel{w}{\to}$	Weak convergence
ϕ	Characteristic function
sign	Signum function, $\text{sign}(x) = 1$ if $x > 0$, $\text{sign}(x) = -1$ if $x < 0$ and $\text{sign}(x) = 0$ if $x = 0$, page 28
$\vee, \wedge$	Maximum, minimum
$o_x(1)$	A function that tends to zero as x does, page 45
$\bar{\mathbb{R}}, \bar{\mathbb{R}}_+, \bar{\mathbb{R}}_{+,0}$	$\mathbb{R} \cup \{-\infty, \infty\}, (0, \infty], [0, \infty]$
$\mathbb{B}, \bar{\mathbb{B}}$	Borel-σ-Algebra on $\mathbb{R}$, Borel-σ-Algebra on $\bar{\mathbb{R}}$
$\mathbb{C}$	Cone, $x \in \mathbb{C} \Rightarrow tx \in \mathbb{C}$, $\forall t \in \mathbb{R}$, page 36
$\mathbb{S}^{d-1}, \mathbb{S}_+^{d-1}$	$\{\mathbf{x} \in \mathbb{R}^d \big\vert \|\mathbf{x}\| = 1\}, \{\mathbf{x} \in \mathbb{R}^d \big\vert \|\mathbf{x}\| = 1, \mathbf{x} \geq 0\}$
$M_+(\mathbb{C})$	Space of all non-negative Radon measures on $\mathbb{C}$, page 37
∂A	Boundary of the set A
$\mathcal{L}$	Probability law
N_{μ,σ^2}	Normal distribution with mean μ and variance σ^2, stands also for its distribution function
$\text{Par}(\kappa)$	Pareto distribution with index $\kappa > 0$, page 26
$\text{Unif}[a, b]$	Uniform distribution on $-\infty < a < b < \infty$

Chapter 1

Introduction: On Light Tails, Heavy Tails and Extremes

1.1 Preliminaries

The aim of this thesis is to find and explore new connections between two classes of distributions, those with heavy tails and those with light ones, with a special focus on their extremal behavior. Before we start to explain the course of analysis and the selected connections in more detail, we will take the time to introduce the types of distributions, clarify what we mean when speaking of "light" and "heavy" tails of distributions and note some of the connections between the three types which have been studied so far.

First of all, note that the word "tail" has a double meaning in probability theory: The tail function of a random variable X with distribution function F is given by $\overline{F} := 1-F$. Second, the word "tail" is often used as an abbreviation for "tail behavior" of a distribution and describes the asymptotic behavior of $\overline{F}$ as $x \to \infty$ (right tail of the distribution) and of F as $x \to -\infty$ (left tail of the distribution). As we will see later on, the tail of a distribution always comes into play when we deal with limit theorems in probability, for example for sums or maxima of i.i.d. random variables. At the latest since the evolving of extreme value theory one could notice a tendency towards clustering different types of distributions which then got the label "light-tailed" or "heavy-tailed". To start with, the phrase "light-tailed" is a term often used in the literature but seldom exactly defined. Often, this expression stands for distribution functions F for which $\int e^{\alpha|x|}\, F(dx) < \infty$ holds for an $\alpha > 0$. This assumption implies that the random variable X which has distribution function F possesses finite moments of all orders. This includes all bounded random variables and those which have tails that decrease at least exponentially fast, meaning that there exists a $\beta > 0$ such that both

$$\lim_{x\to\infty} e^{\beta x}\overline{F}(x) = 0 \quad \text{and} \quad \lim_{x\to-\infty} e^{-\beta x}F(x) = 0$$

hold. As for the "heavy-tailed" random variables, this characteristic is often associ-

ated with so-called regular variation of the respective distribution function. We call a function $f : \mathbb{R} \to \mathbb{R}$ regularly varying (at $+\infty$) with index $\lambda \in \mathbb{R}$ if

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\lambda \tag{1.1.1}$$

holds for all $t > 0$. If $\lambda = 0$ then the function is called "slowly varying". See [6] for an extensive treatment of this topic. A non-negative random variable X is called regularly varying with index $\alpha > 0$ if the respective tail distribution function is regularly varying with index $-\alpha$. Extensions of this definition to real valued random variables and random vectors have been made and will later be introduced when needed.

The next two sections deal with some already known connections between light and heavy tails which constitute the background for our further analysis.

1.2 Connections through Limit Theorems

Probably the most common example of a "typical" light-tailed distribution is the normal distribution, which we denote as N_{μ,σ^2}. Its predominant role in probability theory is demonstrated by the Central Limit Theorem:

Theorem 1.2.1 (Central Limit Theorem according to Lindeberg, cf. [20])
Let $X_1, X_2, \cdots \in \mathbb{R}$ be mutually independent random variables with distribution functions $F_1, F_2, \ldots$ such that

$$\mathbb{E}(X_k) = 0, \quad \mathrm{Var}(X_k) = \sigma_k^2 > 0,$$

and put

$$s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2, \quad n \in \mathbb{N}.$$

Assume that for each $t > 0$

$$s_n^{-2} \sum_{k=1}^{n} \int_{|y|\geq t s_n} y^2 \, F_k(dy) \to 0, \quad n \to \infty, \tag{1.2.1a}$$

or, what amounts to the same, that

$$s_n^{-2} \sum_{k=1}^{n} \int_{|y|< t s_n} y^2 \, F_k(dy) \to 1, \quad n \to \infty. \tag{1.2.1b}$$

Then the distribution of the normalized sum

$$S_n^* = \frac{X_1 + \cdots + X_n}{s_n}$$

converges to the standard normal distribution $N_{0,1}$.

Often, this theorem is applied to the special case of identically distributed random variables where $s_n = n\sigma_1^2 \to \infty$. It is thus evident from (1.2.1a) that the limit behavior of the normalized sums S_n^* is only determined by the tails of the distribution function F_1 and unaffected by its behavior in regions which are bounded away from $-\infty$ and ∞. It is well known that in the i.i.d. case a finite variance of X_1 is sufficient for the Central Limit Theorem to hold. Therefore, all light-tailed distributions as introduced in the beginning of this chapter (except for the degenerate ones) satisfy the conditions of the Central Limit Theorem. Before the rise of extreme value theory and the recently developed sensitivity for non light-tailed distributions, the Central Limit Theorem was often used as a justification for the assumption of normally distributed random variables in models: The normal distribution emerges in all cases where the variable in question is the sum of many small, light-tailed influences and where we can only observe the final result, which may then be modelled by the the normal distribution.

More generally, limit theorems for sums of i.i.d. random variables may also hold if they possess infinite variances and in those cases the limit may well differ from the normal distribution. This leads us to the class of α-stable distributions which can be described by their characteristic function ϕ of the type (cf. [21], Chapter 34)

$$\log(\phi(t)) = iat - b|t|^\alpha \left(1 + ic\,\mathrm{sign}(t)\omega(t,\alpha)\right),$$

with $a \in \mathbb{R}, b \geq 0, c \in [-1,1], \alpha \in (0,2]$ and

$$\omega(t,\alpha) = \begin{cases} \tan\left(\frac{\pi}{2}\alpha\right), & \text{if } \alpha \neq 1, \\ \frac{2}{\pi}\log|t|, & \text{if } \alpha = 1. \end{cases}$$

The normal distribution is a special case of this class of distributions with $\alpha = 2$. It is now well known that if $X_1, X_2, \ldots$ are i.i.d. random variables and if there exist sequences of norming constants $a(n)$ and $b(n), n \in \mathbb{N}$, such that

$$\frac{\sum_{i=1}^n X_i - b(n)}{a(n)} \tag{1.2.2}$$

converges in distribution (to a non-degenerate limit), this limit distribution is necessarily α-stable (cf. [21], Chapter 33). Thus, for $\alpha = 2$ the limit distribution is itself light-tailed, while one can show that for $\alpha < 2$ the limit distribution has regularly varying tails with index α (cf. [20], Chapter 17.5). By comparing the light-tailed normal distribution and the heavy-tailed rest of the α-stable distributions we may therefore observe the first connections between the aforementioned classes of distributions.

Now, apart from results about the uniqueness of possible limit distributions, there has always been great interest in necessary and sufficient conditions on the distribution of the X_i in order to ensure the weak convergence of (1.2.2). Starting from a certain α-stable limit distribution this problem deals with the characterisation of the so-called "(sum-)domains of attraction". Again, regular variation is an important characteristic in this context. For non-negative i.i.d. random variables $X_1, X_2, \ldots$ with distribution

function F there exist sequences $a(n) > 0, b(n) \in \mathbb{R}, n \in \mathbb{N}$, such that (1.2.2) converges weakly to a non-degenerate limit distribution if and only if the so-called truncated moment function

$$\mu(x) = \int_0^x y^2 \, F(dy)$$

is regularly varying with index $2 - \alpha, \alpha \in (0, 2]$, where the value of α corresponds to the index of the α-stable limit distribution. For real valued random variables an additional, but mild assumption about the balance of positive and negative tail is necessary, cf. [20], Chapter 17. Thus, both the light-tailed and heavy-tailed distributions play an important role in limit theorems for sums of i.i.d. random variables and relevant distributions of both groups can be seen as special representatives of the parametric family of α-stable distributions.

A second branch of limit theorems are those for maxima of i.i.d. random variables, which build the foundation of extreme value theory. In contrast to (1.2.2), one is interested in weak convergence of the distribution of

$$\frac{\max(X_1, \ldots, X_n) - b(n)}{a(n)} \tag{1.2.3}$$

for suitable sequences of norming constants $a(n) > 0$ and $b(n) \in \mathbb{R}, n \in \mathbb{N}$. It is a fundamental theorem in extreme value theory that the only possible non-degenerate limit distributions (up to a linear transformation) are of the form (cf. [17], Chapter 3.4)

$$F_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), & \gamma \neq 0 \\ \exp\left(-\exp(-x)\right), & \gamma = 0 \end{cases}, \quad \text{for } 1 + \gamma x > 0, \quad F_\gamma(x) = 0, \quad \text{else.}$$

Note that $\lim_{\gamma \to 0}(1 + \gamma x)^{-1/\gamma} = \exp(-x)$ for all $x \in \mathbb{R}$, thus we have a smooth parametrization. One often distinguishes the cases $\gamma < 0$ (Weibull), $\gamma = 0$ (Gumbel) and $\gamma > 0$ (Fréchet) - see Chapter 2 for details. We note that for $\gamma \leq 0$ the distribution F_γ is light-tailed (meaning that all moments exist) while for $\gamma > 0$ it is heavy-tailed and regularly varying with index γ^{-1}. Again, we have a transition from light-tailed distributions to heavy-tailed ones in the spectrum of possible limit distributions.

Like in the case of sums, one is interested not only in possible limit distributions but also in necessary and sufficient conditions on the distribution of the X_i in order to ensure the weak convergence of (1.2.3). Starting from an extreme value distribution F_γ this problem deals with the identification of the so-called "(max-)domains of attraction". Again, regular variation is an important characteristic in this context, especially for the case $\gamma > 0$. For i.i.d. random varibales $X_1, X_2, \ldots$ with distribution function F there exists a sequence $a(n) > 0, n \in \mathbb{N}$, such that (1.2.3) (with $b(n) \equiv 0$) converges weakly to F_γ with $\gamma > 0$ if and only if $\overline{F}$ is regularly varying at $+\infty$ with index $1/\gamma$. The characterisation for the case $\gamma < 0$ also involves regular variation, but in a modified form since this case only includes bounded random variables. Light-tailed random

variables like the normal oder exponential distribution are typical representatives of the max domain of attraction for the case $\gamma = 0$ (Gumbel distribution), which is the most tedious case of the three regarding its charaterization. Chapter 2 deals with this case in more detail.

Noticing the similarities between the limit theorems for sums and for maxima of i.i.d. random variables one could furthermore ask whether it is possible to combine those limit theorems in a more general framework. Based on [38], this has been done in Chapter 2.

1.3 Connections through Random Difference Equations

The last section introduced some functional similarities between special cases of light-tailed and heavy-tailed distributions regarding both limit distributions for sums and maxima of i.i.d. random variables and their respective domains of attraction. Note, however, that in general light-tailed distributions of the random variables $X_1, X_2, \ldots$ lead to light-tailed limit distributions, while heavy-tailed distributions of the random variables $X_1, X_2, \ldots$ lead to heavy-tailed limit distributions both for sums and for maxima. In this section, in contrast, we will examine how light-tailed random variables can be used as "input" for a time series model which then has heavy-tailed random variables as "output" in the form of a stationary solution. While a multivariate version of this model will be studied in Chapter 4, we will introduce the univariate version here first (cf. [22]).

Let $(A_t, B_t)_{t\in\mathbb{N}}$ be an i.i.d. sequence of $\mathbb{R}^2$-valued random vectors and let a real-valued time series $(X_t)_{t\in\mathbb{N}_0}$ satisfy the recursive equation

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{N}, \tag{1.3.1}$$

where X_0 is supposed to be independent of $(A_t, B_t)_{t\in\mathbb{N}}$. Equation (1.3.1) is then called "random difference equation" (RDE). Applications of this type of model are numerous, cf. [42], and imply that X_t represents the value of certain objects (e.g. the price of an asset, the amount of a certain good or the size of a population) at time t, where B_t is a quantity that is added or subtracted and A_t determines the proportional growth or decay of the stock from time $t-1$ to t. Now, an important question for this kind of equations is whether a stationary distribution of $(X_t)_{t\in\mathbb{N}}$ exists. Note that a random variable X with this stationary distribution would satisfy the equation

$$X \stackrel{d}{=} AX + B, \tag{1.3.2}$$

where (A, B) has the same distribution as $(A_t, B_t), t \in \mathbb{N}$, and is independent of X (cf. [42], Lemma 1.1).

Iterating the recursion we get that

$$X_t = B_t + A_t B_{t-1} + A_t A_{t-1} B_{t-2} + \cdots + A_t A_{t-1} \dots A_2 B_1 + A_t A_{t-1} \dots A_1 X_0, \quad t \in \mathbb{N}.$$

A sufficient condition for a stationary solution to (1.3.2) is the existence of a $\kappa > 0$ such that

$$E|A|^\kappa = 1, \quad E|A|^\kappa \log^+ |A| < \infty, \quad 0 < E|B|^\kappa < \infty,$$

together with some mild assumptions about certain distributions related to A and B being non-lattice (cf. [28]). These assumptions are typically satisfied for light-tailed distributions of A and B. Although it is in general hard to make precise statements about the distribution of a stationary solution to (1.3.2), it is well-known that the solution has certain asymptotic properties. One can show that under the assumptions made above the distribution of X solving (1.3.2) asymptotically behaves like the distribution of $\max_{n\in\mathbb{N}} |\prod_{i=1}^n A_i|$ (cf. [28]). Taking logarithms, we may identify the transformed random variable as the maximum of a random walk $(\max_{n\in\mathbb{N}} \sum_{i=1}^n \log |A_i|)$ whose distribution can be found using renewal theory (cf. [19], Chapter 11). One can then show that there exist constants c_- and c_+ such that

$$P(X > x) \sim c_+ x^{-\kappa} \quad \text{and} \quad P(X < -x) \sim c_- |x|^{-\kappa} \tag{1.3.3}$$

for large values of x. Thus, by starting with light-tailed random variables $(A_t, B_t)_{t\in\mathbb{N}}$ and using a simple recursive definition we end up with heavy-tailed random variables as the stationary solution to the process studied. Examples for applications of the model (1.3.1) and especially its multivariate extensions encompass the popular GARCH(p, q) models for financial time series and are in line with the observations of heavy tails of marginal distributions for financial data. More on this will be laid out in Chapter 3. It is devoted to the extension of the work done in [39] to two connected time series. It deals with the so-called "tail chain" of a process, which is an instrument to describe the behavior of a time series $(Y_t)_{t\in\mathbb{Z}}$ given that we have an extreme observation at time zero. Putting it more formally, we look at the limit

$$\lim_{u\to\infty} \mathcal{L}\left(\frac{Y_{-n}}{u}, \dots, \frac{Y_0}{u}, \dots, \frac{Y_m}{u} \,\middle|\, |Y_0| > u \right) \tag{1.3.4}$$

for $m, n \in \mathbb{N}$. One can show that the existence of a non-degenerate weak limit of (1.3.4) is closely related to multivariate regular variation of the time series. Furthermore, if the time series behaves asymptotically like a random difference equation (the meaning of "asymptotically" in this context is explained in [39] and Chapter 3) one can show that the limit law in (1.3.4) has an especially simple form, like a multiplicative random walk with light-tailed increments.

It becomes clear from Equation (1.3.3) that κ, the tail index of regular variation for X, is the key characteristic for the asymptotic behavior of the distribution of the stationary solution to (1.3.2). In the univariate case, we recall from the assumptions made

above that κ is uniquely determined by being the solution of the equation $E|A|^\kappa = 1$. While this equation is analytically solvable for some distributions of A, it can always be tackled numerically (at least as long as one can simulate from the distribution of $|A|$), and therefore the derivation of κ bears no problems in the one-dimensional setting. This changes dramatically as one passes over to multivariate RDEs (cf. Chapter 4), where the finding of the right κ for a given distribution of the matrix $\mathbf{A}$ is connected to the solution of a relation much more complicated than $E|A|^\kappa = 1$ and a simple formula for κ depending on the distribution of $\mathbf{A}$ is only known in a few special cases. A new way of assessing the index of regular variation by a combination of theory and Monte Carlo simulation is proposed in Chapter 4.

Chapter 2

Limit Laws for Power Sums and Norms of i.i.d. Samples

2.1 Introduction

The aim of this chapter is to explore connections between limit laws for sums and limit laws for maxima by combining both sums and maxima in the broader setting of l_p-norms. This chapter is based on the article [26].

The use of l_p-norms is a well-known way of measuring vectors in $\mathbb{R}^n$. In the following, we study l_p-norms of random samples, where $X_1, X_2, \ldots$ are i.i.d. positive random variables. Let the l_p-norm $\|\cdot\|_p$ of a positive sample $X_{1n} = (X_1, X_2, \ldots, X_n)$ be defined as

$$\|X_{1n}\|_p = \left(\sum_{i=1}^{n} X_i^p\right)^{1/p}, \; p > 0, \tag{2.1.1}$$

and $\|X_{1n}\|_\infty = \max\{X_1, \ldots, X_n\}$. (We do not exclude $0 < p < 1$ in (2.1.1) although $\|\cdot\|_p$ is not a norm in that case.) For $p = 1$ or $p = \infty$ and $n \to \infty$ the limit behavior of these norms is well explored in terms of limit results for sums and for maxima of i.i.d. samples which have already been mentioned in the introduction. Here, however, we allow p to grow with n, thus looking at $l_{p(n)}$-norms of samples as $n \to \infty$. Our analysis of $l_{p(n)}$-norms of samples is based on the closely related behavior of power sums of the form

$$\left(\|X_{1n}\|_{p(n)}\right)^{p(n)} = \sum_{i=1}^{n} X_i^{p(n)}.$$

For both norms and power sums, the limit behavior is determined by the distribution of X_1 and the behaviour of $p(n)$ as $n \to \infty$. Typically, an important role will be played by the largest summand, especially for heavy-tailed distributions of X_1 and fast-growing sequences $p(n)$. We show that nontrivial limit laws emerge if $p(n)$ is chosen in accordance with the tail behavior of X_1. The resulting families of limit laws

can be considered as links between the limit theory for sums on the one hand and extreme value theory on the other hand.

For samples X_{1n} of positive i.i.d. random variables and properly chosen sequences $p(n) \to \infty$ the limit behavior of the appropriately normalized power sums

$$\frac{\sum_{i=1}^{n} X_i^{p(n)} - \hat{b}(n)}{\hat{a}(n)}$$

has been studied recently in [4] and [7]. Both articles are based on a Tauberian approach and assume a certain asymptotic tail behavior of X_1. Surprisingly, the emerging families of limit distributions are identical for the different tail behaviors studied there. We will demonstrate that this generality is no coincidence, that is, we will show that the limit behavior of power sums and norms of samples is basically governed by max domains of attraction. Now, since the distributions studied in [4] and [7] belong to the Gumbel max domain of attraction, they yield the same family of limit laws.

While it is illustrated in [4] that the analysis of power sums has applications, for example, in branching processes and the Random Energy Model (see [11]), the limit distributions of norms are of theoretical interest, since they build a smooth transition between limit laws for sums and for maxima.

For samples X_{1n} of i.i.d. positive random variables from the max domain of attraction of the Weibull or Fréchet distribution the limit laws for the suitably normalized $l_{p(n)}$-norms

$$\frac{\|X_{1n}\|_{p(n)} - \tilde{b}_c(n)}{\tilde{a}_c(n)}$$

have been derived in [38]. Schlather's method, which is based on the asymptotic tail behavior of the distribution of X_1, works well for the cases studied there but fails for the Gumbel domain of attraction as it contains distributions with a wide range of different tail behavior.

Using a connection between extreme value theory and limit theorems for sums in our proofs, we follow a different approach. This method is applicable to each of the three max domains of attraction in the same way as demonstrated in Section 2. For the Gumbel case, by use of the von Mises representation of its max domain of attraction, we obtain convergence to the family of distributions found in [4] and [7], which is stated in Section 3, and proved in Section 4. However, in some cases this convergence is restricted to certain subsequences. Section 5 provides such an example as well as some further results for the Gumbel case. Section 6 is dedicated to the Weibull and the Fréchet case.

2.2 Fundamentals

Our approach is based on extreme value theory. Consequently, all distributions that are considered here are assumed to be in the max domain of attraction of an extreme value distribution.

Let $X_1, X_2, \ldots$ be i.i.d. with distribution function F. Then F is said to be in the max domain of attraction of an extreme value distribution with distribution function G if there exist norming constants $a(n)$ and $b(n)$ such that

$$\lim_{n\to\infty} P\left(\frac{\max\{X_1,\ldots,X_n\} - b(n)}{a(n)} \leq x\right) = G(x) \tag{2.2.1}$$

for all $x \in \mathbb{R}$. Then G belongs to one of the three possible extreme value distributions (see, e.g., [37]), namely

- the Weibull distribution with parameter $\alpha > 0$

$$\Psi_\alpha(x) = \exp(-(-x)^\alpha), \qquad x \leq 0,$$

- the Fréchet distribution with parameter $\alpha > 0$

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \qquad x \geq 0,$$

- the Gumbel distribution

$$\Lambda(x) = \exp(-e^{-x}), \qquad x \in \mathbb{R}.$$

We write $F \in \mathcal{D}_\infty(G)$ if F is in the max domain of attraction of G.

Both limit theorems for maxima and for sums can be formulated in such a way that part of the necessary and sufficient conditions depends on the behavior of the tails of the distribution functions. This connection between sums and maxima, which, to our knowledge, has been established first in [29], p. 315, will form the base of our analysis.

One can show (see, e.g., [17], Proposition 3.3.2) that equation (2.2.1) is equivalent to

$$\lim_{n\to\infty} n\overline{F}(a(n)x + b(n)) = -\log(G(x)), \qquad x \in \mathbb{R}. \tag{2.2.2}$$

On the other hand, for sums

$$S_n = \xi_{n,1} + \ldots + \xi_{n,k_n} - b(n)$$

of independent and infinitesimal random variables $\xi_{n,l}$ with distribution functions F_n, $n \in \mathbb{N}, 1 \leq l \leq k_n$, a criterion for convergence as formulated in [21], p. 116-117, includes an expression similar to (2.2.2). Here, it is necessary that there exist nondecreasing functions

$$M(\cdot) \text{ with } M(-\infty) = 0 \quad \text{and} \quad N(\cdot) \text{ with } N(+\infty) = 0,$$

defined on $[-\infty, 0)$ and $(0, +\infty]$, respectively, such that at every continuity point of $M(u)$ and $N(u)$ it holds that

$$\lim_{n\to\infty} k_n F_n(u) = M(u), \qquad u < 0, \tag{2.2.3}$$

$$\lim_{n\to\infty} k_n \overline{F}_n(u) = -N(u), \qquad u > 0. \tag{2.2.4}$$

The similarity between (2.2.2) and (2.2.4) is fundamental to our analysis. In the next section, it will be used for the Gumbel max domain of attraction and we will derive limit laws for power sums. As a second step, we then use the following lemma by Bogachev [7] to derive the limit distributions for the corresponding norms.

Lemma 2.2.1 ([7], Lemma 9.1) *Let $\{S(t), t \geq 0\}$ be a family of positive random variables, such that for some (non-negative) functions $B(t)$, $A(t)$ and a non-degenerate random variable Y,*

$$S^*(t) := \frac{S(t) - B(t)}{A(t)} \xrightarrow{d} Y \qquad (t \to \infty),$$

where $\xrightarrow{d}$ stands for convergence in distribution. Set $R(t) := S(t)^{1/t}$ and $B^*(t) := B(t)/A(t)$.

(a) If $B^*(t) \to \infty$ as $t \to \infty$, then

$$tB^*(t)\left(\frac{R(t)}{B(t)^{1/t}} - 1\right) \xrightarrow{d} Y \qquad (t \to \infty).$$

(b) If $B^*(t) \equiv 0$ then

$$t\left(\frac{R(t)}{A(t)^{1/t}} - 1\right) \xrightarrow{d} \log Y \qquad (t \to \infty).$$

2.3 Main results for the Gumbel case

Throughout this section let $X_1, X_2, \ldots$ be a sequence of i.i.d. positive random variables with distribution function $F \in \mathcal{D}_\infty(\Lambda)$. Then, according to (2.2.2), there exist norming constants $a(n)$ and $b(n)$, such that

$$\lim_{n\to\infty} n\overline{F}(a(n)x + b(n)) = -\log(\Lambda(x)) = \exp(-x). \tag{2.3.1}$$

If the summands $\xi_{n,k} := (X_k - b(n))/a(n)$ were infinitesimal, equation (2.3.1) would ensure that the convergence criterion (2.2.4) is met for the sums $\sum_{i=1}^n \xi_{n,i}$. However, since

$$\begin{aligned} \sup_{1\leq k\leq n} P\left(\left|\frac{X_k - b(n)}{a(n)}\right| \geq \epsilon\right) &\geq \sup_{1\leq k\leq n} P\left(\frac{X_k - b(n)}{a(n)} \leq -\epsilon\right) \\ &\geq P\left(\frac{\max_{1\leq k\leq n} X_k - b(n)}{a(n)} \leq -\epsilon\right) \\ &\to \Lambda(-\epsilon) > 0, \qquad n \to \infty, \end{aligned} \tag{2.3.2}$$

we know that this is not the case. To apply the connection between (2.2.2) and (2.2.4), we use a power transformation of the summands. If follows from the domain of the Gumbel distribution, which is $(-\infty, \infty)$, that

$$b(n)/a(n) \to \infty \tag{2.3.3}$$

as $n \to \infty$. Let us replace x in (2.3.1) by

$$g_n(x) := \frac{b(n)}{a(n)}(x^{a(n)/cb(n)} - 1)$$

with $c > 0$. Because of (2.3.3) this function converges pointwise to $c^{-1} \log x$ as $n \to \infty$. Therefore,

$$\begin{aligned} n\overline{F}\left(b(n)x^{a(n)/cb(n)}\right) &= n\overline{F}(a(n)g_n(x) + b(n)) \\ &= n\overline{F}\left(a(n)\left(c^{-1}\log x + o(1)\right) + b(n)\right) \\ &\sim -\log\left(\Lambda\left(c^{-1}\log x\right)\right) = x^{-1/c}, \qquad n \to \infty, \end{aligned} \tag{2.3.4}$$

where the asymptotic relation follows from the local uniformity of weak convergence to a continuous limit (see [17], p. 149). Note that formula (2.3.4) equals condition (2.2.4) for sums of the form

$$\sum_{i=1}^{n} \left(\frac{X_i}{b(n)}\right)^{cb(n)/a(n)}, \tag{2.3.5}$$

which consist of infinitesimal summands. Consequently, we will analyze limit theorems for power sums of the form (2.3.5). An analogous procedure can be applied to the Weibull case, while the power transformation is not needed in the Fréchet case.

For further investigation of the Gumbel case, the so-called von Mises representation of a distribution function $F \in \mathcal{D}_\infty(\Lambda)$ will be used.

Lemma 2.3.1 ([37], Proposition 1.4) *A distribution function F with upper endpoint $x_\infty \leq \infty$ belongs to the Gumbel max domain of attraction if and only if for $x \in (-\infty, x_\infty)$ there exists a representation*

$$\overline{F}(x) = \overline{c}(x) \exp\left(-\int_{z_0}^{x} \frac{\mathbb{1}_{(z_0, x_\infty)}(u)}{f(u)}\, du\right) \tag{2.3.6}$$

for some $z_0 < x_\infty$ with $\lim_{x \to x_\infty} \overline{c}(x) = \overline{c} > 0$ and an absolutely continuous strictly positive function f on $[z_0, x_\infty)$ with $\lim_{x \to x_\infty} f'(x) = 0$.

We are now able to formulate our main result.

Theorem 2.3.2 *Let $X_1, X_2, \ldots$ be positive i.i.d. random variables with distribution function $F \in \mathcal{D}_\infty(\Lambda)$. Choose norming constants for the sequence of maxima*

$$b(n) = F^{\leftarrow}(1 - 1/n) \quad \textit{and} \quad a(n) = f(b(n))$$

with $F^{\leftarrow}(x) = \inf\{t \in \mathbb{R} : F(t) \geq x\}$, f *from the von Mises representation* (2.3.6) *and define* $p(n) := b(n)/a(n)$. *There exists a family* F_c *of distribution functions of the form given below and a subsequence* $n_k \to \infty$ *such that for every* $c > 0$ *there exist norming constants* $\hat{a}_c(n_k), \hat{b}_c(n_k), k \in \mathbb{N}$, *so that*

$$F_c(x) = \lim_{k\to\infty} P\left(\frac{\sum_{i=1}^{n_k} X_i^{cp(n_k)} - \hat{b}_c(n_k)}{\hat{a}_c(n_k)} \leq x\right).$$

The norming constants $\hat{a}_c(n_k)$ *and* $\hat{b}_c(n_k)$ *can be chosen according to Table* 2.1.

c	$\hat{a}_c(n)$	$\hat{b}_c(n)$
$0 < c \leq \frac{1}{2}$	$\sqrt{n\mathrm{Var}(X_1^{cp(n)}\mathbb{1}_{\{X_1 \leq b(n)\}})}$	$nE(X_1^{cp(n)}\mathbb{1}_{\{X_1^{cp(n)} \leq \hat{a}_c(n)\}})$
$\frac{1}{2} < c < 1$	$b(n)^{cp(n)}$	$nE(X_1^{cp(n)}\mathbb{1}_{\{X_1 \leq b(n)\}}) + \frac{b(n)^{cp(n)}}{1-c}$
$c = 1$	$b(n)^{cp(n)}$	$nE(X_1^{cp(n)}\mathbb{1}_{\{X_1 \leq b(n)\}})$
$c > 1$	$b(n)^{cp(n)}$	0

Table 2.1: Norming constants

If $0 < c \leq \frac{1}{2}$, *the* F_c *equal the standard normal distribution function* $N_{0,1}$.
If $\frac{1}{2} < c < \infty$, *the* F_c *are given by* $F_c(x) = G_{1/c}(x)$, *where* $G_{1/c}$ *is an* α*-stable distribution function with* $\alpha = 1/c$, *skewness* $\beta = 1$ *and characteristic function*

$$\phi_{1/c}(u) = \begin{cases} \exp\left(-\Gamma(1-1/c)|u|^{1/c}\exp\left(-\frac{i\pi}{2c}\operatorname{sign}(u)\right)\right), & c \neq 1 \\ \exp\left(iu(1-\gamma) - \frac{\pi}{2}|u|\left(1 + i\operatorname{sign}(u)\frac{2}{\pi}\log|u|\right)\right), & c = 1. \end{cases} \tag{2.3.7}$$

Here, γ *is the Euler constant, and* $\Gamma(\cdot)$ *is the gamma function.*

The limit laws for $l_{p(n)}$-norms, which are missing in [38], are readily obtained with the help of Lemma 2.2.1.

Corollary 2.3.3 *Let* X_{1n} *and* $p(n)$ *be as in Theorem* 2.3.2. *There exists a family* $\tilde{F}_c$ *of distribution functions of the form given below and a subsequence* $n_k \to \infty$ *such that for every* $c > 0$ *there exist norming constants* $\tilde{a}_c(n_k), \tilde{b}_c(n_k), k \in \mathbb{N}$, *so that*

$$\tilde{F}_c(x) = \lim_{k\to\infty} P\left(\frac{\|X_{1n_k}\|_{cp(n_k)} - \tilde{b}_c(n_k)}{\tilde{a}_c(n_k)} \leq x\right).$$

If $0 < c \leq \frac{1}{2}$, *the* $\tilde{F}_c$ *equal the standard normal distribution function* $N_{0,1}$.
If $\frac{1}{2} < c \leq 1$, *the* $\tilde{F}_c$ *are given by* $\tilde{F}_c(x) = G_{1/c}(x)$.
If $1 < c < \infty$, *the* $\tilde{F}_c$ *are given by* $\tilde{F}_c(x) = G_{1/c}(\exp(x))$.

Furthermore,

$$\tilde{F}_c(cx) \to \Lambda(x), \qquad c \to \infty. \tag{2.3.8}$$

Proof. The existence of norming constants and the form of the limit distributions follow from Lemma 2.2.1. We use part (a) of the lemma if $c \leq 1$ and part (b) if $c > 1$. Relation (2.3.8) has been shown in [7], Theorem 10.2. □

2.4 Proof of Theorem 2.3.2

4.1 Main idea

We use the abovementioned theorem for limit laws of sums (see [21], p. 116-117). It has already been shown that the summands in (2.3.5) are infinitesimal. Let F_n^c denote the distribution function of $(X_1/b(n))^{cp(n)}$. Since $F_n^c(u) = 0$ for all $u < 0$, from (2.2.3) we readily obtain $M(\cdot) \equiv 0$. For $u > 0$ it follows from (2.2.4) and (2.3.4) that for any subsequence $n_k \to \infty$,

$$N(u) = -\lim_{k\to\infty} n_k \overline{F_{n_k}^c}(u) = -u^{-1/c}, \qquad u > 0. \tag{2.4.1}$$

We are now left to prove (see [21], p. 116) that, along a certain subsequence n_k,

$$\begin{aligned} &\lim_{\epsilon\to 0} \liminf_{k\to\infty} n_k \left(\int_0^\epsilon x^2 dF_{n_k}^c - \left(\int_0^\epsilon x\, dF_{n_k}^c(x) \right)^2 \right) \\ &= \lim_{\epsilon\to 0} \limsup_{k\to\infty} n_k \left(\int_0^\epsilon x^2 dF_{n_k}^c - \left(\int_0^\epsilon x\, dF_{n_k}^c(x) \right)^2 \right) =: \sigma^2 \in [0,\infty). \end{aligned} \tag{2.4.2}$$

Note that

$$0 \leq \int_0^\varepsilon x^2 dF_{n_k}^c(x) - \left(\int_0^\varepsilon x dF_{n_k}^c(x) \right)^2 \leq \int_0^\varepsilon x^2 dF_{n_k}^c(x),$$

and we will in fact show that, for an appropriate subsequence n_k,

$$\lim_{k\to\infty} n_k \int_0^\varepsilon x^p dF_{n_k}^c(x) = \frac{1}{cp-1}\varepsilon^{p-1/c}, \qquad p > 1/c, \tag{2.4.3}$$

which will readily imply that $\sigma^2 = 0$. The limit (2.4.3) would follow immediately if one could interchange limit and integration. We will choose a proper subsequence in order to justify this interchange by dominated convergence.

4.2 Choice of a proper subsequence

Let f be given by (2.3.6). Choose a sequence $b_k \to x_\infty$ such that for the function $g(x) := f(x)/x$ the inequality $g(y) \geq g(b_k)$ holds for all $y \in (z_0, b_k)$ and all $k \in \mathbb{N}$. For instance, since g is continuous and converges to zero (see [37], Lemma 1.2), set

$b_k := \min\{x \in [z_0, x_\infty) | g(x) \leq 1/k\}$. Setting $n_k := \lfloor 1/\overline{F}(b_k) \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x, it follows that

$$\begin{aligned} n_k \overline{F}(f(b_k)x + b_k) &\sim \frac{\overline{F}(f(b_k)x + b_k)}{\overline{F}(b_k)} \\ &= \frac{\overline{c}(f(b_k)x + b_k)}{\overline{c}(b_k)} \exp\left(- \int_{b_k}^{f(b_k)x + b_k} \frac{\mathbb{1}_{(z_0, x_\infty)}(u)}{f(u)}\, du \right) \\ &\sim \exp(-x), \qquad k \to \infty, \end{aligned} \tag{2.4.4}$$

with the same argumentation as in the proof of Lemma 2.3.1 (see [37], p. 42). Therefore, both sequences $(f(b_k), b_k)$ and $(a(n_k), b(n_k))$ can be used as norming constants for the maxima along the subsequence n_k. By the Convergence to Types Theorem, it follows that

$$\lim_{k\to\infty} \frac{a(n_k))}{f(b_k)} = \lim_{k\to\infty} \frac{f(b(n_k))}{f(b_k)} = 1 \qquad \text{and} \qquad \lim_{k\to\infty} \frac{b_k - b(n_k)}{f(b_k)} = 0. \tag{2.4.5}$$

In the following let n_k be chosen as described above.

4.3 Application of dominated convergence

Partial integration in (2.4.3) yields

$$\begin{aligned} n_k \int_0^\varepsilon x^p dF^c_{n_k}(x) &= -n_k \int_0^\varepsilon x^p d\overline{F^c_{n_k}}(x) \\ &= -n_k \varepsilon^p \overline{F^c_{n_k}}(\varepsilon) + n_k p \int_0^\varepsilon x^{p-1} \overline{F^c_{n_k}}(x) dx, \end{aligned} \tag{2.4.6}$$

where the first term in (2.4.6) converges to $-\varepsilon^{p-1/c}$ for every subsequence $n_k \to \infty$ by (2.3.4). In what follows let $\beta^c_k := \frac{f(b(n_k))}{cb(n_k)}$. Now,

$$\begin{aligned} n_k \overline{F^c_{n_k}}(x) &= n_k \overline{F}(b(n_k) x^{\beta^c_k}) \\ &\sim \frac{\overline{F}(b(n_k) x^{\beta^c_k})}{\overline{F}(b(n_k))} \\ &= \frac{\overline{c}(b(n_k) x^{\beta^c_k})}{\overline{c}(b(n_k))} \exp\left(\int_{b(n_k)x^{\beta^c_k}}^{b(n_k)} \frac{\mathbb{1}_{(z_0, x_\infty)}(u)}{f(u)}\, du \right) \\ &= \frac{\overline{c}(b(n_k) x^{\beta^c_k})}{\overline{c}(b(n_k))} \exp\Bigg(\frac{1}{c} \int_{\log x}^0 \frac{f(b(n_k)) \exp(\beta^c_k u)}{f(b(n_k) \exp(\beta^c_k u))} \\ &\qquad \times \mathbb{1}_{(z_0, x_\infty)}(b(n_k) \exp(\beta^c_k u))\, du \Bigg). \end{aligned}$$

With the previously defined g we get

$$n_k \overline{F^c_{n_k}}(x) \sim \frac{\overline{c}(b(n_k) x^{\beta^c_k})}{\overline{c}(b(n_k))} \exp\left(\frac{1}{c} \int_{\log x}^0 \frac{g(b(n_k))}{g(b(n_k) \exp(\beta^c_k u))} \right.$$

$$\times \mathbb{1}_{(z_0,x_\infty)}\left(b(n_k)\exp\left(\beta_k^c u\right)\right)\,du\Bigg). \tag{2.4.7}$$

For $\epsilon \in (0,\overline{c})$ there exists a $k(\epsilon) \in \mathbb{N}$ such that $|\overline{c}(b(n_k)) - \overline{c}| < \epsilon$ for all $k > k(\epsilon)$. Since $\overline{F}$ lies between 0 and 1 and the exponential part of $\overline{F}$ in (2.3.6) is monotonically decreasing, it follows that

$$\overline{c}(y) \le \exp\left(\int_{z_0}^{b(n_{k(\epsilon)})} \frac{\mathbb{1}_{(z_0,x_\infty)}(u)}{f(u)}\,du\right)$$

for $y < b(n_{k(\epsilon)})$. Therefore, we can find a constant $C > 0$ such that

$$\frac{\overline{c}(b(n_k)x^{\beta_k^c})}{\overline{c}(b(n_k))} \le C$$

for all $x < 1$ if k is large enough. For such x and k we get

$$\begin{aligned} n_k\overline{F_{n_k}^c}(x) &\le C\exp\left(\frac{1}{c}\int_{\log x}^{0} \frac{g(b_k)(1+\epsilon)}{g(b(n_k)\exp(\beta_k^c u))}\mathbb{1}_{(z_0,x_\infty)}\left(b(n_k)\exp\left(\beta_k^c u\right)\right)\,du\right) \\ &\le C\exp\left(-\frac{1+\epsilon}{c}\log x\right) \end{aligned}$$

because of (2.4.5), the choice of n_k and since $b(n_k) \le b_k$.

Taking the limit in (2.4.6), we can interchange limit and integral by Lebesgue's Theorem to obtain

$$\begin{aligned} \lim_{k\to\infty} n_k\int_0^\varepsilon x^p dF_{n_k}^c(x) &= -\varepsilon^{p-1/c} + p\int_0^\varepsilon x^{p-1}x^{-1/c}dx \\ &= -\varepsilon^{p-1/c} + \frac{cp}{cp-1}\varepsilon^{p-1/c} = \frac{1}{cp-1}\varepsilon^{p-1/c}. \end{aligned}$$

4.4 Limit distributions and norming constants for $c > \frac{1}{2}$

For $c > \frac{1}{2}$ both (2.2.4) and (2.4.2) are met and the limit law has characteristic function $\phi(\cdot)$, where

$$\log(\phi(t)) = i\mu t - \frac{\sigma^2}{2}t^2 + \int_{|u|>0}\left(e^{iut} - 1 - \frac{iut}{1+u^2}\right)dN(u) \tag{2.4.8}$$

(see [21], p. 117), with

$$N(u) = -u^{-1/c}, \qquad \sigma^2 = 0, \tag{2.4.9}$$

and location parameter μ yet to be determined. Hence, the limit law is α-stable (see [21], p. 164) with $\alpha = 1/c$ and skewness parameter $\beta = 1$. The value of μ depends on the choice of additive norming constants. We choose

$$B_c(n) = \begin{cases} 0 & \text{for } c > 1 \\ nE(X_1^{cp(n)}\mathbb{1}_{\{X_1\le b(n)\}}) & \text{for } c = 1 \\ nE(X_1^{cp(n)}\mathbb{1}_{\{X_1\le b(n)\}}) + \frac{b(n)^{cp(n)}}{1-c} & \text{for } 1/2 < c < 1. \end{cases}$$

Then, for $\tau > 0$, it follows that (see [21], pp. 84, 117)

$$\mu = \lim_{k\to\infty}\left[n_k E\left(\left(\frac{X_1}{b(n_k)}\right)^{cp(n_k)} \mathbb{1}_{\left\{\left(\frac{X_1}{b(n_k)}\right)^{cp(n_k)} \le \tau\right\}}\right) - \frac{B_c(n_k)}{b(n_k)^{cp(n_k)}}\right] - \int_0^\tau \frac{x^3}{1+x^2} dN(x) + \int_\tau^\infty \frac{x}{1+x^2} dN(x).$$

Using (2.4.9) and considering various cases for the constant c we obtain

- For $c > 1$:

$$\begin{aligned}\mu &= \lim_{k\to\infty} n_k \int_0^\tau x\, dF^c_{n_k}(x) - 0 - \frac{1}{c}\int_0^\tau \frac{x^{2-1/c}}{1+x^2}\, dx + \frac{1}{c}\int_\tau^\infty \frac{x^{-1/c}}{1+x^2}\, dx \\ &= \frac{1}{c}\int_0^\tau \left(x^{-1/c} - \frac{x^{2-1/c}}{1+x^2}\right) dx + \frac{1}{c}\int_\tau^\infty \frac{x^{-1/c}}{1+x^2}\, dx \\ &= \frac{1}{c}\int_0^\infty \frac{x^{-1/c}}{1+x^2} = \frac{\pi}{2c\cos\left(\frac{\pi}{2c}\right)}\end{aligned}$$

  by (2.4.3) and [23], # 3.241(2).

- For $c = 1$:

$$\begin{aligned}\mu &= \lim_{k\to\infty} n_k \left(\int_0^\tau x\, dF^c_{n_k}(x) - \int_0^1 x\, dF^c_{n_k}(x)\right) \\ &\qquad - \frac{1}{c}\int_0^\tau \frac{x}{1+x^2}\, dx + \frac{1}{c}\int_\tau^\infty \frac{x^{-1}}{1+x^2}\, dx \\ &= \frac{1}{c}\int_1^\tau x^{-1}\, dx - \frac{1}{c}\int_{1/\tau}^\tau \frac{x}{1+x^2}\, dx = \frac{1}{c}\log\tau - \frac{1}{c}\log\tau = 0,\end{aligned}$$

  since limit and integration are interchangeable on the interval $(1, \tau)$, and because of [23], # 2.145(2).

- For $1/2 < c < 1$:

$$\begin{aligned}\mu &= \lim_{k\to\infty} n_k \left(\int_0^\tau x\, dF^c_{n_k}(x) - \int_0^1 x\, dF^c_{n_k}(x)\right) - \frac{1}{1-c} \\ &\qquad - \frac{1}{c}\int_0^\tau \frac{x^{2-1/c}}{1+x^2}\, dx + \frac{1}{c}\int_\tau^\infty \frac{x^{-1/c}}{1+x^2}\, dx \\ &= -\frac{1}{c}\int_\tau^\infty \left(x^{-1/c} - \frac{x^{-1/c}}{1+x^2}\right) dx - \frac{1}{c}\int_0^\tau \frac{x^{2-1/c}}{1+x^2}\, dx \\ &= -\frac{1}{c}\int_0^\infty \frac{x^{2-1/c}}{1+x^2}\, dx = \frac{\pi}{2c\cos\left(\frac{\pi}{2c}\right)}\end{aligned}$$

  again by (2.4.3) and [23], # 3.241(2).

It is shown in [4], Theorem 6.2, that the characteristic functions (2.4.8) may be written in the form (2.3.7) with constants as stated in Theorem 2.3.2.

4.5 Limit distributions and norming constants for $c \leq \frac{1}{2}$

The case $c \leq \frac{1}{2}$ is studied seperately, as we have to change the multiplicative norming constants to keep the sum of the truncated variances bounded. According to [21], p. 130-131, for a given subsequence n_k, it suffices to find a sequence $C_{n_k} \to \infty$ such that

$$\lim_{k\to\infty} n_k \int_{|x|>C_{n_k}} dF^c_{n_k} = 0 \tag{2.4.10}$$

and

$$\lim_{k\to\infty} \frac{n_k}{C^2_{n_k}} \left(\int_{|x|<C_{n_k}} x^2 dF^c_{n_k}(x) - \left(\int_{|x|<C_{n_k}} x dF^c_{n_k}(x) \right)^2 \right) = \infty. \tag{2.4.11}$$

To satisfy condition (2.4.11), we subtract the median m_n from each variable (2.3.5) as done in [19]. It follows from (2.3.4) that m_n tends to zero. From local uniform convergence we obtain

$$n_k \overline{F^{c,m}_{n_k}}(x) \sim x^{-1/c}, \qquad x > 0, \tag{2.4.12}$$

where $F^{c,m}_{n_k}$ denotes the distribution function of the median-subtracted variables. Hence, (2.4.10) is valid for every sequence C_{n_k} that tends to infinity. To verify (2.4.11) we make use of the median-normalization. Feller shows (see [19], p. 527) that it suffices to find a sequence C_{n_k} such that

$$\lim_{k\to\infty} \frac{n_k}{C^2_{n_k}} \left(\int_{|x|<C_{n_k}} x^2 dF^{c,m}_{n_k}(x) \right) = \infty.$$

But it follows from (2.4.12) and Fatou's Lemma that

$$\lim_{k\to\infty} n_k \int_{|x|<\epsilon} x^2 dF^{c,m}_{n_k}(x) = \infty$$

for any $\epsilon > 0$. Hence, an appropriate sequence C_{n_k} can be constructed. We can choose norming constants as given in the theorem.

2.5 Further results for the Gumbel case

The following corollary shows that in some cases the restriction to certain subsequences is not necessary.

Corollary 2.5.1 *Let $X_1, X_2, \ldots$ be i.i.d. with distribution function $F \in \mathcal{D}_\infty(\Lambda)$ with representation (2.3.6) and let $g(x) := f(x)/x$ be ultimately monotone. Let $p(n)$ be*

defined as in Theorem 2.3.2. Then, for every $c > 0$, there exist norming constants $\hat{a}_c(n), \hat{b}_c(n)$ such that

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i^{cp(n)} - \hat{b}_c(n)}{\hat{a}_c(n)} \le x\right) = F_c(x),$$

with F_c defined as in Theorem 2.3.2. The norming constants may be chosen according to Table 1.

Proof. If g is ultimately monotone, the constant z_0 can be shifted in such a way that g is monotonically decreasing on (z_0, x_∞). The proof of Theorem 2.3.2 then shows the convergence for the whole sequence. □

The following example illustrates that the restriction in Theorem 2.3.2 is not purely technical.

Example 2.5.2. Let $X_1, X_2, \ldots$ be i.i.d. with tail distribution function

$$\overline{F}(x) = \exp\left(-\int_0^x \frac{u}{1.5 + \sin(u)}\, du\right).$$

Then $F \in \mathcal{D}_\infty(\Lambda)$, but there exist $c > 0$ and a sequence $n_k \to \infty$ such that no norming constants exist that ensure convergence of the power sums $\sum_{i=1}^{n_k} X_i^{cp(n_k)}$ to the limit distribution function F_c of Theorem 2.3.2.

Proof. Set $f(x) = x^{-1}(1.5 + \sin(x))$. Since $f'(x) \to 0$ as $x \to x_\infty = \infty$, it follows from Lemma 2.3.1 that F has a von Mises representation (2.3.6), and therefore $F \in \mathcal{D}_\infty(\Lambda)$. Now, let us choose a subsequence that does not allow for the interchange of limit and integral on the left hand side of (2.4.3). Here, the idea is to find a sequence n_k such that the integrand in (2.4.7) is large for small values of x. By reasoning similar to the proof of Theorem 2.3.2, it is possible to choose a subsequence n_k such that $b(n_k) = (k+0.5)\pi$ are suitable norming constants for the maxima. With $g(x) = f(x)/x = x^{-2}(1.5 + \sin(x))$, along this sequence we have $g(b(n_k)) = 2.5((k + 0.5)\pi)^{-2}$. It follows for the second term on the right hand side of (2.4.6) that

$$\begin{aligned} 2n_k \int_0^\varepsilon x\overline{F_{n_k}^c}(x)\, dx &\sim 2\int_0^\varepsilon \exp\left(\log x + \int_{b(n_k)x^{\beta_k^c}}^{b(n_k)} \frac{1}{f(u)}\, du\right) dx \\ &= 2\int_{-\infty}^{\log\varepsilon} \exp\left(2y + \int_{b(n_k)\exp(\beta_k^c y)}^{b(n_k)} \frac{1}{f(u)}\, du\right) dy. \end{aligned} \tag{2.5.1}$$

Next, we will show that $-(2+\epsilon)y$ is a lower bound for the inner integral in (2.5.1) in certain regions of the integration range. To this end, we analyze the behavior of the inner integral for y in a neighborhood of

$$y_k := \log\left(\frac{k-0.5}{k+0.5}\right)\frac{(k+0.5)^2\pi^2}{2.5} = -\frac{(k+0.5)\pi^2}{2.5} + o(1).$$

Choose $\delta > 0$ and $y \in (y_k - \delta, y_k + \delta)$ with $\delta(y) := y - y_k$. It follows that

$$\begin{aligned}
-\frac{1}{y} & \int_{b(n_k)\exp\left((y_k+\delta(y))\beta_k^c\right)}^{b(n_k)} \frac{1}{f(u)}\,du \\
= & \int_{(k-0.5)\pi\exp(\delta(y)g(b(n_k))/c)}^{(k+0.5)\pi} \frac{u}{1.5+\sin(u)}\,du\ \frac{2.5}{(k+0.5)\pi^2}(1+o(1)) \\
= & \int_{(k-0.5)\pi+o(1)}^{(k+0.5)\pi} \frac{1}{1.5+\sin(u)}\,du\left(\frac{2.5}{\pi}+o(1)\right) \\
= & \int_{-0.5\pi+o(1)}^{0.5\pi} \frac{1}{1.5+\sin(u)}\,du\left(\frac{2.5}{\pi}+o(1)\right) \\
= & \frac{2}{\sqrt{1.25}}\left(\arctan\left(\frac{1.5\tan(\pi/4)+1}{\sqrt{1.25}}\right)-\arctan\left(\frac{1.5\tan(-\pi/4)+1}{\sqrt{1.25}}\right)\right) \\
& \times\left(\frac{2.5}{\pi}+o(1)\right) \approx 2.236+o(1),
\end{aligned}$$

by use of [23], #2.551(3).

Hence, for every $\epsilon > 0$ and $\delta > 0$ there exists a $k(\delta, \epsilon) \in \mathbb{N}$ such that

$$\int_{b(n_k)\exp\left(y\frac{a(n_k)}{cb(n_k)}\right)}^{b(n_k)} \frac{1}{f(u)}\,du \geq -(2.2-\epsilon)y$$

for all $y \in (y_k - \delta, y_k + \delta)$ with $k > k(\delta, \epsilon)$. The integrand in (2.5.1) is therefore bounded from below by $\exp(-(0.2-\epsilon)y)$ in a region of length 2δ that tends to $-\infty$ as $k \to \infty$. This prevents the convergence of the integral on the right hand side of (2.4.3) as $k \to \infty$. Now, since we have shown that the lim sup in (2.4.2) is infinite, we can argue similarly to the case $c \leq 1/2$ in the proof of Theorem 2.3.2. Namely, the distributions of the power sums converge to a normal distribution along this subsequence. By the Convergence to Types Theorem this prevents convergence to F_c for $c > 1/2$. □

Note the analogy of this example to distributions that have a rather light tail but are not in $\mathcal{D}_\infty(\Lambda)$ because of their discreteness (e.g., a binomial distribution). Similar to discrete laws, the tail distribution function given in the example above has a somewhat stairlike appearance since f is an oscillating function.

The next theorem shows that the limit distributions that are not covered by Theorem 2.3.2 are quite similar to the ones studied so far.

Theorem 2.5.3 *Let X_{1n} and $p(n)$ be defined as in Theorem 2.3.2. Assume that there exist $c > 0$, a subsequence $n_k \to \infty$, and norming constants $\hat{a}_c(n_k)$ and $\hat{b}_c(n_k)$ such that a limit distribution function*

$$\hat{F}_c(x) = \lim_{k\to\infty} P\left(\frac{\sum_{i=1}^{n_k} X_i^{cp(n_k)} - \hat{b}_c(n_k)}{\hat{a}_c(n_k)} \leq x\right)$$

exists and is not of the same type as the F_c in Theorem 2.3.2, i.e., there exists no linear transformation $l(\cdot)$ such that $\hat{F}_c(x) = F_c(l(x))$ for all $x \in \mathbb{R}$. Then, $c > 1/2$ and $\hat{F}_c$ is of the same type as $N_{0,1}$ or $G_{1/c} \star N_{0,\sigma^2}$, where $\star$ denotes the convolution of distribution functions.

Proof. Every limit law has to be infinitely divisible and is uniquely determined by its Lévy measure and σ^2. The Lévy measure is defined by $M(\cdot)$ and $N(\cdot)$ in (2.2.3) and (2.2.4). If we choose multiplicative norming constants as in (2.3.5), cf. the proof of Theorem 2.3.2, we have $M(\cdot) \equiv 0$, and $N(\cdot)$ of the form (2.4.1). With asymptotically different norming constants we obtain $M(\cdot) \equiv N(\cdot) \equiv 0$. Apart from the arbitrary location and scale parameters the limiting law is determined by whether it has a normal component (i.e. $\sigma^2 > 0$ in (2.4.8)) or not. To determine this component, consider the lim sup in (2.4.2). If it is infinite, then, similarly to the case $c \leq 1/2$ in the proof of Theorem 2.3.2, we obtain convergence to a normal distribution along a certain subsubsequence; if it equals zero, we obtain convergence to an α-stable distribution; if it is finite and positive, by choosing the corresponding subsubsequence we obtain convergence to an α-stable distribution convoluted with a normal distribution. Hence, the subsequence converges to a distribution of the stated form. □

2.6 The Weibull and Fréchet cases

The limit laws of norms for the Weibull and the Fréchet max domain of attraction are studied in [38]. With the technique stated above we find limit distributions for power sums and are able to simplify some of the proofs in [38].

Let $X_1, X_2, \ldots$ be i.i.d. positive random variables with distribution function $F \in \mathcal{D}_\infty(\Psi_\alpha)$ and upper endpoint $x_\infty < \infty$. Without loss of generality, $x_\infty = 1$. Now, possible norming constants for the sequence of maxima are given by $a(n) = 1 - F^{\leftarrow}(1 - 1/n)$ and $b(n) \equiv 1$ (see [17], Theorem 3.3.12). Similar to (2.3.2) we can conclude that the summands $\xi_{n,k} := (X_k - a(n))/b(n)$ are not infinitesimal. Since $b(n)/a(n) \to \infty$ as $n \to \infty$, a power transformation allows us to analyze sums of the form (2.3.5). We obtain $M(\cdot) \equiv 0$ and

$$N(u) = -\lim_{n\to\infty} n\overline{F^c_n}(u) = \log(\Psi_\alpha(\log(u^{1/c}))) = -\left(-\frac{\log u}{c}\right)^\alpha, \quad u \in (0,1),$$

for the Lévy measure of the limit law of the power sums (2.3.5). To show that condition (2.4.2) is met we make use of a special representation of F:

$$\overline{F}(x) = \overline{c}(x) \exp\left(-\int_0^x \frac{\delta(t)}{1-t}\, dt\right) \tag{2.6.1}$$

for $x \in (0,1)$ with $\delta : \mathbb{R}^+ \to \mathbb{R}^+$, $\delta(t) \to \alpha$ and $\overline{c}(t) \to \overline{c} > 0$ as $t \to 1$ (see [37], Corollary 1.14).

Equation (2.4.2) is again used to derive σ^2. Interchanging limit and integral in

$$\lim_{n\to\infty} np \int_0^\epsilon x^{p-1} \overline{F_n^c}(x)\, dx, \qquad p > 0,$$

is allowed because of dominated convergence: Choose the functions $\overline{c}$ and δ in (2.6.1) such that $|\delta(t) - \alpha| < \varepsilon$ for all $t \in (0,1)$ and some ε with $0 < \varepsilon < \alpha$. Now, we obtain

$$\begin{aligned}
n\overline{F_n^c}(x) &\sim \frac{\overline{c}(x^{1/cp(n)})}{\overline{c}(F^{\leftarrow}(1-1/n))} \exp\left(-\int_{F^{\leftarrow}(1-1/n)}^{x^{1/cp(n)}} \frac{\delta(t)}{1-t}\, dt\right) \\
&\le C \max\left[\exp\left(-(\alpha-\varepsilon)\int_{F^{\leftarrow}(1-1/n)}^{x^{1/cp(n)}} \frac{1}{1-t}\, dt\right),\right. \\
&\qquad \left.\exp\left(-(\alpha+\varepsilon)\int_{F^{\leftarrow}(1-1/n)}^{x^{1/cp(n)}} \frac{1}{1-t}\, dt\right)\right] \\
&= C \max\left\{\exp\left((\alpha-\varepsilon)\log\left(p(n)(1-x^{1/cp(n)})\right)\right),\right. \\
&\qquad \left.\exp\left((\alpha+\varepsilon)\log\left(p(n)(1-x^{1/cp(n)})\right)\right)\right\} \\
&\le C \max\left\{(-\log(x^{1/c}))^{\alpha-\varepsilon}, (-\log(x^{1/c}))^{\alpha+\varepsilon}\right\}.
\end{aligned}$$

It follows from (2.4.2) that $\sigma^2 = 0$.

Theorem 2.6.1 *Let $X_1, X_2, \ldots$ be positive i.i.d. random variables with distribution function $F \in \mathcal{D}_\infty(\Psi_\alpha)$, $\alpha > 0$, and upper endpoint $0 < x_\infty < \infty$. Let*

$$b(n) \equiv x_\infty, \quad a(n) = x_\infty - F^{\leftarrow}(1-1/n) \quad \textit{and} \quad p(n) = b(n)/a(n).$$

Then, there exists a family F_c of distribution functions of the form given below such that for every $c > 0$ there exist norming constants $\hat{a}_c(n), \hat{b}_c(n), n \in \mathbb{N}$, so that

$$F_c(x) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i^{cp(n)} - \hat{b}_c(n)}{\hat{a}_c(n)} \le x\right).$$

The limit distribution functions are of the type $F_c(x) = H_c(x)$, where H_c has characteristic function of the form (2.4.8) with $\sigma^2 = 0$ and $N(u) = -c^{-\alpha}\,(-\log u)^\alpha\, \mathbb{1}_{(0,1)}(u)$.

Corollary 2.6.2 ([38], Theorem 2.2) *Let X_{1n} and $p(n)$ be as in Theorem 2.6.1. Then, for every $c > 0$, there exist norming constants $\tilde{a}_c(n), \tilde{b}_c(n), n \in \mathbb{N}$, such that*

$$\tilde{H}_c(\exp(x)) = \lim_{n\to\infty} P\left(\frac{\|X_{1n}\|_{cp(n)} - \tilde{b}_c(n)}{\tilde{a}_c(n)} \le x\right).$$

Let us conclude by considering the Fréchet max domain of attraction. In contrast to the cases studied above, a purely multiplicative normalization is sufficient for the

sequences of maxima, i.e. $a(n) = F^{\leftarrow}(1-1/n)$ and $b(n) \equiv 0$ (see [17], Theorem 3.3.7). Since $a(n) \to \infty$ as $n \to \infty$, the variables $\xi_{n,k} = (X_k - b(n))/a(n)$ are infinitesimal, here.

Let X_1 be a random variable with distribution function $F \in \mathcal{D}_\infty(\Phi_\alpha)$. Then, the random variable $X_1^c, c > 0$, belongs to the sum domain of attraction of an α-stable distribution (see, e.g., [17], Theorem 2.2.8). Hence, there exist norming constants $\hat{a}_c(n)$ and $\hat{b}_c(n)$ and a limit distribution function F_c such that

$$F_c(x) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i^c - \hat{b}_c(n)}{\hat{a}_c(n)} \le x\right)$$

for all $x \in \mathbb{R}$. If $c < \alpha/2$, the F_c equal the normal distribution function $N_{0,1}$, whereas for $c \ge \alpha/2$ they equal an α-stable distribution function $G_{\alpha/c}$. This connection between limit laws for sums and for maxima of random variables with regularly varying tails is in fact well known and has been studied in [13]. Schlather's result for the limit laws of norms is stated below for completeness.

Corollary 2.6.3 ([38], Theorem 2.3) *Let $X_{1n} = (X_1, X_2, \ldots, X_n)$ where $X_1, X_2, \ldots$ are positive i.i.d. random variables with distribution function $F \in \mathcal{D}_\infty(\Phi_\alpha)$, $\alpha > 0$. There exists a family $\tilde{F}_c$ of distribution functions of the form given below such that for every $c > 0$ there exist norming constants $\tilde{a}_c(n), \tilde{b}_c(n), n \in \mathbb{N}$, so that*

$$\tilde{F}_c(x) = \lim_{n\to\infty} P\left(\frac{\|X_{1n}\|_c - \tilde{b}_c(n)}{\tilde{a}_c(n)} \le x\right).$$

If $0 < c \le \frac{\alpha}{2}$, the $\tilde{F}_c$ equal the standard normal distribution function $N_{0,1}$.
If $\frac{\alpha}{2} < c \le \alpha$, the $\tilde{F}_c$ are given by $\tilde{F}_c(x) = G_{\alpha/c}(cx)$.
If $\alpha < c < \infty$, the $\tilde{F}_c$ are given by $\tilde{F}_c(x) = G_{\alpha/c}(x^c)$.

Note the similarity between the family of laws that we obtained for the Gumbel case and the family for the Fréchet case with $\alpha = 1$. For $c \le 1$ the limit distribution functions $\tilde{F}_c$ are equal for the two cases whereas for $c > 1$ the limit random variables in the Gumbel case are logarithms of the limit random variables in the Fréchet case. For $c \to \infty$ this reflects the relation $\Lambda = \Phi_1 \circ \exp$.

Closing the gap which existed for the Gumbel case we have thus shown that for all three max domains of attraction a family of limit distributions exists which builds a smooth transition from the limit laws for sums to the limit laws for maxima.

Chapter 3

Extremal Behavior of two Connected RDE-like Time Series

3.1 Introduction

The previous chapter dealt with connections between limit laws for sums and limit laws for maxima. While the three classes of extreme value distributions have been studied seperately we have nevertheless been able to notice similarities between the diffrent cases, cf. for example the remark on the families of distributions for the Gumbel and the Fréchet case at the end of the last chapter. In this section we allow for an interplay of heavy-tailed and light-tailed distributions where the latter are used as input for a time series which will show heavy-tailed stationary behavior.

It has already been mentioned in the introduction that heavy-tailed distributions emerge as the stationary solutions to so-called random difference equations (RDEs). In the following we will study univariate RDEs while their multivariate analogs will be analyzed in Chapter 4. To this end, let $(X_t)_{t\in\mathbb{Z}}$ satisfy the equation

$$X_t = A_t X_{t-1} + B_t, \quad t \in \mathbb{Z}, \tag{3.1.1}$$

where $(A_t, B_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in $\mathbb{R}^2$ with (A_t, B_t) independent of $(X_s)_{s<t}$. A well known sufficient condition for the existence of a stationary solution to (3.1.1) is stated in the following theorem.

Theorem 3.1.1 (cf. [22], Lemma 2.2 and Theorem 4.1) *Let* (3.1.1) *be given as above and let there exist a* $\kappa^* > 0$*, such that*

$$E\left(|A_1|^{\kappa^*}\right) = 1,$$

$$E\left(|A_1|^{\kappa^*} \log^+ |A_1|\right) < \infty$$

and let the conditional law of $\log|A_1|$*, given* $A_1 \neq 0$ *be nonarithmetic. If furthermore*

$$E\left(|B_1|^{\kappa^*}\right) < \infty,$$

then there exists a unique stationary solution $(X_t)_{t\in\mathbb{Z}}$ *to* (3.1.1) *and there exist constants* $C_+, C_- \in \mathbb{R}_0^+$ *such that*

$$P(X_t > x) \sim C_+ x^{-\kappa^*}, \quad P(X_t < -x) \sim C_- x^{-\kappa^*}, \quad x \to \infty. \tag{3.1.2}$$

Here, $C_+ = 0$ *or* $C_- = 0$ *indicate that* $P(X_t > x) = o(x^{-\kappa^*})$ *or* $P(X_t < -x) = o(x^{-\kappa^*})$*, respectively, and* $C_- + C_+ > 0$ *if and only if*

$$P(B_1 = (1 - A_1)c) < 1 \tag{3.1.3}$$

for all $c \in \mathbb{R}$*.*

Thus, if $C_+ > 0$ or $C_- > 0$ the upper or, respectively, lower tail of the stationary solution to (3.1.1) is regularly varying with index κ^*.

In this chapter we are going to study the extremal behavior of the process $(X_t)_{t\in\mathbb{Z}}$ and related processes. For this purpose, we are not only interested in the extremal behavior of a single observation at a given time but in the behavior of the whole process, given that we have an extreme observation of, say, X_0. Thus, we are looking at the finite-dimensional distributions of the properly scaled process, conditioning on the event $|X_0| > u$ as $u \to \infty$. Let us, for a start, assume that the conditions of Theorem 3.1.1 together with (3.1.3) are satisfied and that $(X_t)_{t\in\mathbb{Z}}$ is a stationary solution to (3.1.1). The regular variation of the distribution $F^{|X_0|}$ then guarantees that the weak limit

$$\lim_{u\to\infty} \mathcal{L}\left(\frac{|X_0|}{u}\middle| |X_0| > u\right)$$

exists and equals a Pareto distribution with parameter κ^* (Par(κ^*), for short) since

$$\lim_{u\to\infty} P\left(\frac{|X_0|}{u} > x\middle| |X_0| > u\right) = \lim_{u\to\infty} \frac{\overline{F}^{|X_0|}(xu)}{\overline{F}^{|X_0|}(u)} = x^{-\kappa^*}$$

for all $x > 1$. Using the constants C_- and C_+ we may also analyze the behavior of the upper and the lower tail of X_0 separately, namely

$$\lim_{u\to\infty} P\left(\frac{X_0}{u} > x\middle| |X_0| > u\right) = \frac{C_+}{C_- + C_+} x^{-\kappa^*}$$

and

$$\lim_{u\to\infty} P\left(\frac{X_0}{u} < -x\middle| |X_0| > u\right) = \frac{C_-}{C_- + C_+} x^{-\kappa^*}$$

which leads to

$$\mathcal{L}\left(\frac{X_0}{u}\middle| |X_0| > u\right) \xrightarrow{w} \mathcal{L}(M \cdot X),$$

for $u \to \infty$, where M and X are independent random variables with

$$P(M = -1) = \frac{C_-}{C_- + C_+}, \quad P(M = 1) = \frac{C_+}{C_- + C_+},$$

and $X \sim \mathrm{Par}(\kappa^*)$.

Analyzing the process following an extreme observation at time zero we may apply the continuous mapping theorem to derive that for $n \in \mathbb{N}$:

$$\mathcal{L}\left(\frac{X_0}{u}, \frac{X_1}{u}, \ldots, \frac{X_n}{u} \middle| |X_0| > u\right) \xrightarrow{w} \mathcal{L}\left(MX, MX\tilde{A}_1, \ldots, MX\prod_{i=1}^{n}\tilde{A}_i\right), \tag{3.1.4}$$

for $u \to \infty$ with random variables M, X as above and $\tilde{A}_i, i = 1, \ldots, n$, i.i.d. and independent of (M, X) with the same distribution as A_1. Obviously, just like the original process defined in (3.1.1), the limit process in (3.1.4) is a Markov chain and since it describes the extremal behavior of the process it is called the "tail chain" of the process (cf. [40]). It was shown in [39] that the "forward" tail chain in (3.1.4) can be complemented by a "backward" tail chain which describes the conditional distribution of the properly scaled process *before* the extreme event at time zero. Again, this process has a simple Markovian structure. Segers' approach now amplifies the analysis of RDEs to processes which behave asymptotically like RDEs in the sense that, roughly speaking, $X_{t+1} \sim X_t A_t$ for large values of X_t, where A_t is a random variable independent of X_t. For comparison and motivation we will recapture some of Segers' results, which we will sometimes slightly modify where it is useful for our later extensions.

3.2 Segers' results for a single RDE-like time series

Let us consider the real-valued stochastic process $(Y_t)_{t\in\mathbb{Z}}$ defined by the recursive equation

$$Y_t = \Phi(Y_{t-1}, \epsilon_t), \quad t \in \mathbb{Z}, \tag{3.2.1}$$

where

$$\begin{aligned} &\text{(i)} \ (\epsilon_t)_{t\in\mathbb{Z}} \in \mathbb{R}^d \text{ are i.i.d. random vectors and } \epsilon_t, t \in \mathbb{Z}, \text{ is independent} \\ &\quad \text{of } (Y_s)_{s<t}, \\ &\text{(ii)} \ \Phi \text{ is a measurable function from } \mathbb{R} \times \mathbb{R}^d \text{ to } \mathbb{R}. \end{aligned} \tag{3.2.2}$$

Since we leave the framework of RDEs we cannot deduce the existence of a stationary solution to (3.2.1), furthermore, if one exists, we do not know about its regular variation properties. Therefore, we will assume that a stationary solution to (3.2.1) exists and fullfills the following two conditions.

Condition 3.1 a) *There exists* $0 < \kappa < \infty$ *such that*

$$\lim_{y\to\infty} \frac{P(|Y_0| > yx)}{P(|Y_0| > y)} = x^{-\kappa}, \quad 0 < x < \infty.$$

Moreover, there exists $0 \le p \le 1$ *such that*

$$\lim_{y\to\infty} \frac{P(Y_0 > y)}{P(|Y_0| > y)} = p.$$

Condition 3.2 a) *There exists a function* $\phi : \mathbb{R}^d \times \{-1, 1\} \to \mathbb{R}$ *such that*

$$\lim_{y\to\infty} \frac{\Phi(yw(y), v(y))}{y} = w\phi(v, \text{sign}(w))$$

for all $w(y) \to w \neq 0, v(y) \to v \in \mathbb{R}^d$ *and*

$$\lim_{y\to\infty} \frac{\Phi(yw(y), v(y))}{y} = 0$$

for $w(y) \to 0, v(y) \to v \in \mathbb{R}^d$.

To be in line with our subsequent modifications on the theorems the assumptions on Φ are slightly stronger than the ones originally formulated in [39]. This also motivates the assumption $\epsilon_t \in \mathbb{R}^d$ instead of the broader setting of Segers where the ϵ_t are just assumed to be random elements of a measurable space. The motivation for Condition 3.2 a) becomes clear by writing

$$\lim_{y\to\infty} \mathcal{L}\left(\frac{|Y_0|}{y}, \frac{Y_0}{|Y_0|}, \frac{Y_1}{|Y_0|} \middle| |Y_0| > y\right) = \lim_{y\to\infty} \mathcal{L}\left(\frac{|Y_0|}{y}, \frac{Y_0}{|Y_0|}, \frac{\Phi\left(|Y_0|\frac{Y_0}{|Y_0|}, \epsilon_1\right)}{|Y_0|} \middle| |Y_0| > y\right),$$

where the condition admits the application of the continuous mapping theorem (cf. Theorem 4.27, [27]). Induction then allows to derive the so-called "forward tail chain" of the process $(Y_t)_{t\in\mathbb{N}_0}$, cf. [39], pp. 4-5, for details of the proof.

Proposition 3.2.1 (cf. [39], Theorem 2.3) *Let* $(Y_t)_{t\in\mathbb{Z}}$ *(not necessarily stationary) be given by (3.2.1)-(3.2.2) and let Conditions 3.1 a) and 3.2 a) hold. Then for* $t \in \mathbb{N}$, *as* $y \to \infty$,

$$\mathcal{L}\left(\frac{|Y_0|}{y}, \frac{Y_0}{|Y_0|}, \ldots, \frac{Y_t}{|Y_0|} \middle| |Y_0| > y\right) \xrightarrow{w} \mathcal{L}(Y, M_0, \ldots, M_t), \tag{3.2.3}$$

with

$$M_j = h(M_{j-1}, A_j, B_j), \quad j = 1, 2, \ldots,$$

where $h : \mathbb{R}^3 \to \mathbb{R}$, $h(y, a, b) := y\left(a\mathbb{1}_{(0,\infty)}(y) + b\mathbb{1}_{(-\infty,0)}(y)\right)$ *and* $Y, M_0, (A_1, B_1), (A_2, B_2), \ldots$ *are independent with*

(i) $Y \sim Par(\kappa)$,

(ii) $P(M_0 = 1) = p = 1 - P(M_0 = -1)$,

(iii) $\mathcal{L}(A_i, B_i) = \mathcal{L}(\phi(\epsilon_t, 1), \phi(\epsilon_t, -1)), i \in \mathbb{N}$.

If $(Y_t)_{t\in\mathbb{Z}}$ is additionally assumed to be stationary, there also exists a so-called "backward tail chain" which has a surprisingly simple form as well.

Proposition 3.2.2 (cf. [39], Theorem 5.2) *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process given by (3.2.1)-(3.2.2) and let Conditions 3.1 a) and 3.2 a) hold. Then for all $s,t \in \mathbb{N}$, as $y \to \infty$,*

$$\mathcal{L}\left(\frac{|Y_0|}{y}, \frac{Y_{-s}}{|Y_0|}, \dots, \frac{Y_t}{|Y_0|} \,\middle|\, |Y_0| > y\right) \xrightarrow{w} \mathcal{L}(Y, M_{-s}, \dots, M_t), \tag{3.2.4}$$

with

(i) $Y \sim Par(\kappa)$, independent of $(M_t)_{t\in\mathbb{Z}}$,

(ii) $(M_t)_{t\in\mathbb{Z}}$ is a BFTC(κ,μ) where $\mu = \mathcal{L}(M_0, M_1)$ with (M_0, M_1) as in Proposition 3.2.1.

The expression BFTC stands for back-and-forth tail chain and is defined as follows

Definition 3.2.3 (cf. [39], Definition 4.1)**.** A discrete-time process $(M_t)_{t\in\mathbb{Z}}$ is said to be a back-and-forth tail chain with index $0 < \kappa < \infty$ and forward transition law μ, notation BFTC(κ,μ), if

(i) $\mathcal{L}(M_0, M_1) = \mu$,

(ii) $\mu^* := \mathcal{L}(M_0, M_{-1})$ is adjoint to μ, meaning that

$$E[(xM_0)_+^\kappa \wedge (yM_1)_+^\kappa] = E[(xM_{-1})_+^\kappa \wedge (yM_0)_+^\kappa], \quad \forall\, x, y \in \mathbb{R}, \tag{3.2.5}$$

(iiia) for all integer $t \geq 1$ and all real $x_{t-1}, x_{t-2}, \dots,$

$$\mathcal{L}(M_t | M_{t-1} = x_{t-1}, M_{t-2} = x_{t-2}, \dots) = \mathcal{L}(h(x_{t-1}, A_1, B_1)),$$

where A_1 and B_1 are independent with

$$\mathcal{L}(A_1) = \mathcal{L}\left(\frac{M_1}{M_0} \,\middle|\, M_0 = 1\right), \quad \mathcal{L}(B_1) = \mathcal{L}\left(\frac{M_1}{M_0} \,\middle|\, M_0 = -1\right).$$

(iiib) for all integer $t \geq 1$ and all real $x_{-t+1}, x_{-t+2}, \dots,$

$$\mathcal{L}(M_{-t} | M_{-t+1} = x_{-t+1}, M_{-t+2} = x_{-t+2}, \dots) = \mathcal{L}(h(x_{-t+1}, A_{-1}, B_{-1}),$$

where A_{-1} and B_{-1} are independent with

$$\mathcal{L}(A_{-1}) = \mathcal{L}\left(\frac{M_{-1}}{M_0} \,\middle|\, M_0 = 1\right), \quad \mathcal{L}(B_{-1}) = \mathcal{L}\left(\frac{M_{-1}}{M_0} \,\middle|\, M_0 = -1\right).$$

While this definition seems to be a bit "ad hoc" at first sight, it can be shown that the definition of the adjoint measure as in (3.2.5) and the definition of the BFTC as in 3.2.3 guarantees that the following proposition holds.

Proposition 3.2.4 (cf. [39], Proposition 4.2) *Let $(M_t)_{t\in\mathbb{Z}}$ be a BFTC(κ,μ). For all integers $s \geq 1$ and $t \geq 0$ and for every bounded, measurable function $f : \mathbb{R}^{s+t+1} \to \mathbb{R}$ such that $f(x_{-s},\dots,x_t) = 0$ as soon as $x_{-s} = 0$, the numbers*

$$E\left[f\left(\frac{M_{-s+i}}{|M_i|},\dots,\frac{M_{t+i}}{|M_i|}\right)|M_i|^\kappa\right], \quad i = 0,\dots,s$$

are all the same.

It can be shown that a process with this property has the same structure as the limit law

$$\lim_{y\to\infty}\mathcal{L}\left(\frac{Y_{-s}}{|Y_0|},\dots,\frac{Y_t}{|Y_0|}\,\middle|\,|Y_0|>y\right).$$

Since for a given measure μ it is not clear from the definition if the adjoint measure μ^* actually exists it is useful to know that the unique existence of μ^* is guaranteed if μ is the distribution of a random vector (M_0, M_1) such that

(i) $P(M_0 = 1) = p = 1 - P(M_0 = -1)$ and

(ii) $E\left[(\sigma M_1)_+^\kappa\right] \leq P(M_0 = \sigma)$ for $\sigma \in \{-1, 1\}$ (cf. [39], Proposition 3.1).

Since Equation (3.2.5) is quite unhandy to derive μ^* from μ for practical purposes (for example for simulations), we may use that the relation

$$\begin{aligned} & E\left[\mathbb{1}_{\{M_0^*=\sigma\}}f(M_1^*)\right] \\ = \;& E\left[f(M_0/|M_1|)(\sigma M_1)_+^\kappa\right] + \left(P(M_0=\sigma) - E((\sigma M_1)_+^\kappa)\right)f(0) \end{aligned}$$

holds for all M_1^*-integrable functions f and for $\sigma \in \{-1, 1\}$ with $\mathcal{L}(M_0, M_1) = \mu$ and $\mathcal{L}(M_0^*, M_1^*) = \mu^*$ (cf. [39], Proposition 3.1).

Taken together, Propositions 3.2.1 and 3.2.2 show that the limit distribution

$$\lim_{y\to\infty}\mathcal{L}\left(\left(\frac{Y_0}{y},\frac{Y_{-s}}{|Y_0|},\dots,\frac{Y_t}{|Y_0|}\right)\,\middle|\,|Y_0|>y\right)$$

of the scaled conditional process has a very simple structure, where multiplication with independent factors leads us to the next value of the chain. We may think of the tail chain as a process consisting of two components: The scale of the process is determined by a Pareto distributed random variable Y while its relative behavior is determined by a multiplicative random walk which solely depends on $p \in [0, 1]$ and the distribution of $\phi(\epsilon_t, \pm 1)$. Many applications of this and related models (see, for example, Section 3.7) imply a light-tailed distribution of $\phi(\epsilon_t, -1)$ and $\phi(\epsilon_t, 1)$. Thus, while the marginal distribution of Y_t is heavy-tailed and, as is shown in Section 6 of [39], the stationary process $(Y_t)_{t\in\mathbb{Z}}$ given by (3.2.1)-(3.2.2) is regularly varying if it satisfies Conditions 3.1 a) and 3.2 a), its extreme value behavior is strongly influenced by the light-tailed distributions of $\phi(\epsilon, \pm 1)$.

3.3 Extension to two connected time series

We will now extend the setting of [39] to the scenario of two connected time series. When modelling processes it is sometimes convenient to use two separate time series which are connected in a simple way. This holds especially true for financial time series - think for example of the ARCH ("autoregressive conditional heteroscedasticity") and GARCH ("generalized ARCH") modelling where the price of an asset is driven by an underlying volatility process. A GARCH(p, q) process with $p, q \in \mathbb{N}_0$ consists of two time series $(X_t)_{t\in\mathbb{Z}}$ and $(\sigma_t)_{t\in\mathbb{Z}}$ which satisfy the following two equations:

$$X_t = \sigma_t \epsilon_{t+1}, \quad t \in \mathbb{Z}, \tag{3.3.1}$$

with $\epsilon_t, t \in \mathbb{Z}$, i.i.d., ϵ_t independent of $(X_{s-1}, \sigma_s)_{s<t}$ and $E(\epsilon_t) = 0, \mathrm{Var}(\epsilon_t) = 1$ (often, the $\epsilon_t, t \in \mathbb{Z}$, are assumed to be standard normally distributed). The so-called "volatility process" $(\sigma_t)_{t\in\mathbb{Z}} \geq 0$ is given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \tag{3.3.2}$$

with $\alpha_i, \beta_j \geq 0$ and $\alpha_p \beta_q \neq 0$. A GARCH$(p, 0)$ model corresponds to an ARCH(p) model. Under certain assumptions about the distribution of ϵ_t and the values of the α_i's and β_j's one can show that a stationary solution to (3.3.1) and (3.3.2) exists. For example, a sufficient condition is given by $\sum_{i=1}^{p} \alpha_i + \sum_{j=i}^{q} \beta_j < 1$ (cf. [1], p. 190). Compare [8] and [41] for the origins of GARCH(p, q) models and [1] for a compound treatment of their characteristics. See also Chapter 4 where we will come back to regular variation properties of GARCH(p, q) processes.

One can interprete these two time series as a visible one ($(X_t)_{t\in\mathbb{Z}}$, the price of the asset) and a hidden one, which cannot be observed directly ($(\sigma_t)_{t\in\mathbb{Z}}$, the volatility). Figure 3.1 shows a part of the S&P 500 time series and the estimated corresponding volatility. While the GARCH(p, q) models have become extremely popular tools in financial mathematics the most used specification is the simplest form of the model, the GARCH$(1, 1)$ process.

The aim of our extensions of Segers' theorems is to analyze the joint behavior of two connected time series like the ones of a GARCH$(1, 1)$ process with the help of a tail chain approach. Motivated by the setting of a visible process (further denoted by $(X_t)_{t\in\mathbb{Z}}$) and an underlying process (further denoted by $(Y_t)_{t\in\mathbb{Z}}$) we are interested in the behavior of the underlying process, given that we have an extreme observation of the visible process. We will assume that the underlying process satisfies the assumptions made in [39] just like the volatility process of a GARCH$(1, 1)$ model does.

Therefore, let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary time series given by (3.2.1)-(3.2.2) and let Conditions 3.1 a) and 3.2 a) be satisfied. Furthermore, let

$$X_t = \Psi(Y_t, \epsilon_{t-s_-}, \ldots, \epsilon_{t+s_+}), \quad t \in \mathbb{Z}, \tag{3.3.3}$$

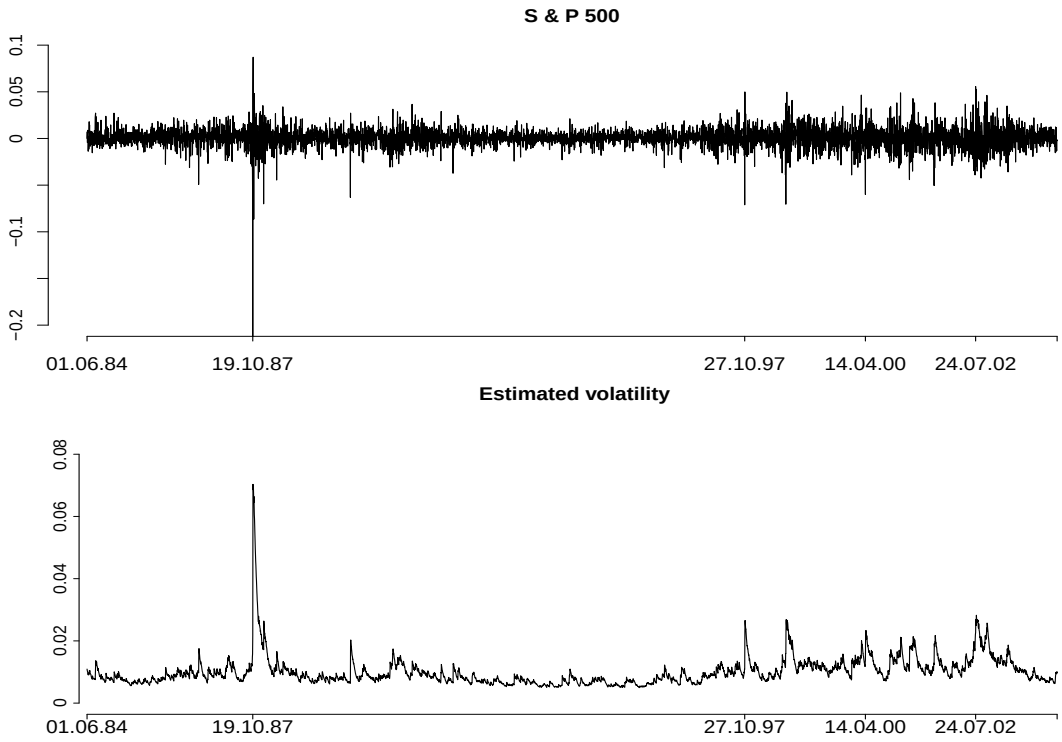


Figure 3.1: S&P 500 index and estimated volatility using a GARCH$(1,1)$ model and [43].

with $-s_- - 1 \leq 0 \leq s_+$ and measurable $\Psi : \mathbb{R}^{s_-+s_++2} \to \mathbb{R}$. Thus, the visible process $(X_t)_{t\in\mathbb{Z}}$ depends on the underlying process $(Y_t)_{t\in\mathbb{Z}}$ and both share the innovations $\epsilon_t, t \in \mathbb{Z}$, as a source for random development. Note that the recursive definition of Y_t would allow us to find a function $\tilde{\Psi} : \mathbb{R}^{s_-+s_++2} \to \mathbb{R}$ such that $X_t = \tilde{\Psi}(Y_{t-s_-}, \epsilon_{t-s_-}, \ldots, \epsilon_{t+s_+})$. This would allow us to set $\tilde{s}_+ = s_- + s_+$ and $\tilde{s}_- = 0$, but the original definition is often more suitable to examples and will therefore be kept.

An example for time series described by (3.2.1)-(3.2.2) and (3.3.3) is given by a GARCH(1,1) process, which has the simple form

$$X_t = \sigma_t \epsilon_{t+1}, \quad t \in \mathbb{Z}, \tag{3.3.4}$$

(for technical reasons we will mainly look at

$$X_t^2 = \sigma_t^2 \epsilon_{t+1}^2, \quad t \in \mathbb{Z},$$

instead) and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 \epsilon_t^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z}, \tag{3.3.5}$$

with suitable constants $\alpha_0, \alpha_1, \beta_1 > 0$ and $\epsilon_t, t \in \mathbb{Z}$, i.i.d. with $E(\epsilon_t) = 0, \mathrm{Var}(\epsilon_t) = 1$. Taking $(\sigma_t^2)_{t\in\mathbb{Z}}$ as $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t^2)_{t\in\mathbb{Z}}$ as $(X_t)_{t\in\mathbb{Z}}$ this example satisfies equations (3.2.1)-(3.2.2) and (3.3.3) with

$$\Phi(x, e) = \alpha_0 + \alpha_1 x e^2 + \beta_1 x, \quad \Psi(x, e) = xe, \quad s_- = -1, \ s_+ = 1.$$

We will assume that Ψ satisfies a condition analogue to 3.2 a), where we have to keep in mind that $(X_t)_{t\in\mathbb{Z}}$ is in general not a Markov chain like $(Y_t)_{t\in\mathbb{Z}}$.

Condition 3.2 b) *There exists a function* $\psi : \mathbb{R}^{s_-+s_++1} \times \{-1,1\} \to \mathbb{R}$ *such that*

$$\lim_{y\to+\infty} \frac{\Psi(yw(y), v(y))}{y} = w\psi(v, \operatorname{sign}(w))$$

for all $w(y) \to w \neq 0, v(y) \to v \in \mathbb{R}^{s_-+s_++1}$ *and*

$$\lim_{y\to+\infty} \frac{\Psi(yw(y), v(y))}{y} = 0$$

for $w(y) \to 0, v(y) \to v \in \mathbb{R}^{s_-+s_++1}$.

Furthermore, we will assume that X_0 and Y_0 are tail equivalent.

Condition 3.1 b) *There exists* $C > 0$ *such that*

$$\lim_{x\to\infty} \frac{P(|X_0| > x)}{P(|Y_0| > x)} = C.$$

If it is possible to find two independent random variables $\tilde{Y}$ and $\tilde{Z}$, such that $|Y_0| \stackrel{d}{=} \tilde{Y}$ and $|X_0| \stackrel{d}{=} \tilde{Y}\tilde{Z}$ and if $0 < E(|\tilde{Z}|^{\kappa+\varepsilon}) < \infty$ holds for an $\varepsilon > 0$, then Condition 3.1 b) follows from Condition 3.1 a) as a consequence of Breiman's Lemma with $C = E(|\tilde{Z}|^{\kappa})$ (cf. [12]). Of course, Conditions 3.1 a) and 3.1 b) imply that

$$\lim_{x\to\infty} \frac{P(|X_0| > yx)}{P(|X_0| > x)} = y^{-\kappa}, \quad 0 < y < \infty.$$

Given these assumptions we now want to derive the form of the limit law

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_t}{x} \,\middle|\, |X_0| > x\right), \tag{3.3.6}$$

for $s, t \in \mathbb{N}$. Now, the difficulty consists in the interconnection of the time series $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$. A large observation of X_0 depends on Y_0 but also on $(\epsilon_{-s_-}, \ldots, \epsilon_{s_+})$. But then these ϵ_t's also directly influence the values of $Y_{-s_-}, \ldots, Y_{s_+}$. Therefore, it has to be evaluated if a tail chain approach is suitable for the analysis of (3.3.6), and, if so, which parts of the time series $(Y_t)_{t\in\mathbb{Z}}$ can be treated with this approach. We will show that the tail chain approach is suitable for $Y_t, t < s_- - 1$, and $Y_t, t > s_+$, and split up the time series accordingly. For further investigations, we have to assume that the conditional limit distribution of $(Y_{-s_--1}, \ldots, Y_{s_+})$ exists.

Condition 3.3 *There exists a random vector* $(Y^X_{-s_--1}, \ldots, Y^X_0, \ldots, Y^X_{s_+})$ *such that*

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--1}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x} \,\middle|\, |X_0| > x\right) = \mathcal{L}(Y^X_{-s_--1}, \ldots, Y^X_0, \ldots, Y^X_{s_+}). \tag{3.3.7}$$

From a theoretical point of view one can show that the family of distributions defined by the left hand side of (3.3.7) and indexed by x is tight if Conditions 3.1 a) and 3.1 b) are fulfilled and if $(Y_t)_{t\in\mathbb{Z}}$ is stationary, therefore a weak accumulation point must exist but is not necessarily unique. From a practical point of view the verification of Condition 3.3 may get cumbersome, therefore we provide some Lemmata to facilitate this task. To start with, it is often easier and sufficient to check whether the limit of the conditional distribution of $(Y_{-s_--1}, \epsilon_{-s_-}, \dots, \epsilon_{s_+})$, given that we have an extreme observation of $|X_0|$, exists:

Lemma 3.3.1 *Let $(Y_t, X_t)_{t\in\mathbb{Z}}$ be given by (3.2.1)-(3.2.2) and (3.3.3) and let Condition 3.2 a) be satisfied. If there exists a random vector $(\hat{Y}^X_{-s_--1}, \epsilon^X_{-s_-}, \dots, \epsilon^X_{s_+})$ such that*

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--1}}{x}, \epsilon_{-s_-}, \dots, \epsilon_{s_+} \,\middle|\, |X_0| > x\right) = \mathcal{L}(\hat{Y}^X_{-s_--1}, \epsilon^X_{-s_-}, \dots, \epsilon^X_{s_+}), \tag{3.3.8}$$

then Condition 3.3 is satisfied with

$$\mathcal{L}\left(Y^X_{-s_--1}, \dots, Y^X_{s_+}\right) = \mathcal{L}\left(\hat{Y}^X_{-s_--1}, \dots, \hat{Y}^X_{s_+}\right), \tag{3.3.9}$$

where

$$\hat{Y}^X_{-s_--1+n} = \hat{Y}^X_{-s_--1+n-1}\phi\left(\epsilon^X_{-s_--1+n}, \operatorname{sign}(\hat{Y}^X_{-s_--1+n-1})\right) \tag{3.3.10}$$

for all $n \in \{1, \dots, s_- + s_+ + 1\}$.

Proof. Let (3.3.8) hold. We will analyze the joint limit behavior of Y_t and $\epsilon_t, t \in \mathbb{Z}$, and show that for all $i \in \{0, \dots, s_- + s_+ + 1\}$

$$\begin{aligned} &\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{-s_--1+i}}{x}, \epsilon_{-s_-}, \dots, \epsilon_{s_+} \,\middle|\, |X_0| > x\right) \\ = \;& \mathcal{L}(\hat{Y}^X_{-s_--1}, \dots, \hat{Y}^X_{-s_--1+i}, \epsilon^X_{-s_-}, \dots, \epsilon^X_{s_+}) \end{aligned} \tag{3.3.11}$$

where $\hat{Y}^X_{-s_--1+n}$ is defined as in (3.3.10) for $0 \le n \le i$. The claim then follows for $i = s_- + s_+ + 1$, the proof is by induction on i. For a start, if $i = 0$, then (3.3.11) follows from (3.3.8).

For the induction step let (3.3.11) hold for $i \in \{0, \dots, s_- + s_+\}$ and set for abbreviation $\boldsymbol{\epsilon} = (\epsilon_{-s_-}, \dots, \epsilon_{s_+})$ and $\boldsymbol{\epsilon}^X = (\epsilon^X_{-s_-}, \dots, \epsilon^X_{s_+})$, then

$$\begin{aligned} &\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{-s_-+i}}{x}, \boldsymbol{\epsilon} \,\middle|\, |X_0| > x\right) \\ = \;& \lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{-s_-+i-1}}{x}, \frac{\Phi\left(x\frac{Y_{-s_-+i-1}}{x}, \epsilon_{-s_-+i}\right)}{x}, \boldsymbol{\epsilon} \,\middle|\, |X_0| > x\right) \\ = \;& \lim_{x\to\infty} \mathcal{L}\left(g_x\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{-s_-+i-1}}{x}, \boldsymbol{\epsilon}\right) \,\middle|\, |X_0| > x\right), \end{aligned}$$

with

$$\begin{aligned} & g_x(y_{-s_--1},\dots,y_{-s_-+i-1},\mathbf{e}) \\ := & \left(y_{-s_--1},\dots,y_{-s_-+i-1},\frac{\Phi\left(xy_{-s_-+i-1},e_{-s_-+i}\right)}{x},\mathbf{e}\right) \end{aligned}$$

with $\mathbf{e}=(e_{-s_-},\dots,e_{s_+})$. As x tends to infinity, this function converges pointwise to

$$\begin{aligned} & g(y_{-s_--1},\dots,y_{-s_-+i-1},\mathbf{e}) \\ := & \left(y_{-s_--1},\dots,y_{-s_-+i-1},y_{-s_-+i-1}\phi(e_{-s_-+i},\operatorname{sign}(y_{-s_-+i-1})),\mathbf{e}\right) \end{aligned}$$

because of the assumptions made in Condition 3.2 a). Together with the induction hypothesis this lets us apply a version of the continuous mapping theorem (Theorem 4.27, [27]),

$$\begin{aligned} & \lim_{x\to\infty}\mathcal{L}\left(\frac{Y_{-s_--1}}{x},\dots,\frac{Y_{-s_-+i}}{x},\boldsymbol{\epsilon}\,\middle|\,|X_0|>x\right) \\ = & \lim_{x\to\infty}\mathcal{L}\left(g_x\left(\frac{Y_{-s_--1}}{x},\dots,\frac{Y_{-s_-+i-1}}{x},\boldsymbol{\epsilon}\right)\,\middle|\,|X_0|>x\right) \\ = & \mathcal{L}\left(g(\hat{Y}^X_{-s_--1},\dots,\hat{Y}^X_{-s_-+i-1},\boldsymbol{\epsilon}^X)\right) \\ = & \mathcal{L}\left(\hat{Y}^X_{-s_--1},\dots,\hat{Y}^X_{-s_-+i-1},\hat{Y}^X_{-s_-+i-1}\phi(\epsilon^X_{-s_-+i},\operatorname{sign}(\hat{Y}^X_{-s_-+i-1})),\boldsymbol{\epsilon}^X\right). \end{aligned}$$

This proves (3.3.11) for all $i\in\{1,\dots,s_-+s_++1\}$ and shows that Condition 3.3 is satisfied. □

The next lemma might be helpful to derive the distribution in (3.3.8).

Lemma 3.3.2 *Let $(Y_t,X_t)_{t\in\mathbb{Z}}$ be given by (3.2.1)-(3.2.2) and (3.3.3), stationary, and let Conditions 3.1 a) and b) be satisfied. Let $\boldsymbol{\epsilon},\boldsymbol{\epsilon}^X$ and $\mathbf{e}$ stand for the same vectors as in the proof of Lemma 3.3.1. If there exist functions f_+ and f_-, both mapping $\mathbb{R}^{s_-+s_++1}\times\mathbb{R}_+$ to $[0,1]$ such that*

$$\lim_{x\to\infty}P\left(\frac{|X_0|}{x}>1\,\middle|\,\boldsymbol{\epsilon}=\mathbf{e},\frac{Y_{-s_--1}}{x}>y\right) = f_+(\mathbf{e},y)$$

and

$$\lim_{x\to\infty}P\left(\frac{|X_0|}{x}>1\,\middle|\,\boldsymbol{\epsilon}=\mathbf{e},\frac{Y_{-s_--1}}{x}<-y\right) = f_-(\mathbf{e},y)$$

for all $(\mathbf{e},y)\in\mathbb{R}^{s_-+s_++1}\times\mathbb{R}_+$, then (3.3.8) is satisfied with

$$P\left(\hat{Y}^X_{-s_--1}>y,\boldsymbol{\epsilon}^X\in A\right)=\frac{p}{C}y^{-\kappa}\int_A f_+(\mathbf{e},y)\,P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e}) \tag{3.3.12}$$

and

$$P\left(\hat{Y}^X_{-s_--1} < -y, \boldsymbol{\epsilon}^X \in A\right) = \frac{1-p}{C} y^{-\kappa} \int_A f_-(\mathbf{e}, y)\, P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e}), \tag{3.3.13}$$

for all $y > 0, A \in \mathbb{B}^{s_-+s_++1}$.

Proof. To see that (3.3.12) holds under the above circumstances, let us write

$$\begin{aligned}
& \lim_{x\to\infty} P\left(\frac{Y_{-s_--1}}{x} > y, \boldsymbol{\epsilon} \in A \,\middle|\, |X_0| > x\right) \\
= \quad & \lim_{x\to\infty} \frac{P\left(\frac{Y_{-s_--1}}{x} > y, \boldsymbol{\epsilon} \in A, |X_0| > x\right)}{P(|X_0| > x)} \\
= \quad & \lim_{x\to\infty} \frac{P\left(Y_{-s_--1} > xy\right) \int_A P(|X_0| > y \,|\boldsymbol{\epsilon} = \mathbf{e}, Y_{-s_--1} > xy) P^{\boldsymbol{\epsilon}\,|Y_{-s_--1}>xy}(\mathrm{d}\mathbf{e})}{P(|X_0| > x)} \\
\overset{\boldsymbol{\epsilon}, Y_{-s_--1}\text{ind.}}{=} \quad & \lim_{x\to\infty} \frac{P\left(Y_{-s_--1} > xy\right) \int_A P(|X_0| > y \,|\boldsymbol{\epsilon} = \mathbf{e}, Y_{-s_--1} > xy) P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e})}{P(|X_0| > x)} \\
\overset{\text{stationarity}}{=} \quad & \lim_{x\to\infty} \frac{P\left(Y_0 > xy\right) \int_A P(|X_0| > y \,|\boldsymbol{\epsilon} = \mathbf{e}, Y_{-s_--1} > xy) P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e})}{P(|X_0| > x)} \\
\overset{\text{Domin. conv.}}{=} \quad & \lim_{x\to\infty} \frac{P\left(Y_0 > xy \,\middle|\, |Y_0| > xy\right) P(|Y_0| > xy)}{P(|X_0| > x)} \int_A f_+(\mathbf{e}, y) P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e}) \\
\overset{\text{3.1 a), b)}}{=} \quad & \frac{p}{C} y^{-\kappa} \int_A f_+(\mathbf{e}, y) P^{\boldsymbol{\epsilon}}(\mathrm{d}\mathbf{e}).
\end{aligned}$$

Equation (3.3.13) follows analogously. □

Note that Lemma 3.3.2 gives a complete characterization of $(Y^X_{-s_--1}, \epsilon^X_{-s_-}, \ldots, \epsilon^X_{s_+})$ only in the case that $P(Y^X_{-s_--1} = 0) = 0$.

3.4 Multivariate regular variation

Although Lemmas 3.3.1 and 3.3.2 give a notion when Condition 3.3 is satisfied, this condition looks like a very strong assumption. Nevertheless, it is closely related to the theory of multivariate regular variation which is quite well explored for large classes of common time series models. From the equivalent definitions of multivariate regular variation which are known, we choose the same as is used in [39], although we will use the slightly more general concept of multivariate regular variation on a cone instead of multivariate regular variation on $\bar{\mathbb{R}}^d$ (cf. [36], Chapter 6). Let $\mathbb{C}$ be a cone, i.e. a non empty subset of $\bar{\mathbb{R}}^d$, such that $\mathbf{x} \in \mathbb{C}$ implies that $t\mathbf{x} \in \mathbb{C}$ for all $t > 0$. We call a random vector $\mathbf{Y} \in \mathbb{R}^d$ multivariate regularly varying with index $\kappa > 0$ on $\mathbb{C}$ if there exists a univariate regularly varying function $U : \mathbb{R}_+ \to \mathbb{R}_+$ with index $-\kappa$ (cf. (1.1.1)) and a non-degenerate, non-zero Radon measure ν on $\mathbb{C}$ such that

$$P\left(\mathbf{Y} \in x\cdot\right) / U(x) \xrightarrow{v} \nu(\cdot), \qquad x \to \infty, \tag{3.4.1}$$

where $\xrightarrow{v}$ stands for vague convergence (cf. [36], p. 49) in $M_+(\mathbb{C})$, the space of all non-negative Radon measures on $\mathbb{C}$. The choice of $\mathbb{C}$ is often crucial for the analysis of regular variation, cf. [36], Chapter 9.4. Popular choices for $\mathbb{C}$ are $\mathbb{C} = \bar{\mathbb{R}}^d \setminus \{\mathbf{0}\}$ or $\mathbb{C} = \bar{\mathbb{R}}_+^d \setminus \{\mathbf{0}\}$, in the following we will choose $\mathbb{C} = \left(\bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^d\right) \setminus \{\mathbf{0}\}$. Excluding the zero implies that the compact sets in $\mathbb{C}$ are those which are bounded away from $\mathbf{0}$ (cf. [36], Proposition 6.1), so if we want to prove multivariate regular variation of a random vector $\mathbf{Y}$ on $\mathbb{C}$ we have to check if the relation

$$\lim_{x\to\infty} \frac{P(\mathbf{Y} \in xA)}{U(x)} = \nu(A)$$

holds for all $A \subset \mathbb{C}$ bounded away from $\mathbf{0}$ with $\nu(\partial A) = 0$. One can show that the limit measure ν is necessarily homogeneous, meaning that

$$\nu(xA) = x^{-\kappa}\nu(A)$$

holds for all $x > 0, A \subset \mathbb{C}$ (cf. [36], p. 168).

In this section we will analyze how Condition 3.3 is connected to multivariate regular variation of the vector $(|X_0|, Y_{-s_--1}, \ldots, Y_{s_+})$ and especially if one implies the other. We will see that although the characteristics look similar, without further assumptions neither of them implies the other.

Let us first assume that $(Y_t, X_t)_{t\in\mathbb{Z}}$ is stationary and given by (3.2.1)-(3.2.2) and (3.3.3). Further, let us assume that $(|X_0|, Y_{-s_--1}, \ldots, Y_{s_+})$ is multivariate regularly varying on $\mathbb{C} = \left(\bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^{s_-+s_++2}\right) \setminus \{\mathbf{0}\}$ (which is, for example, known to be the case for GARCH(p,q) processes if we look at $(\tilde{X}_t, \tilde{Y}_t)_{t\in\mathbb{Z}}$ with $\tilde{X}_t := X_t^2, \tilde{Y}_t := \sigma_t^2$, cf. [1], p. 191, see also Chapter 4). Comparing the definition of multivariate regular variation and Condition 3.3 we note that we can of course write the latter as

$$\begin{aligned} & \lim_{x\to\infty} P\left(\left(\frac{Y_{-s_--1}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x}\right) \in A \,\middle|\, |X_0| > x\right) \\ = & \lim_{x\to\infty} \frac{P\left(\left(\frac{Y_{-s_--1}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x}\right) \in A, |X_0| > x\right)}{P(|X_0| > x)} \\ = & P\left((Y^X_{-s_--1}, \ldots, Y^X_{s_+}) \in A\right) \end{aligned} \tag{3.4.2}$$

for a random vector $(Y^X_{-s_--1}, \ldots, Y^X_{s_+})$ and all $A \in \mathbb{R}^{s_-+s_++2}$ with $P((Y^X_{-s_--1}, \ldots, Y^X_{s_+}) \in \partial A) = 0$. By continuity from below it suffices to look at those A which are bounded away from $\mathbf{0}$ in order to derive Condition 3.3 from (3.4.2). The assumption of multivariate regular variation of $(|X_0|, Y_{-s_--1}, \ldots, Y_{s_+})$ now guarantees the existence of a function $U : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\begin{aligned} & \lim_{x\to\infty} \frac{P\left(\left(\frac{Y_{-s_--1}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x}\right) \in A, |X_0| > x\right)}{P(|X_0| > x)} \\ = & \lim_{x\to\infty} \frac{P\left(\left(\frac{Y_{-s_--1}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x}\right) \in A, |X_0| > x\right)}{U(x)} \frac{U(x)}{P(|X_0| > x)} \end{aligned}$$

$$= \frac{\nu((1,\infty) \times A)}{\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2})} \tag{3.4.3}$$

if the denominator is positive. One easily checks that (3.4.3) defines a probability measure for $A \in \mathbb{B}^{s_- + s_+ + 2}$ and may be set as the law of the random vector $(Y^X_{-s_- - 1}, \ldots, Y^X_{s_+})$ if $\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) > 0$. Because of the aforementioned homogeneity of ν we note the equivalence

$$\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0 \Leftrightarrow \nu((\delta,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0 \ \ \forall\, \delta > 0. \tag{3.4.4}$$

Thus, $\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0$ implies that the mass of ν is concentrated on the hyperplane $\{0\} \times \mathbb{R}^{s_- + s_+ + 2}$. Unfortunately, this is not excluded by the definition of regular variation. Nevertheless, since we know that ν is non-degenerate and since the process $(Y_t)_{t \in \mathbb{Z}}$ is stationary, we may derive that

$$\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0 \Rightarrow \lim_{x \to \infty} \frac{P(|Y_0| > x)}{U(x)} > 0.$$

Now, $\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0$ implies that

$$\lim_{x \to \infty} \frac{P(|X_0| > x)}{U(x)} = 0.$$

Hence, $\nu((1,\infty) \times \mathbb{R}^{s_- + s_+ + 2}) = 0$ entails that $|X_0|$ and $|Y_0|$ are not tail equivalent, thus it is a contradiction to Condition 3.1 b). Taken the other way round, this leads us to the following proposition.

Proposition 3.4.1 *Let $(|X_0|, Y_{-s_- - 1}, \ldots, Y_{s_+}) \in \mathbb{R}_+ \times \mathbb{R}^{s_- + s_+ + 2}$ be a multivariate regularly varying vector with index κ and let Condition 3.1 b) hold, then Condition 3.3 is satisfied.*

Next, we will analyze the opposite direction, which means the question if Condition 3.3 implies multivariate regular variation. In a first step, let us study the limit behavior of

$$\mathcal{L}\left(\frac{|X_0|}{x}, \frac{Y_{-s_- - 1}}{x}, \ldots, \frac{Y_{s_+}}{x} \,\middle|\, |X_0| > x \right) \tag{3.4.5}$$

as $x \to \infty$, extending the view to the joint limit behavior of $|X_0|$ and $(Y_{-s_- - 1}, \ldots, Y_{s_+})$. Let us assume that Condition 3.3 holds and in addition

$$\lim_{t \to \infty} \frac{P(|X_0| > xt)}{P(|X_0| > t)} = x^{-\kappa} \tag{3.4.6}$$

for all $x > 0$ (which follows of course if we would assume Condition 3.1 a) and b)). We will show that the family of probability distributions in (3.4.5) is tight. Let therefore

$\epsilon > 0$ and let $B \subset \mathbb{R}^{s_-+s_++2}$ be a compact set such that $P((Y^X_{-s_--1}, \dots, Y^X_{s_+}) \in B) \geq 1 - \epsilon/3$. Then

$$\begin{aligned} & P\left(\frac{|X_0|}{x} \in [1, (\epsilon/3)^{-1/\kappa}], \left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \in B \,\middle|\, |X_0| > x\right) \\ \geq\ & 1 - P\left(\frac{|X_0|}{x} > (\epsilon/3)^{-1/\kappa} \,\middle|\, |X_0| > x\right) - P\left(\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \in B^c \,\middle|\, |X_0| > x\right) \\ \geq\ & 1 - \epsilon/3 - \epsilon/3 - \epsilon/3 = 1 - \epsilon \end{aligned}$$

for x large enough. The family of multivariate distributions in (3.4.5) is therefore tight and there exists a subsequence $x_n, n \in \mathbb{N}$, along which convergence to a probability measure holds (cf. [5], Theorem 25.10). Let now $u \geq 1$ and $B \in \mathbb{B}^{s_-+s_++2}$ with $P((Y^X_{-s_--1}, \dots, Y^X_{s_+}) \in \partial(u^{-1}B)) = 0$, then

$$\begin{aligned} & \lim_{x\to\infty} P\left(\left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \in (u,\infty) \times B \,\middle|\, |X_0| > x\right) \\ =\ & \lim_{x\to\infty} \frac{P\left(\left(\frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \in B, |X_0| > ux\right)}{P(|X_0| > ux)} \frac{P(|X_0| > ux)}{P(|X_0| > x)} \\ =\ & \lim_{x\to\infty} P\left(\left(\frac{Y_{-s_--1}}{ux}, \dots, \frac{Y_{s_+}}{ux}\right) \in u^{-1}B \,\middle|\, |X_0| > ux\right) \frac{P(|X_0| > ux)}{P(|X_0| > x)} \\ =\ & u^{-\kappa} P\left((Y^X_{-s_--1}, \dots, Y^X_{s_+}) \in u^{-1}B\right). \end{aligned} \tag{3.4.7}$$

By the tightness of distributions, the r.h.s. of the last equation now defines a probability distribution on $[1,\infty) \times \mathbb{R}^{s_-+s_++2}$ and the convergence holds independent of the chosen subsequence. Define a random variable $|X_0|^X$ living on the same probability space as $\left(Y^X_{-s_--1}, \dots, Y^X_{s_+}\right)$ such that

$$P\left(|X_0|^X > u, \left(Y^X_{-s_--1}, \dots, Y^X_{s_+}\right) \in B\right) = u^{-\kappa} P\left(\left(Y^X_{-s_--1}, \dots, Y^X_{s_+}\right) \in u^{-1}B\right)$$

holds for all $u \geq 1$ and $B \in \mathbb{B}^{s_-+s_++2}$. Then

$$\mathcal{L}\left(\left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \,\middle|\, |X_0| > x\right) \to \mathcal{L}\left(\left(|X_0|^X, Y^X_{-s_--1}, \dots, Y^X_{s_+}\right)\right), \quad x \to \infty.$$

Having thus shown that the limit of the conditional distributions in (3.4.5) exists, we will next derive a useful characteristic of it. Therefore, apply the continuous mapping theorem to the function

$$f : \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2} \to \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2}, f(x_0, y_{-s_--1}, \dots, y_{s_+}) = \left(x_0, \frac{y_{-s_--1}}{x_0}, \dots, \frac{y_{s_+}}{x_0}\right),$$

which leads to

$$\lim_{x\to\infty} P\left(\left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{|X_0|}, \dots, \frac{Y_{s_+}}{|X_0|}\right) \in (u,\infty) \times B \,\middle|\, |X_0| > x\right)$$

$$= \lim_{x\to\infty} P\left(\left(\frac{Y_{-s_--1}}{|X_0|},\dots,\frac{Y_{s_+}}{|X_0|}\right)\in B \middle| |X_0|>ux\right)\frac{P(|X_0|>xu)}{P(|X_0|>x)}$$

$$\overset{\text{cont. mapping}}{=} u^{-\kappa}P\left(\left(\frac{Y^X_{-s_--1}}{|X_0|^X},\dots,\frac{Y^X_{s_+}}{|X_0|^X}\right)\in B\right)$$

for $u\geq 1$ and $B\in\mathbb{B}^{-s_-+s_++2}$ with $P\left(\left(\frac{Y^X_{-s_--1}}{|X_0|^X},\dots,\frac{Y^X_{s_+}}{|X_0|^X}\right)\in\partial B\right)=0$, thus leading to the following lemma.

Lemma 3.4.2 *Let $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t)_{t\in\mathbb{Z}}$ be defined as in (3.2.1)-(3.2.2) and (3.3.3) and let Condition 3.3 and Equation (3.4.6) hold. Then there exists a random variable $|X_0|^X$, living on the same probability space as $(Y^X_{-s_--1},\dots,Y^X_{s_+})$ and having marginal distribution $P(|X_0|^X>u)=u^{-\kappa}$ for $u\geq 1$, such that*

$$\lim_{x\to\infty}\mathcal{L}\left(\left(\frac{|X_0|}{x},\frac{Y_{-s_--1}}{|X_0|},\dots,\frac{Y_{s_+}}{|X_0|}\right)\middle| |X_0|>x\right)$$
$$= \mathcal{L}\left(|X_0|^X\right)\times\mathcal{L}\left(\left(\frac{Y^X_{-s_--1}}{|X_0|^X},\dots,\frac{Y^X_{s_+}}{|X_0|^X}\right)\right),$$

where "$\times$" denotes the product measure.

Next, we will use this lemma to derive that $(|X_0|,Y_{-s_--1},\dots,Y_{s_+})$ is regularly varying on the cone $\bar{\mathbb{R}}_+\times\bar{\mathbb{R}}^{s_-+s_++2}$ if it satisfies the assumptions of Lemma 3.4.2, where the method of our proof is based on the one used in [25] for the proof of Lemma 6.1 there. Let us define sets

$$V_{u,B}:=\{(x,y_{-s_--1},\dots,y_{s_+})\in\mathbb{R}_+\times\mathbb{R}^{s_-+s_++2}: x>u, x^{-1}(y_{-s_--1},\dots,y_{s_+})\in B\}$$

for $u>0, B\in\mathbb{B}^{s_-+s_++2}$. Set $\mathcal{A}:=\{V_{u,B}\setminus V_{v,B}: 0<u\leq v, B\in\mathbb{B}^{s_-+s_++2}\}$, then $\mathcal{A}$ is a semiring. Define set-functions $\mu,\mu_x, x\in\mathbb{R}_+$, on $\mathcal{A}$ by

$$\mu_x(V_{u,B}\setminus V_{v,B}):=\frac{P\left(\left(\frac{|X_0|}{x},\frac{Y_{-s_--1}}{x},\dots,\frac{Y_{s_+}}{x}\right)\in V_{u,B}\right)}{P(|X_0|>x)}-\frac{P\left(\left(\frac{|X_0|}{x},\frac{Y_{-s_--1}}{x},\dots,\frac{Y_{s_+}}{x}\right)\in V_{v,B}\right)}{P(|X_0|>x)}$$

and

$$\mu(V_{u,B}\setminus V_{v,B}) := u^{-\kappa}P\left(\left(\frac{Y^X_{-s_--1}}{|X_0|^X},\dots,\frac{Y^X_{s_+}}{|X_0|^X}\right)\in B\right)$$
$$-v^{-\kappa}P\left(\left(\frac{Y^X_{-s_--1}}{|X_0|^X},\dots,\frac{Y^X_{s_+}}{|X_0|^X}\right)\in B\right). \qquad (3.4.8)$$

Now, by Carathéodory's Extension Theorem (see, for example, [27], Theorem 2.5) all of these measures can be extended to the generated sigma-algebra $\mathbb{B}_+\otimes\mathbb{B}^{s_-+s_++2}$. The extensions are unique, since all measures are σ-finite on $\mathbb{R}_+\times\mathbb{R}^{s_-+s_++2}=\bigcup_{k=1}^{\infty}V_{1/k,\mathbb{R}^{s_-+s_++2}}$.

We extend μ and $\mu_x, x \in \mathbb{R}_+$, to $\bar{\mathbb{B}}_+ \otimes \bar{\mathbb{B}}^{s_-+s_++2}$ by setting

$$\mu_x(\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2} \setminus \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2}) = \mu(\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2} \setminus \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2}) = 0.$$

Finally, to show vague convergence on the μ-continuity sets of $\mathcal{A}$ we note first that $\mu(\{u\} \times B) = 0$ and therefore $\mu(\partial V_{u,B}) = \mu(V_{u,\partial B})$ for all $u > 0, B \in \mathbb{B}^{s_-+s_++2}$. Therefore, for each set $V_{u,B} \setminus V_{v,B} \in \mathcal{A}$ with $\mu(\partial V_{u,B}) = 0$ we get from Lemma 3.4.2 that

$$\mu_x(V_{u,B} \setminus V_{v,B}) = \mu_x(V_{u,B}) - \mu_x(V_{v,B}) \to \mu(V_{u,B}) - \mu(V_{v,B}) = \mu(V_{u,B} \setminus V_{v,B})$$

as $x \to \infty$. Since $\mu(\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2} \setminus \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2}) = 0$, this implies vague convergence of

$$\frac{P\left(\left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \in \cdot\right)}{P(|X_0| > x)}$$

to the Radon measure μ on $\bar{\mathbb{B}}_+ \otimes \bar{\mathbb{B}}^{s_-+s_++2}$ and therefore multivariate regular variation of $(|X_0|, Y_{-s_--1}, \dots, Y_{s_+})$ on the cone $\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2}$.

Lemma 3.4.3 *Let $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t)_{t\in\mathbb{Z}}$ be defined as in (3.2.1)-(3.2.2) and (3.3.3) and let Condition 3.3 and Equation (3.4.6) hold. Then $(|X_0|, Y_{-s_--1}, \dots, Y_{s_+})$ is multivariate regularly varying with index κ on the cone $\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2}$.*

However, this alone does not necessarily lead to regular variation on $\bar{\mathbb{R}}^{s_-+s_++3} \setminus \{\mathbf{0}\}$. One could think of situations where the conditioning on $|X_0| > zx$ for a $z > 0$ as $x \to \infty$ may guarantee a certain tail-behavior of the remaining components of $(Y_{-s_--1}, \dots, Y_{s_+})$ which gets lost if we drop this condition. This may be a special case of so-called hidden regular variation, cf. [36], Chapter 9.4. However, the picture changes if we add more assumptions about the stationary sequences $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t)_{t\in\mathbb{Z}}$. If we assume that Conditions 3.1 a)-b) and 3.2 a) hold in addition to Condition 3.3 we may derive the following. Let $x_0 > 0, -\infty \le y_i^- < 0, 0 < y_i^+ \le \infty, i = -s_- - 1, \dots, s_+$ and set

$$\mathbf{x}^- = (0, y^-_{-s_--1}, \dots, y^-_{s_+}), \quad \mathbf{x}^+ = (x_0, y^+_{-s_--1}, \dots, y^+_{s_+}).$$

To verify that the multivariate regular variation can be extended to $\bar{\mathbb{R}}^{s_-+s_++3} \setminus \{\mathbf{0}\}$ we have to show that

$$\lim_{x\to\infty} \frac{P\left(\left[\mathbf{x}^- \le \left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \le \mathbf{x}^+\right]^c\right)}{P(|X_0| > x)} \xrightarrow{v} \mu_1([\mathbf{x}^-, \mathbf{x}^+]^c) \tag{3.4.9}$$

holds for all such choices of $\mathbf{x}^-, \mathbf{x}^+$ and a Radon measure μ_1 on $\bar{\mathbb{B}}^{s_-+s_++3} \setminus \{\mathbf{0}\}$ with $\mu_1\big|_{\bar{\mathbb{B}}_+ \times \bar{\mathbb{B}}^{s_-+s_++2}} = \mu$. By inclusion-exclusion we get that

$$\frac{P\left(\left[\mathbf{x}^- \le \left(\frac{|X_0|}{x}, \frac{Y_{-s_--1}}{x}, \dots, \frac{Y_{s_+}}{x}\right) \le \mathbf{x}^+\right]^c\right)}{P(|X_0| > x)}$$

$$= \frac{P\left(\{|X_0| > xx_0\} \cup \bigcup_{i=-s_--1}^{s_+} \left(\{Y_i < xy_i^-\} \cup \{Y_i > xy_i^+\}\right)\right)}{P(|X_0| > x)}$$

$$= \frac{P(|X_0| > xx_0)}{P(|X_0| > x)} + \frac{P\left(\bigcup_{i=-s_--1}^{s_+} \left(\{Y_i < xy_i^-\} \cup \{Y_i > xy_i^+\}\right)\right)}{P(|X_0| > x)} -$$

$$\frac{P\left(\{|X_0| > xx_0\} \cap \bigcup_{i=-s_--1}^{s_+} \left(\{Y_i < xy_i^-\} \cup \{Y_i > xy_i^+\}\right)\right)}{P(|X_0| > x)}.$$

Here, the first and the third term in the sum converge because of Conditions 3.1 a), 3.1 b) and Condition 3.3. For the second term we may use that Conditions 3.1 a) and 3.2 a) induce multivariate regular variation of the vector $(Y_{-s_--1}, \dots, Y_{s_+})$ on $\bar{\mathbb{R}}^{s_-+s_++1} \setminus \{\mathbf{0}\}$ (cf. [39], Theorem 6.1). Then by Condition 3.1 b) it must be the case that $P(|X_0| > x)$ is a proper scaling function for $P(x^{-1}(Y_{-s_--1}, \dots, Y_{s_+}) \in \cdot)$. Therefore, the second term also converges.

Proposition 3.4.4 *Let $(Y_t, X_t)_{t\in\mathbb{Z}}$ be a stationary sequence given by (3.2.1)-(3.2.2) and (3.3.3) and let Conditions 3.1 a)+b), 3.2 a)+b) and 3.3 be satisfied. Then, the vector $(|X_0|, Y_{-s_--1}, \dots, Y_{s_+})$ is multivariate regularly varying with index κ on $\left(\bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^{s_-+s_++2}\right) \setminus \{\mathbf{0}\}$.*

Of course, the limit measure on $\left(\bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^{s_-+s_++2}\right) \setminus \{\mathbf{0}\}$ can trivially be continued on $\bar{\mathbb{R}}^{s_-+s_++3} \setminus \{\mathbf{0}\}$ since the first component of the vector is necessarily non-negative.

Although Conditions 3.1-3.3 seem to involve quite a lot of assumptions, it is often easy to prove that Conditions 3.1 and 3.3 are satisfied, since they can be derived from well-known results about the stationary behavior of time series, which often allow us to draw conclusions about univariate and multivariate regular variation of the time series $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t)_{t\in\mathbb{Z}}$. These results are then helpful to show the existence of the limit distribution in Condition 3.3, although explicit calculations for further analysis might get tedious. Condition 3.2 is in most cases easy to check.

Because of the aforementioned homogeneity the measure μ in (3.4.8) (or its extension defined in (3.4.9)) has a spectral form and can be written as

$$\mu\left(\{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\| > r, \|\mathbf{x}\|^{-1}\mathbf{x} \in A\}\right) = r^{-\kappa} S(A), \tag{3.4.10}$$

(with $\mathbb{C} = \bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}^{s_-+s_++2}$ or $\mathbb{C} = \bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^{s_-+s_++2} \setminus \{\mathbf{0}\}$) for all $r > 0$, where S is a Radon measure on $\mathbb{S}_\mathbb{C} := \{\mathbf{x} \in \mathbb{C} : \|\mathbf{x}\| = 1\}$, $A \subset \mathbb{S}_\mathbb{C}$ (cf. [30], p. 36). For $\mathbb{C} = \bar{\mathbb{R}}_{+,0} \times \bar{\mathbb{R}}^{s_-+s_++2} \setminus \{\mathbf{0}\}$ we know in addition that S is finite (cf. [37], p. 281). Here, $\|\cdot\|$ denotes an arbitrary norm, for example the l_2-norm. Lemma 3.4.2 shows that in our case this decomposition also holds for the seminorm

$$f : \mathbb{R}_+ \times \mathbb{R}^{s_-+s_++2} \mapsto \mathbb{R}_+, \quad f(x_0, y_{-s_--1}, \dots, y_{s_+}) := x_0.$$

This can be interpreted as follows: The limit process can be split up in two parts, which are independent of each other. The first one determines the scale of the limit

process, while the second one determines its "direction". The use of this decomposition for both simulations and calculations is highly dependent on the form of the spectral measure. In some cases, the spectral measure and therefore the limit process admits a very simple representation, like in the case of the existence of a forward and backward tail chain for a single process.

In the next section, we will show that outside of the range of direct dependence (meaning to the left of $-s_- - 1$ and to the right of s_+) the results of Propositions 3.2.1 and 3.2.2 can be carried over to the case where we condition on $|X_0| > x$ instead of $|Y_0| > x$.

3.5 Main result

Before we state and prove our main result, we should again point out that an important difference between the single time series model analyzed in [39] and the correlated time series analyzed here is the double role of the ϵ_i which influence both the progress of $(Y_t)_{t\in\mathbb{Z}}$ and the relation of $(Y_t)_{t\in\mathbb{Z}}$ and $(X_t)_{t\in\mathbb{Z}}$. Note that in Proposition 3.2.1 the distribution of ϵ_t does not directly appear and only the distribution of $\phi(\epsilon_t)$ is needed. Therefore, as a first step, we modify Proposition 3.2.1.

Lemma 3.5.1 *Let $(Y_t)_{t\in\mathbb{Z}}$ be given by (3.2.1) and (3.2.2) and let the conditions of Proposition 3.2.2 be satisfied. Then for all $s, t \in \mathbb{N}$*

$$\begin{aligned} &\lim_{y\to\infty} \mathcal{L}\left(\frac{Y_{-s}}{y}, \dots, \frac{Y_{-1}}{y}, \frac{|Y_0|}{y}, \frac{Y_0}{y}, \epsilon_1, \frac{Y_1}{y}, \dots, \epsilon_t, \frac{Y_t}{y} \,\middle|\, |Y_0| > y\right) \\ = \;& \mathcal{L}(Y^Y_{-s}, \dots, Y^Y_{-1}, Y, Y^Y_0, \epsilon^Y_1, Y^Y_1, \epsilon^Y_2, Y^Y_2, \dots, \epsilon^Y_t, Y^Y_t), \end{aligned}$$

with

- *$Y \sim Par(\kappa), \epsilon^Y_i \overset{d}{=} \epsilon_i$, M_0 with $P(M_0 = 1) = p = 1 - P(M_0 = -1)$ all independent and $Y^Y_0 = Y M_0$,*
- *$Y^Y_i = h(Y^Y_{i-1}, A_i, B_i)$, with $A_i = \phi(\epsilon^Y_i, 1), B_i = \phi(\epsilon^Y_i, -1), i \geq 1$,*
- *$Y^Y_{-i} = h(Y^Y_{-i+1}, A_{-i}, B_{-i}), i \geq 1$, with*

$$\mathcal{L}(A_{-i}) = \mathcal{L}\left(\frac{M_{-1}}{M_0} \,\middle|\, M_0 = 1\right), \quad \mathcal{L}(B_{-i}) = \mathcal{L}\left(\frac{M_{-1}}{M_0} \,\middle|\, M_0 = -1\right), i \geq i,$$

where $\mathcal{L}(M_{-1}, M_0) = \mu^*$, the adjoint measure to $\mathcal{L}(M_0, M_1)$. The $A_{-i}, B_{-i}, i \geq 1$, are independent and independent of $(Y, Y^Y_0, \epsilon^Y_1, Y^Y_1, \dots)$.

Proof. This follows directly from Proposition 3.2.2 if $t = 0$. For $t > 0$ we note that the $\epsilon_i, i \geq 1$, are independent of $(Y_{-s}, \dots, Y_{-1}, Y_0)$ and there exist suitable functions $g_y : \mathbb{R}^{t+1} \to \mathbb{R}^{2t}, y > 0$, such that

$$\left(\epsilon_1, \frac{Y_1}{y}, \epsilon_2, \frac{Y_2}{y}, \dots, \epsilon_t, \frac{Y_t}{y}\right) = g_y\left(\frac{Y_0}{y}, \epsilon_1, \dots, \epsilon_t\right).$$

Analogously to the proof of Proposition 3.2.1 in [39] we may then apply the continuous mapping theorem which leads to the result. □

We are now ready to prove our main result.

Theorem 3.5.2 *Let $(Y_t, X_t)_{t\in\mathbb{Z}}$ be given by (3.2.1)-(3.2.2) and (3.3.3), stationary and let Conditions 3.1 a)+b), 3.2 a)+b) and 3.3 be satisfied. Then, for all integers $m > 0, n \geq 0$*

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_--m}}{x}, \dots, \frac{Y_{s_++n}}{x} \middle| |X_0| > x\right) = \mathcal{L}(Y^X_{-s_--m}, \dots, Y^X_{s_++n})$$

with $(Y^X_{-s_--1}, \dots, Y^X_{s_+})$ from Condition 3.3 and

$$\begin{aligned} Y^X_t &= h(Y^X_{t-1}, A_t, B_t), && t > s_+, \\ Y^X_{-t} &= h(Y^X_{-t+1}, A_{-t}, B_{-t}), && t > s_- + 1, \end{aligned}$$

where $(A_t, B_t), t \in \mathbb{Z}$, are independent and independent of $(Y^X_{-s_--1}, \dots, Y^X_{s_+})$ with

$$\mathcal{L}(A_t, B_t) = \mathcal{L}(A_1, B_1),\ t \geq 1, \quad \mathcal{L}(A_t, B_t) = \mathcal{L}(A_{-1}, B_{-1}),\ t \leq -1,$$

with $\mathcal{L}(A_1, B_1)$ and $\mathcal{L}(A_{-1}, B_{-1})$ defined as in Proposition 3.2.2.

This theorem indicates that outside of the range of influence of X_0 the tail chain behaves "normally" (i.e., like the one that is conditioned on $\{|Y_0| > x\}$).

Proof. Let $f : \mathbb{R}^{s_-+s_++m+n+1} \to \mathbb{R}$ be bounded and continuous. We will show that

$$\begin{aligned} &\lim_{x\to\infty} E\left(f\left(\frac{Y_{-s_--m}}{x}, \dots, \frac{Y_{s_++n}}{x}\right) \middle| |X_0| > x\right) \\ = \ & E\left(f(Y^X_{-s_--m}, \dots, Y^X_{s_++n})\right) \end{aligned} \tag{3.5.1}$$

with $(Y^X_{-s_--m}, \dots, Y^X_{s_++n})$ as defined in the theorem for all f.

The idea of the proof is to shift the time series, using its stationarity, thereby clarifying the dependence structure between the $\epsilon_i, i \in \mathbb{Z}$, and both $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$.

Note that for $m = 1$ we may conduct the proof analogously to the proof of Proposition 3.2.1 in [39]. Since $(\epsilon_{s_++1}, \epsilon_{s_++2}, \dots)$ is independent of $(X_0, Y_{-s_--1}, \dots, Y_{s_+})$ the continuous mapping theorem can be applied to derive (3.5.1) and leads to the multiplicative structure with independent increments.

Let now $m > 1$ and $n \geq 0$. Let us further assume that $f(x) = 0$ as soon as the first component of x equals 0. Note that an arbitrary function $f : \mathbb{R}^{s_-+s_++m+n+1} \to \mathbb{R}$ can be additively split up in two functions f_1 and f_2 such that

$$f(x_{-s_--m}, \dots, x_{s_++n}) = \underbrace{f(x_{-s_--m}, \dots, x_{s_++n}) - f(0, x_{-s_--m+1}, \dots, x_{s_++n})}_{=: f_1(x_{-s_--m}, \dots, x_{s_-+n})}$$

$$+ \underbrace{f(0, x_{-s_- -m+1}, \ldots, x_{s_+ +n})}_{=: f_2(x_{-s_- -m+1}, \ldots, x_{s_- +n}))},$$

so f_1 satisfies this assumption and f_2 is merely a function of $(x_{-s_- -m+1}, \ldots, x_{s_+ +n})$. Since (3.5.1) is satisfied by a function of $(x_{-s_- -1}, \ldots, x_{s_+ +n})$ we may work inductively and the assumption about the structure of f is no loss of generality.

Then

$$\begin{aligned} &\lim_{x\to\infty} E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \middle| |X_0| > x \right) \\ = \; & o_\delta(1) + \lim_{x\to\infty} E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \mathbb{1}_{\{|Y_{-s_- -m}| > \delta x\}} \middle| |X_0| > x \right), \end{aligned}$$

where the first term is bounded by

$$\sup\{f(x_{-s_- -m}, \ldots, x_{s_+ +n}) | (x_{-s_- -m+1}, \ldots, x_{s_+ +n}) \in \mathbb{R}^{s_- + s_+ + n + m}, x_{-s_- -m} < \delta\}$$

which tends to zero as $\delta \to 0$ because of the assumptions made about f. With Condition 3.1 b) and with stationarity we get

$$\begin{aligned} & \lim_{x\to\infty} E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \mathbb{1}_{\{|Y_{-s_- -m}| > \delta x\}} \middle| |X_0| > x \right) \\ = \; & \lim_{x\to\infty} \frac{E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \mathbb{1}_{\{|X_0| > x\}} \mathbb{1}_{\{|Y_{-s_- -m}| > \delta x\}} \right)}{P(|X_0| > x)} \\ \overset{3.1\,b)}{=} \; & \frac{\delta^{-\kappa}}{C} \lim_{x\to\infty} \frac{E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \mathbb{1}_{\{|X_0| > x\}} \mathbb{1}_{\{|Y_{-s_- -m}| > \delta x\}} \right)}{P(|Y_{-s_- -m}| > \delta x)} \\ = \; & \frac{\delta^{-\kappa}}{C} \lim_{x\to\infty} E\left(f\left(\frac{Y_{-s_- -m}}{x}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+ +n}}{x} \right) \mathbb{1}_{\{|X_0| > x\}} \middle| |Y_{-s_- -m}| > \delta x \right) \\ \overset{\text{Stationarity}}{=} \; & \frac{\delta^{-\kappa}}{C} \lim_{x\to\infty} E\left(f\left(\frac{Y_0}{x}, \ldots, \frac{Y_{s_- +m}}{x}, \ldots, \frac{Y_{s_+ +n+s_- +m}}{x} \right) \right. \\ & \qquad \left. \mathbb{1}_{\{|X_{s_- +m}| > x\}} \middle| |Y_0| > \delta x \right) \\ \overset{y := \delta x}{=} \; & \frac{\delta^{-\kappa}}{C} \lim_{y\to\infty} E\left(f\left(\delta\frac{Y_0}{y}, \ldots, \delta\frac{Y_{s_- +m}}{y}, \ldots, \delta\frac{Y_{s_+ +n+s_- +m}}{y} \right) \right. \\ & \qquad \left. \mathbb{1}_{\{\delta |\Psi(Y_{s_- +m}, \epsilon_m, \ldots, \epsilon_{s_- + s_+ + m})| > y\}} \middle| |Y_0| > y \right). \end{aligned}$$

An application of the continuous mapping theorem in connection with Lemma 3.5.1 yields that this expression equals

$$\frac{\delta^{-\kappa}}{C} E\Big(f\left(\delta Y_0^Y, \ldots, \delta Y_{s_- +m}^Y, \ldots, \delta Y_{s_+ +n+s_- +m}^Y \right)$$

$$\mathbb{1}_{\{\delta|Y^Y_{s_-+m}\psi(\epsilon^Y_m,\dots,\epsilon^Y_{s_-+s_++m})|>1\}}\Bigg)$$

$$= \frac{\delta^{-\kappa}}{C}E\Bigg(f\left(\delta Y M_0,\dots,\delta Y M_{m-1}\prod_{i=m}^{s_-+m}\frac{h(M_{i-1},\phi(\epsilon_i^Y,1),\phi(\epsilon_i^Y,-1))}{M_{i-1}},\dots,\right.$$
$$\left.\delta Y M_{m-1}\prod_{i=m}^{s_++n+s_-+m}\frac{h(M_{i-1},\phi(\epsilon_i^Y,1),\phi(\epsilon_i^Y,-1))}{M_{i-1}}\right)$$
$$\mathbb{1}_{\{\delta|YM_{m-1}\prod_{i=m}^{s_-+m}\frac{h(M_{i-1},\phi(\epsilon_i^Y,1),\phi(\epsilon_i^Y,-1))}{M_{i-1}}\psi(\epsilon^Y_m,\dots,\epsilon^Y_{s_-+s_++m})|>1\}}\Bigg),$$

with $Y,\epsilon_1^Y,\epsilon_2^Y,\dots$ as in Lemma 3.5.1 and $P(M_0=1)=p=1-P(M_0=-1)$,

$$M_i = h(M_{i-1},\phi(\epsilon_i^Y,1),\phi(\epsilon_i^Y,-1)),\quad i\in\mathbb{N}.$$

Since Y is independent of all other variables with $Y\overset{d}{=}U^{-1/\kappa}$ for $U\sim\text{Unif}[0,1]$ and the $\epsilon_m^Y,\epsilon_{m+1}^Y,\dots$ are independent of $M_0,M_1,\dots,M_{m-1}$ this equals

$$\frac{\delta^{-\kappa}}{C}\int_{\mathbb{R}^{s_-+s_++n+1}}E\Bigg(\int_0^1 f\left(\delta u^{-1/\kappa}M_0,\dots,\delta u^{-1/\kappa}M_{m-1}h_1\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right),\right.$$
$$\left.\dots,\delta u^{-1/\kappa}M_{m-1}h_2\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)\right)$$
$$\mathbb{1}_{\{\delta^\kappa|M_{m-1}^\kappa h_3\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)^\kappa|>u\}}\,\mathrm{d}u\Bigg)P^{(\epsilon^Y_m,\dots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e})$$

for suitable, measurable functions h_1,h_2,h_3. Now, substituting v for $\delta^{-\kappa}|M_{m-1}|^\kappa u$, we get that this expression equals

$$\frac{1}{C}\int_{\mathbb{R}^{s_-+s_++n+1}}E\Bigg(\int_0^{\delta^{-\kappa}|M_{m-1}|^\kappa} f\left(v^{-1/\kappa}\frac{M_0}{|M_{m-1}|},\dots,v^{-1/\kappa}\frac{M_{m-1}}{|M_{m-1}|}h_1\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right),\right.$$
$$\left.\dots,v^{-1/\kappa}\frac{M_{m-1}}{|M_{m-1}|}h_2\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)\right)|M_{m-1}|^\kappa$$
$$\mathbb{1}_{\{|h_3\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)^\kappa|>v\}}\,\mathrm{d}v\Bigg)P^{(\epsilon^Y_m,\dots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e}).$$

Letting δ tend to zero eliminates $|M_{m-1}|$ in the bounds of the integral and interchange of expectation and integral leads to

$$\frac{1}{C}\int_{\mathbb{R}^{s_-+s_++n+1}}\int_0^\infty E\Bigg(f\left(v^{-1/\kappa}\frac{M_0}{|M_{m-1}|},\dots,v^{-1/\kappa}\frac{M_{m-1}}{|M_{m-1}|}h_1\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right),\right.$$

$$\ldots, v^{-1/\kappa}\frac{M_{m-1}}{|M_{m-1}|}h_2\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)\Big)\,|M_{m-1}|^\kappa$$
$$\mathbb{1}_{\{|h_3\left(\frac{M_{m-1}}{|M_{m-1}|},\mathbf{e}\right)^\kappa|>v\}}\Bigg)\,\mathrm{d}v\,P^{(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e}).$$

We are now able to apply Proposition 4.2 of [39] (stated before as Proposition 3.2.4) with $s = m-1, t = 0$ and $i = m-1$ and $i = 0$ respectively to transform the integrand of this expression. It then equals

$$\frac{1}{C}\int_{\mathbb{R}^{s_-+s_++n+1}}\int_0^\infty E\Bigg(f\left(v^{-1/\kappa}\frac{M_{-m+1}}{|M_0|},\ldots,v^{-1/\kappa}\frac{M_0}{|M_0|}h_1\left(\frac{M_0}{|M_0|},\mathbf{e}\right),\right.$$
$$\left.\ldots,v^{-1/\kappa}\frac{M_0}{|M_0|}h_2\left(\frac{M_0}{|M_0|},\mathbf{e}\right)\right)\mathbb{1}_{\{|h_3\left(\frac{M_0}{|M_0|},\mathbf{e}\right)^\kappa|>v\}}\Bigg)\,\mathrm{d}v\,P^{(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e}),$$

where $(M_{-m+1},\ldots,M_0)$ is the BFTC as in Proposition 3.2.2. Note that $|M_0| = 1$ and that $(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})$ is independent of $(M_{-m+1},\ldots,M_0)$ (cf. Lemma 3.5.1) which allows us to leave the measure unchanged in the two expressions. After this essential time shift we are doing every step of the proof "backwards", starting with inserting δ again. The above expression then equals the limit for $\delta\to 0$ of

$$\frac{1}{C}\int_{\mathbb{R}^{s_-+s_++n+1}}\int_0^{\delta^{-\kappa}} E\Bigg(f\left(v^{-1/\kappa}M_{-m+1},\ldots,v^{-1/\kappa}M_0h_1\left(M_0,\mathbf{e}\right),\right.$$
$$\left.\ldots,v^{-1/\kappa}M_0h_2\left(M_0,\mathbf{e}\right)\right)\mathbb{1}_{\{|h_3(M_0,\mathbf{e})^\kappa|>v\}}\Bigg)\,\mathrm{d}v\,P^{(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e}).$$

$$\overset{y:=v\delta^\kappa}{=}\quad\frac{\delta^{-\kappa}}{C}\int_{\mathbb{R}^{s_-+s_++1}}\int_0^1 E\Bigg(f\left(\delta y^{-1/\kappa}M_{-m+1},\ldots,\delta y^{-1/\kappa}M_0h_1\left(M_0,\mathbf{e}\right),\right.$$
$$\left.\ldots,\delta y^{-1/\kappa}M_0h_2\left(M_0,\mathbf{e}\right)\right)\mathbb{1}_{\{\delta^\kappa|h_3(M_0,\mathbf{e})^\kappa|>y\}}\Bigg)\,\mathrm{d}y\,P^{(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})}(\mathrm{d}\mathbf{e}).$$

Since $(\epsilon_1,\ldots,\epsilon_{1+s_-+s_++n})$ has the same distribution as $(\epsilon_m,\ldots,\epsilon_{m+s_-+s_++n})$ (and therefore as $(\epsilon^Y_m,\ldots,\epsilon^Y_{m+s_-+s_++n})$) and is independent of $M_0, M_{-1},\ldots$ (cf. Lemma 3.5.1), we may also write

$$=\quad\frac{\delta^{-\kappa}}{C}\int_{\mathbb{R}^{s_-+s_++1}}\int_0^1 E\Bigg(f\left(\delta y^{-1/\kappa}M_{-m+1},\ldots,\delta y^{-1/\kappa}M_0h_1\left(M_0,\mathbf{e}\right),\right.$$
$$\left.\ldots,\delta y^{-1/\kappa}M_0h_2\left(M_0,\mathbf{e}\right)\right)\mathbb{1}_{\{\delta^\kappa|h_3(M_0,\mathbf{e})^\kappa|>y\}}\Bigg)\,\mathrm{d}y\,P^{(\epsilon_1,\ldots,\epsilon_{1+s_-+s_++n})}(\mathrm{d}\mathbf{e})$$

$$
\begin{aligned}
&= \frac{\delta^{-\kappa}}{C} E\left(f\left(\delta Y M_{-m+1}, \ldots, \delta Y M_0 \prod_{i=1}^{s_-+1} \frac{h(M_{i-1}, \phi(\epsilon_i, 1), \phi(\epsilon_i, -1))}{M_{i-1}}, \ldots, \right.\right. \\
&\quad \left.\left. \delta Y M_0 \prod_{i=1}^{s_++n+s_-+1} \frac{h(M_{i-1}, \phi(\epsilon_i, 1), \phi(\epsilon_i, -1))}{M_{i-1}} \right) \right. \\
&\quad \left. \mathbb{1}_{\{\delta |Y M_0 \prod_{i=1}^{s_-+1} \frac{h(M_{i-1}, \phi(\epsilon_i, 1), \phi(\epsilon_i, -1))}{h(M_{i-1}} \psi(\epsilon_1, \ldots, \epsilon_{s_-+s_++1})| > 1\}} \right) \\
&= \lim_{y \to \infty} \frac{\delta^{-\kappa}}{C} E\left(f\left(\delta \frac{Y_0}{y} \tilde{M}_{-m+1}, \ldots, \delta \frac{Y_{s_-+1}}{y}, \ldots, \delta \frac{Y_{s_-+s_++1}}{y} \tilde{M}_{s_-+s_++n} \right) \right. \\
&\quad \left. \mathbb{1}_{\{|\delta X_{s_-+1}| > y\}} \,\middle|\, |Y_0| > y \right)
\end{aligned}
$$

with

$$
\tilde{M}_{-1} = h\left(\operatorname{sign}(Y_0), \tilde{A}_{-1}, \tilde{B}_{-1}\right), \quad \tilde{M}_{-i} = h(\tilde{M}_{-i+1}, \tilde{A}_{-i}, \tilde{B}_{-i}), i \geq 2,
$$

$$
\tilde{M}_{s_-+s_++2} = h\left(\operatorname{sign}(Y_{s_-+s_++1}), \tilde{A}_{s_-+s_++2}, \tilde{B}_{s_-+s_++2}\right),
$$

$$
\tilde{M}_i = h(\tilde{M}_{i-1}, \tilde{A}_i, \tilde{B}_i), i \geq s_- + s_+ + 3,
$$

where $(\tilde{A}_i, \tilde{B}_i)_{i \in \mathbb{Z}}$ are supposed to be independent of $(Y_t, X_t)_{t \in \mathbb{Z}}$ and with the same distribution as $(A_i, B_i)_{i \in \mathbb{Z}}$ in the theorem. Now by stationarity and the assumptions about global independence of $(\tilde{A}_i, \tilde{B}_i)_{i \in \mathbb{Z}}$ this equals

$$
\lim_{y \to \infty} \frac{\delta^{-\kappa}}{C} E\left(f\left(\delta \frac{Y_{-s_--1}}{y} \hat{M}_{-s_--m}, \ldots, \delta \frac{Y_0}{y}, \ldots, \delta \frac{Y_{s_+}}{y} \hat{M}_{s_++n} \right) \mathbb{1}_{\{|\delta X_0| > y\}} \,\middle|\, |Y_{-s_--1}| > y \right)
$$

with

$$
\hat{M}_{-s_--2} = h\left(\operatorname{sign}(Y_{-s_--1}), \tilde{A}_{-s_--2}, \tilde{B}_{-s_--2}\right), \quad \hat{M}_{-i} = h(\hat{M}_{-i+1}, \tilde{A}_{-i}, \tilde{B}_{-i}), i \geq s_- + 3,
$$

$$
\hat{M}_{s_++1} = h\left(\operatorname{sign}(Y_{s_+}), \tilde{A}_{s_++1}, \tilde{B}_{s_++1}\right), \ \hat{M}_i = h(\hat{M}_{i-1}, \tilde{A}_i, \tilde{B}_i), i \geq s_+ + 2.
$$

Finally, this gives us

$$
\begin{aligned}
&= \lim_{x \to \infty} E\left(f\left(\frac{Y_{-s_--1}}{x} \hat{M}_{-s_--m}, \ldots, \frac{Y_0}{x}, \ldots, \frac{Y_{s_+}}{x} \hat{M}_{s_++n} \right) \right. \\
&\quad \left. \mathbb{1}_{\{|Y_{-s_-1}| > \delta x\}} \,\middle|\, |X_0| > x \right) \\
&= o(\delta) + E\left(f\left(Y^X_{-s_--m}, \ldots, Y^X_0, \ldots, Y^X_{s_++n} \right) \right)
\end{aligned}
$$

with $\left(Y^X_{-s_--m}, \ldots, Y^X_0, \ldots, Y^X_{s_++n}\right)$ as stated in the theorem. Now, with $\delta \to 0$ this leads to the proposition. □

3.6 Application for simulating the extremal index

Our analysis of two connected time series was originally motivated by simulation studies to estimate the extremal index and related characteristics of a GARCH$(1,1)$ process (cf. equations (3.3.4) and (3.3.5)) based on the procedure in [14] for ARCH processes. The extremal index is a way to characterize the extremal behavior of a time series, especially its dependencies in the occurence of extremal observations. Many processes have the property that extremal observations come in so-called "clusters". Now, the extremal index is defined as follows.

Definition 3.6.1 (cf. [17], Definition 8.1.2)**.** Let $(X_t)_{t\in\mathbb{N}}$ be a stationary process with marginal distribution function F and θ_X a non-negative number. Set $M_n := \max(X_1,\dots,X_n)$ for $n \in \mathbb{N}$. Assume that for every $\tau > 0$ there exists a sequence $(u_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} n\bar{F}(u_n) = \tau,$$

$$\lim_{n\to\infty} P(M_n \leq u_n) = e^{-\theta_X \tau}.$$

Then θ_X is called the *extremal index* of the sequence $(X_n)_{n\in\mathbb{N}}$.

One can show that $\theta_X \in [0,1]$, where $\theta_X = 1$ for a process of i.i.d. observations. Roughly speaking, the lower the value of θ_X, the more dependence there is in the extremal behavior of the process. A handy interpretation of θ_X can be deduced if we return to the aforementioned clusters of the extremal process, specifying their meaning. One can show that under some mixing conditions on $(X_t)_{t\in\mathbb{N}}$ the family of probability laws

$$\mathcal{L}\left(\sum_{i=1}^{r(n)} \mathbb{1}_{\{X_i>n\}} \,\middle|\, \sum_{i=1}^{r(n)} \mathbb{1}_{\{X_i>n\}} > 0\right) \tag{3.6.1}$$

for a suitable sequence $r(n) \to \infty$ converges as $n \to \infty$ to a probability law which then describes the number of exceedances over a growing threshold, given that we have at least one exceedance over this threshold in a certain part of the time series. If the limit distribution in (3.6.1) is the distribution of a random variable N, then under some mild additional assumptions we get that $\theta_X^{-1} = E(N)$ (cf. [17], p. 421-422).

Furthermore, one can show that θ_X may be written as

$$\theta_X = \lim_{n\to\infty} \lim_{x\to\infty} P\left(\max(X_1,\dots,X_n) < x \mid X_0 > x\right)$$

in the case of a GARCH$(1,1)$ model (cf. [17], p. 422-423, and [15], Section 5.2). A related characteristic is the extremal coefficient function, which is introduced in [18]:

$$\chi(t) = \lim_{x\to\infty} P(X_t > x | X_0 > x), \quad t \in \mathbb{Z}.$$

Under the assumption of a symmetric distribution of the ϵ_t we may easily deduce the extremal behavior of $(X_t)_{t\in\mathbb{Z}}$ from the extremal behavior of $(|X_t|)_{t\in\mathbb{Z}}$. With the help of Theorem 3.5.2 we are now able to find

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{\sigma_{-n}}{x},\dots,\frac{\sigma_0}{x},\dots,\frac{\sigma_m}{x}\,\middle|\, X_0 > x\right), \quad m,n\in\mathbb{N},$$

where the use of the BFTC simplifies both calculations and simulations from this distribution. For simulations, because of the simple form of Ψ, we are able to reconstruct the tail process of $(|X_t|)_{t\in\mathbb{Z}}$ from a given realisation of the tail process of $(\sigma_t^2)_{t\in\mathbb{Z}}$, thus we are able to retrieve realisations from the distribution of

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{|X_{-n}|}{x},\dots,\frac{|X_0|}{x},\dots,\frac{|X_m|}{x}\,\middle|\, |X_0| > x\right), \quad m,n\in\mathbb{N}. \tag{3.6.2}$$

Since for symmetric distributions of ϵ_0 the sign of $\epsilon_t, t\in\mathbb{Z}$, is independent of its absolute value and does neither influence $(\sigma_t)_{t\in\mathbb{Z}}$ nor $(|X_t|)_{t\in\mathbb{Z}}$ we may model the process $(X_t)_{t\in\mathbb{Z}}$ as

$$X_t = |X_t|B_t, t\in\mathbb{Z},$$

where the $B_t, t\in\mathbb{Z}$, are i.i.d. and independent of $(|X_t|)_{t\in\mathbb{Z}}$ and $(\sigma_t)_{t\in\mathbb{Z}}$ with $P(B_0 = 1) = P(B_0 = -1) = 1/2$. Multiplying the realisations from (3.6.2) with the i.i.d. $B_t, t\in\mathbb{Z}$, then allows us to get realisations with distribution

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{X_{-n}}{x},\dots,\frac{X_0}{x},\dots,\frac{X_m}{x}\,\middle|\, |X_0| > x\right), \quad m,n\in\mathbb{N}.$$

Using only the realisations with positive observation at time zero finally leads to simulating from

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{X_{-n}}{x},\dots,\frac{X_0}{x},\dots,\frac{X_m}{x}\,\middle|\, X_0 > x\right), \quad m,n\in\mathbb{N}.$$

These realisations may then be used for Monte Carlo methods to derive the extremal index, the extremal coefficient function or related characteristics for the GARCH$(1,1)$ model. See [15], Chapter 5 (especially Section 5.3) for a detailed analysis of this topic and simulation results. The value of the use of the BFTC becomes especially clear for the estimation of the values of $\chi(t)$ for $t<0$ and a new characteristic derived in [15], Chapter 5.

3.7 Further applications

This last section shall illustrate the use of Theorem 3.5.2 for the analytical derivation of extremal characteristics of two connected time series. For a simple example of two time series $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$, we will analyze the two probabilities

$$\lim_{n\to\infty}\lim_{x\to\infty} P(\max(|Y_1|,\dots,|Y_n|) < x \mid |Y_0| > x) \tag{3.7.1}$$

and

$$\lim_{n\to\infty}\lim_{x\to\infty} P(\max(|Y_1|,\ldots,|Y_n|) < x \mid |X_0| > x). \tag{3.7.2}$$

Since the first limit can be interpreted (under some additional assumptions on the time series $(Y_t)_{t\in\mathbb{Z}}$) as the extremal index of the time series $(|Y_t|)_{t\in\mathbb{Z}}$, the second characteristic can be interpreted as a simple form of a "cross extremal index" for two time series. However, the main goal of this last chapter is to show how the second expression can easily be derived once the right instruments have been found to tackle the first one.

In the following we will deal with the special case of an ARCH(1) model.

To be more precise, let us assume a simple ARCH(1) model of the following form, where we will use the notation σ_t^2 instead of Y_t due to the conventions of financial time series:

$$X_t = \sigma_t\epsilon_{t+1}, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2\epsilon_t^2,$$

for $t \in \mathbb{Z}$ with $\alpha_0 > 0, \alpha_1 > 0$ and let $\epsilon_t^2, t \in \mathbb{Z}$, be i.i.d. uniformly distributed on $(0, \frac{a}{\alpha_1})$, with $1 < a < e$. This simple distribution assumption for ϵ_t^2 ensures (because of $a < e$) that a stationary solution for σ_t^2, and therefore for X_t, exists (cf. [32], Theorem 2) while it also allows us to calculate the extremal index for this model.

As is well known (see [1], p. 192, for example) the stationary distributions of both σ_t^2 and X_t^2 are regularly varying with index κ which is defined by the unique positive solution of the equation

$$E\left[(\alpha_1\epsilon_t^2)^\kappa\right] = 1$$

So, in our case, we may calculate κ from

$$\frac{1}{a}\int_0^a x^\kappa\, dx = 1 \quad \Leftrightarrow \quad \frac{a^\kappa}{\kappa+1} = 1.$$

Now, for calculating the extremal index θ_{σ^2} for $(\sigma_t^2)_{t\in\mathbb{Z}}$ in this model, the equation

$$\theta_{\sigma^2} = \lim_{n\to\infty}\lim_{u\to\infty} P(\max(\sigma_1^2,\ldots,\sigma_n^2) \le u \mid \sigma_0^2 > u)$$

may be used (its applicability to the ARCH(1) model is shown in [34]). The tail chain representation of the conditional time series gives us

$$\theta_{\sigma^2} = \lim_{n\to\infty} P\left(Y\alpha_1\epsilon_1^2 < 1,\ldots,Y\prod_{i=1}^{n}(\alpha_1\epsilon_i^2) < 1\right)$$

where $Y \sim \mathrm{Par}(\kappa)$ and the $\epsilon_i, i \in \mathbb{N}$, are i.i.d. according to the distribution specified above and independent of Y (cf. [34]). We may also write

$$\theta_{\sigma^2} = \lim_{n\to\infty} P\left(\log(Y)+\log(\alpha_1\epsilon_1^2) < 0,\ldots,\log(Y)+\sum_{i=1}^{n}\log(\alpha_1\epsilon_i^2) < 0\right)$$

$$= \lim_{n\to\infty} \int_{-\infty}^{0} \kappa e^{\kappa x} P\left(\max_{n\in\mathbb{N}} \left(\log(Y) + \sum_{i=1}^{n} Y_i \right) < 0 \,\middle|\, \log(Y) = -x \right) dx,$$

where $Y_1, Y_2, Y_3, \ldots$ are independent of Y and i.i.d. with the same distribution as $\log(\alpha_1\epsilon_1^2)$ (cf. [40], p. 41). The independence of Y and $Y_1, Y_2, \ldots$ then leads to

$$P\left(\max_{n\in\mathbb{N}} \left(\log(Y) + \sum_{i=1}^{n} Y_i \right) < 0 \,\middle|\, \log(Y) = -x \right) = P\left(\max_{n\in\mathbb{N}} \left(\sum_{i=1}^{n} Y_i \right) < x \right).$$

Next, define functions $Q_p : \mathbb{R} \to [0,1], p \in \mathbb{N}_0$, as $Q_0 \equiv 1$ and $Q_p(x) := P\left(\max_{1\le n\le p} \left(\sum_{i=1}^{n} Y_i\right) < x\right)$. Then (cf. [40])

$$Q_p(x) = \int_0^\infty Q_{p-1}(y) F^{\log(\alpha_1\epsilon_1^2)}(x - dy).$$

It follows from the definition of the Q_p that $Q_{p+1}(x) \le Q_p(x)$ for all $x \in \mathbb{R}$. Therefore, monotone convergence ensures that $Q_p(x)$ converges to a monotone function $Q(x) : \mathbb{R} \to [0,1]$ (cf. [24], Section 11.5). The limit Q then satisfies the so-called Wiener-Hopf equation

$$\begin{aligned} Q(x) &= \int_0^\infty Q(y) F^{\log(\alpha_1\epsilon_1^2)}(x - dy) \\ &= \int_0^\infty Q(y) \frac{1}{a} e^{x-y} \mathbb{1}_{(-\infty, \log(a)]}(x - y)\, dy \\ &= \frac{1}{a} \int_{\max(0, x-\log(a))}^\infty Q(y) e^{x-y}\, dy. \end{aligned}$$

In a first step, we will show that our limit Q is non-degenerate. Since $(\sum_{i=1}^{n} Y_i)_{n\in\mathbb{N}}$ is a random walk with step distribution function $F^{\log(\alpha_1\epsilon_1^2)}$ and

$$E(\log(\alpha_1\epsilon_1^2)) = a(\log(a) - 1) < 0$$

we may use that

$$Q(x) = P\left(\max_{n\ge 1} \sum_{i=1}^{n} \log(\alpha_1\epsilon_i^2) < x \right) \tag{3.7.3}$$

and

$$\begin{aligned} & P\left(\sum_{i=1}^{n} \log(\alpha_1\epsilon_i^2) > 0 \text{ for infinitely many } n \ge 1 \right) \\ = \; & P\left(\frac{1}{n} \sum_{i=1}^{n} \log(\alpha_1\epsilon_i^2) - E(\log(\alpha_1\epsilon_1^2)) > |E(\log(\alpha_1\epsilon_1^2))| \text{ infinitely often} \right) \\ = \; & 0 \end{aligned}$$

by the strong law of large numbers (cf. [24], Section 11.5). Therefore, there are almost surely only finitely many positive partial sums of the process $(\sum_{i=1}^{n} Y_i)_{n\in\mathbb{N}}$ and therefore the distribution function of $\max_{n\geq 1}\sum_{i=1}^{n}\log(\alpha_1\epsilon_i^2)$ is non-defective.

In the following, we will first analyze the explicit solution of

$$H(x) = \frac{1}{a}\int_{\max(0,x-\log(a))}^{\infty} H(y)e^{x-y}\,dy \tag{3.7.4}$$

and later explore its asymptotic properties.

- **Explicit solution of** (3.7.4)

  To simplify the following calculations, let $\tilde{H}(x) = e^{-x}H(x)$. We then find that $\tilde{H}$ satisfies

  $$\tilde{H}(x) \;=\; \frac{1}{a}\int_{\max(0,x-\log(a))}^{\infty} \tilde{H}(y)\,dy.$$

  The function $\tilde{H}(x)$ is therefore continuous on $\mathbb{R}$ and constant on $(-\infty,\log(a))$. For $x \geq \log(a)$ we may differentiate the last equation on both sides, where in $x = \log(a)$ we may restrict ourselves to taking the right-hand derivative. This leads us to

  $$\tilde{H}'(x) = -\frac{1}{a}\tilde{H}(x-\log(a)), \quad x \geq \log(a).$$

  Next, we have to solve this so-called differential-delay equation. Of course, the solution can only be determined up to a multiplicative constant. So, for $\tilde{H}(x) = y_0, x \leq \log(a)$, we find the solution as (cf. [16], p. 8)

  $$\tilde{H}(x) = y_0 \sum_{i=0}^{\lfloor \frac{x}{\log(a)} \rfloor} \left(-\frac{1}{a}\right)^i \frac{(x-i\log(a))^i}{i!}, \quad x > \log(a). \tag{3.7.5}$$

- **Asymptotic behavior of the solution of** (3.7.4)

  Since we know from (3.7.3) that one solution of (3.7.4) is a distribution function and the general solution is determined up to a multiplicative constant we may conclude that for any solution $H(x)$ of (3.7.4) with initial value y_0 there exists a constant $c(y_0) \in \mathbb{R}$ such that

  $$\lim_{x\to\infty} H(x) = c(y_0).$$

  In the next step we find the uniquely determined initial value x_0 which ensures that $H(x)$ is a distribution function. To this end, we analyze the asymptotic behavior of $H(x)$ as $x \to \infty$.

  Note first that for $x > \log(a)$

  $$H(x) - H(\log(a)) \;=\; \int_{\log(a)}^{x} H'(y)\,dy$$

$$\begin{aligned} &= \int_{\log(a)}^{x} H(y) - H(y-\log(a))\,dy \\ &= \int_{x-\log(a)}^{x} H(y)\,dy - \int_{0}^{\log(a)} H(y)\,dy \\ &= \int_{x-\log(a)}^{x} H(y)\,dy - y_0(a-1). \end{aligned}$$

Therefore, the limit $c(y_0) = \lim_{x\to\infty} H(x)$ has to satisfy the equation

$$c(y_0) - y_0 a = \log(a)c(y_0) - y_0(a-1) \quad \Leftrightarrow \quad c(y_0) = \frac{y_0}{1-\log(a)}.$$

Putting it the other way round, we may thus conclude that $x_0 = 1 - \log(a)$ leads to a solution of (3.7.4) which actually is a distribution function. This finally gives

$$Q(x) = \begin{cases} (1-\log(a))e^x, & x \le \log(a) \\ (1-\log(a))e^x \sum_{i=0}^{\lfloor \frac{x}{\log(a)} \rfloor} \left(-\frac{1}{a}\right)^i \frac{(x-i\log(a))^i}{i!}, & x > \log(a). \end{cases}$$

The extremal index may then be calculated as

$$\theta_{\sigma^2} = \int_{-\infty}^{0} \kappa e^{\kappa x} Q(x)\,dx = \int_{-\infty}^{0} \kappa e^{(1+\kappa)x}(1-\log(a))\,dx = (1-\log(a))\frac{\kappa}{1+\kappa}.$$

Since the extremal index is not affected by monotone transformations we may conclude that

$$\theta_{\sigma^2} = \theta_\sigma$$

Calculating the extremal index or other characteristics of $(\sigma_t^2)_{t\in\mathbb{Z}}$ belongs to the standard analysis of a single time series and can in many cases be simplified by the use of the tail chain approach. Theorem 3.5.2 now allows us to use the tools we have already elaborated for the single time series case for the analysis of two adjoint time series. In our example we may thus exploit the calculations we have made so far to work out characteristics which depend on the joint extremal behavior of $(X_t^2)_{t\in\mathbb{Z}}$ and $(\sigma_t^2)_{t\in\mathbb{Z}}$.

We have seen in section 3.6 that in some cases it might be necessary to condition on the event $X_0^2 > x$ for $x \to \infty$ while looking at the time series σ_t^2, thus finding the limit

$$\theta_{\sigma^2,X^2} := \lim_{n\to\infty} \lim_{x\to\infty} P\left(\sigma_1^2 < x, \ldots, \sigma_n^2 < x, |X_0^2 > x\right). \tag{3.7.6}$$

To calculate the value of θ_{σ^2,X^2} we have to take into account that, given the condition that $X_0^2 = \sigma_0^2\epsilon_1^2$ is large, ϵ_1 no longer has the same distribution as $\epsilon_2, \epsilon_3, \ldots$. However, this condition does not influence the distribution of the ϵ_t with $t \ge 2$. Therefore we may

still make use of our function $Q(x)$ but use a slightly different "starting distribution" to calculate θ_{σ^2,X^2}:

$$\begin{aligned}
\theta_{\sigma^2,X^2} &= \lim_{n\to\infty}\lim_{x\to\infty} P\left(\sigma_1^2 < x,\ldots,\sigma_n^2 < x | X_0^2 > x\right) \\
&= \lim_{n\to\infty}\lim_{x\to\infty}\int_{-\infty}^{0} P\left(\max_{2\le i\le n}\left(\sum_{k=2}^{i}\log\left(\alpha_1\epsilon_k^2\right)\right) \le -s \,\middle|\, \log\left(\frac{\sigma_1^2}{x}\right) = s\right) \\
&\qquad\qquad P^{\log\left(\frac{\sigma_1^2}{x}\right)\,\middle|\,\log\left(\frac{X_0^2}{x}\right)>0}(ds) \\
&= \lim_{n\to\infty}\int_{-\infty}^{0} Q_{n-1}(-s)\lim_{x\to\infty} P^{\log\left(\frac{\sigma_1^2}{x}\right)\,\middle|\,\log\left(\frac{X_0^2}{x}\right)>0}(ds). \qquad (3.7.7)
\end{aligned}$$

For the conditional distribution needed for the integral above, we get

$$\begin{aligned}
& P\left(\log\left(\frac{\sigma_1^2}{x}\right) > u \,\middle|\, \log\left(\frac{X_0^2}{x}\right) > 0\right) \\
= \ & \frac{P\left(o(1) + \alpha_1\sigma_0^2\epsilon_1^2 > xe^u, \sigma_0^2\epsilon_1^2 > x\right)}{P(\sigma_0^2\epsilon_1^2 > x)} \\
= \ & \frac{\int_{-\infty}^{\infty}\int_{\frac{x}{s}\max\left(1,\frac{e^x}{\alpha_1}\right)} F^{\sigma_0^2}(dv)F^{\epsilon_1^2}(ds)}{P(\sigma_0^2\epsilon_1^2 > x)}
\end{aligned}$$

Now, regular variation of $\overline{F}^{\sigma_0^2}$ and Breiman's lemma (cf. [12]) lead us to

$$\begin{aligned}
& \lim_{x\to\infty} P\left(\log\left(\frac{\sigma_1^2}{x}\right) > u \,\middle|\, \log\left(\frac{X_0^2}{x}\right) > 0\right) \\
= \ & \frac{\int_{-\infty}^{\infty} s^{\kappa}\left(\max\left(1,\frac{e^u}{\alpha_1}\right)\right)^{-\kappa} F^{\epsilon_1^2}(ds)}{E\left(\epsilon_1^{2\kappa}\right)} \\
= \ & \left(\max\left(1,\frac{e^x}{\alpha_1}\right)\right)^{-\kappa} \\
= \ & \begin{cases} 1, & x < \log(\alpha_1) \\ \alpha_1^{\kappa}e^{-\kappa x}, & x \ge \log(\alpha_1) \end{cases}. \qquad (3.7.8)
\end{aligned}$$

The value of θ_{σ^2,X^2} now depends on the value of α_1. It is only greater than zero if $\alpha_1 < 1$ since $\sigma_1^2 \sim \alpha_1 X_0^2$ holds for large values of X_0^2. Now, using (3.7.7) and (3.7.8), we get

$$\theta_{\sigma^2,X^2} = \lim_{n\to\infty}\int_{\log(\alpha_1)}^{0} Q_{n-1}(-x)\alpha_1^{\kappa}\kappa e^{-\kappa x}\,dx = \int_{\log(\alpha_1)}^{0} Q(-x)\alpha_1^{\kappa}\kappa e^{-\kappa x}\,dx.$$

In contrast to θ, we need to integrate $Q(x)$ over the positive axis which complicates our calculations since $Q(x)$ is defined in sections. As an example and to keep things

simple, we may choose $\alpha_1 = a^{-1}$, such that we only have to integrate over $(0, \log(a))$. In this case we get

$$\begin{aligned}\theta_{\sigma^2,X^2} &= a^{-\kappa}\kappa \int_0^{\log(a)} e^{\kappa x} e^x (1 - \log(a))\, dx \\ &= \left(a - a^{-\kappa}\right) \frac{\kappa}{1+\kappa}(1 - \log(a)).\end{aligned}$$

For this model specification it means that

$$\theta_{\sigma^2,X^2} = \left(a - a^{-\kappa}\right) \theta_{X^2} = \left((1+\kappa)^{1/\kappa} - (1+\kappa)^{-1}\right) \theta_{X^2}$$

which leads to

$$\frac{\theta_{\sigma^2,X^2}}{\theta_{\sigma^2}} = (1+\kappa)^{1/\kappa} - (1+\kappa)^{-1} > 1$$

for all $\kappa > 0$. A possible interpretation is the following: $|X_0| > x$ is a better indication for large values of σ_t in the future than $|\sigma_0| > x$ would be, although in this context this can mainly be explained by the different scales of the two time series.

This example clarifies that with the help of Theorem 2.3.2 the same procedure for all questions involving the limit distribution

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_- - m}}{x}, \ldots, \frac{Y_{s_+ + n}}{x} \mid |X_0| > x\right)$$

can be handled by

1. exploring the tail chain for the process $(Y_t)_{t\in\mathbb{Z}}$
2. exploring the "starting distribution"

$$\lim_{x\to\infty} \mathcal{L}\left(\frac{Y_{-s_- - 1}}{x}, \ldots, \frac{Y_{s_+}}{x} \mid |X_0| > x\right)$$

3. joining those two pieces together according to Theorem 2.3.2.

However, the purely analytical approach is limited to relatively simple cases due to the complexity of the involved calculations (for example, the solution to the Wiener-Hopf equation as used above is only known explicitly in a few cases). The focus on the application of the results should therefore rather lie on simulations which have proven to be a powerful tool to explore extremal characteristics in the GARCH(1,1) case. Again, in these simulations we can notice the interplay of heavy-tailed and light-tailed random variables. For the single time series model as described in [39] we are able to simulate from the tail of $(Y_t)_{t\in\mathbb{Z}}$ by simulating one heavy-tailed Pareto random variable and then multiply with (for the most applications:) light-tailed random increments. For the two time series model which we have analyzed we are able to simulate from the tail of $(Y_t)_{t\in\mathbb{Z}}$ given that $|X_0|$ is extreme, where we start with simulating from

$(Y^X_{-s_- -1}, \ldots, Y^X_{s_+})$. By Lemma 3.4.2 this vector is the product of a Pareto random variable and a in most cases light-tailed random vector which are independent. Again, we have one heavy-tailed component which determines the scale of the tail-process and light-tailed increments which determine its further direction.

Chapter 4

A New Way of Assessing κ in RDEs

4.1 Introduction

In this chapter we analyze *multivariate* random difference equations (RDEs), which are of the form

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{N}, \tag{4.1.1}$$

for $\mathbf{Y}_t \in \mathbb{R}^d$, where $\mathbf{A}_t$ are random $\mathbb{R}^{d\times d}$-matrices and $\mathbf{B}_t$ are random $\mathbb{R}^d$-vectors. We assume that $(\mathbf{A}_t, \mathbf{B}_t)_{t\in\mathbb{N}}$ are i.i.d. and independent of $\mathbf{Y}_0$. We are interested in the existence of a stationary distribution μ^* of $\mathbf{Y}_t, t \in \mathbb{N}_0$, which means that

$$\mathbf{Y}^* \overset{d}{=} \mathbf{A}\mathbf{Y}^* + \mathbf{B}$$

holds for $\mathbf{Y}^*$ with $\mathcal{L}(\mathbf{Y}^*) = \mu^*$ and $(\mathbf{A}, \mathbf{B}) \overset{d}{=} (\mathbf{A}_1, \mathbf{B}_1)$ independent of $\mathbf{Y}^*$. Of special interest in this setting are conditions for $(\mathbf{A}_1, \mathbf{B}_1)$ to guarantee the existence of a stationary distribution and specific characteristics of it.

In a seminal paper [28], Kesten gave those conditions for $(\mathbf{A}_1, \mathbf{B}_1)$ (mainly for the case of non-negative entries of $\mathbf{A}_1$ and $\mathbf{B}_1$) and furthermore derived a property of the stationary distribution μ^* which could later be shown to imply multivariate regular variation, cf. [10]. While we have already come across multivariate regular variation in Chapter 3, we will use a slightly different but equivalent definition in this context. A random vector $\mathbf{Y}^* \in \mathbb{R}^d$ is multivariate regularly varying with index $\kappa > 0$ if and only if there exists a random vector $\boldsymbol{\theta} \in \mathbb{S}^{d-1} := \{\mathbf{x} \in \mathbb{R}^d \big| \|\mathbf{x}\| = 1\}$ and $C > 0$ such that

$$\frac{P\left(\|\mathbf{Y}^*\| > uy, \frac{\mathbf{Y}^*}{\|\mathbf{Y}^*\|} \in \cdot\right)}{P\left(\|\mathbf{Y}^*\| > y\right)} \xrightarrow{v} Cu^{-\kappa} P(\boldsymbol{\theta} \in \cdot), \quad y \to \infty, \tag{4.1.2}$$

where $\|\cdot\|$ denotes the l_2-norm on $\mathbb{R}^d$ (although, in principle, (4.1.2) holds for any norm on $\mathbb{R}^d$) and $\xrightarrow{v}$ stands for vague convergence in $M_+(\bar{\mathbb{R}}_+ \times \mathbb{S}^{d-1})$. The equivalence to the definition which we have used in Chapter 3 becomes clear if one compares (4.1.2) and (3.4.10). Kesten's Theorem now says that under suitable assumptions about the

distribution of $(\mathbf{A}_1, \mathbf{B}_1)$ a stationary solution $\mathbf{Y}^*$ to (4.1.1) exists and satisfies (4.1.2). Both the law of $\boldsymbol{\theta}$, called the spectral measure, and the index of regular variation κ are important characteristics of the stationary time series defined by (4.1.1), in particular with regard to its extremal behavior. For example, under the same assumptions about $(\mathbf{A}_t, \mathbf{B}_t)_{t\in\mathbb{N}}$ as in [28] it was shown in [35] that for a stationary process $(\mathbf{Y}_t^*)_{t\in\mathbb{N}}$ satisfying (4.1.1) the distribution of $n^{-1/\kappa}\max(\mathbf{Y}_0^*, \mathbf{Y}_1^*, \ldots, \mathbf{Y}_n^*)$ converges weakly to a non-degenerate limit distribution as $n \to \infty$.

However, for most cases and especially for higher-dimensional matrices, the formula for the derivation of κ which is stated in [28] is not analytically solvable. This instance is especially problematic since κ is not only of interest in itself but needed for lots of further derivations and simulations about the extremal behavior of a stationary solution to (4.1.1), for example for the simulations related to GARCH$(1,1)$ processes which have been mentioned in Chapter 4.5 and executed in [15], Section 5.3. This chapter deals with a new method for assessing the index of regular variation which adopts some defining properties of κ used in the proofs of [28] and determines its value with the help of Monte Carlo simulations.

The precise setting of the time series we deal with will be stated in Section 4.2. This mainly recaptures the conditions from [28] and states the most commonly used formula for κ, which is both analytically and numerically unfeasible for most applications. In Section 4.3 we will then describe the theoretical base for our new method to determine κ. Section 4.4 discusses the numerical implementation and Section 4.5 shows some simulation results.

4.2 Setting

In this section, we will introduce the notation and the assumptions made in [28]. Basically, we will make the same assumptions as have been made there. Since the assumptions are quite technical we will also give an example of a class of time series for which the assumptions are satisfied. The original assumptions and the main result of [28] have been properly reformulated in [2] and shall, for convenience, be restated here.

Theorem 4.2.1 (cf. [2], Theorem 2.4) *Let $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of $d \times d$ matrices $\mathbf{A}_n$ with non-negative entries and d-dimensional non-negative-valued random vectors $\mathbf{B}_n \neq \mathbf{0}$ almost surely. Assume that the following conditions hold:*

(a) For some $\epsilon > 0, E\|\mathbf{A}_1\|_{\text{op}}^{\epsilon} < 1$, where $\|\cdot\|_{\text{op}}$ denotes the operator norm on $\mathbb{R}^{d\times d}$ corresponding to the l_2-norm, meaning that $\|\mathbf{A}\|_{\text{op}} := \sup_{\|\mathbf{x}\|=1}\|\mathbf{A}\mathbf{x}\|$.

(b) $\mathbf{A}_1$ has no zero rows almost surely.

(c) The set

$$\{\ln\|\mathbf{a}_n \cdot \ldots \cdot \mathbf{a}_1\|_{\text{op}} : n \geq 1, \mathbf{a}_n \cdot \ldots \cdot \mathbf{a}_1 > 0 \text{ and } \mathbf{a}_n, \ldots, \mathbf{a}_1 \in \text{support of } P^{\mathbf{A}_1}\}$$

generates a dense group in $\mathbb{R}$.

(d) *There exists a* $\kappa_0 > 0$ *such that*

$$E\left(\min_{i=1,\ldots,d}\sum_{j=1}^{d} A_{ij}\right)^{\kappa_0} \geq d^{\kappa_0/2} \tag{4.2.1}$$

and

$$E\left(\|\mathbf{A}_1\|_{\text{op}}^{\kappa_0} \ln^+ \|\mathbf{A}_1\|_{\text{op}}\right) < \infty. \tag{4.2.2}$$

Then the following statements hold:

1. *There exists a unique solution* $\kappa^* \in (0, \kappa_0]$ *to the equation*

$$0 = \lim_{n\to\infty} \frac{1}{n} \ln E\|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1\|_{\text{op}}^{\kappa^*}. \tag{4.2.3}$$

2. *There exists a unique probability measure* μ^* *for* $\mathbf{Y}_0$ *such that the process* $(\mathbf{Y}_t)_{t\in\mathbb{N}}$ *as defined by* (4.1.1) *is stationary.*

3. *If* $E\|\mathbf{B}_1\|^{\kappa^*} < \infty$, *then there exists a random variable* $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$ *such that* $\mathbf{Y}^*$ *with probability distribution* μ^* *satisfies Equation* (4.1.2) *with* $\kappa = \kappa^*$.

Remark 4.2.2.

(a) The first condition given here is a sufficient condition which ensures the existence of a negative Lyapunov exponent γ for the sequence $\mathbf{A}_1, \mathbf{A}_2, \ldots$. A weaker condition, which is obviously more difficult to prove, would be that

$$\gamma := \inf\left\{\frac{1}{n} E \ln \|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1\|_{\text{op}}, n \in \mathbb{N}\right\} < 0$$

holds. If $E \ln^+ \|\mathbf{A}_1\|_{\text{op}} < \infty$ then

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1\|_{\text{op}} \quad \text{a.s.}, \tag{4.2.4}$$

cf. [2], p. 97. Furthermore, if a non-negative random matrix $\mathbf{A}_1$ satisfies $E \ln^+ \|\mathbf{A}_1\|_{\text{op}} < \infty$ and assumptions (b) and (c) of Theorem 4.2.1, then also

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1 \mathbf{x}\| \quad \text{a.s.}, \quad \forall\, \mathbf{x} \in \mathbb{S}_+^{d-1} := \{x \in \mathbb{R}_+^d : \|x\| = 1\}, \tag{4.2.5}$$

cf. [28], p. 224. For a specific choice of $\mathbf{A}_1$ the question whether or not $\gamma < 0$ can often be answered by Monte Carlo simulations and the help of (4.2.4) or (4.2.5).

(b) Inequality (4.2.1) is a sufficient but not necessary condition, cf. Remark 4.3.2 (b).

(c) The third part of the proposition follows from Corollaries 2.6 and 2.7 in [2] if κ^* is not an even integer. For general κ^* this follows from [10], p. 704.

Corollary 4.2.3 *Let the conditions of Theorem 4.2.1 be satisfied and let in addition* $\mathbf{A}_1$ *have no zero columns almost surely and*

$$E\left(\min_{j=1,\dots,d}\sum_{i=1}^{d} A_{ij}\right)^{\kappa_0} \geq d^{\kappa_0/2}. \tag{4.2.6}$$

Then the statements of the Theorem hold not only for the RDE defined by (4.1.1) but also for the RDE defined by

$$\tilde{\mathbf{Y}}_t = \mathbf{A}_t^t\tilde{\mathbf{Y}}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{N}. \tag{4.2.7}$$

The index of regular variation is the same for both RDEs.

Proof. Since $\|\mathbf{A}\|_{\text{op}} = \|\mathbf{A}^t\|_{\text{op}}$ for all $\mathbf{A} \in \mathbb{R}^{d\times d}$ the additional assumptions imply that the conditions of Theorem 4.2.1 are satisfied for the RDE defined in (4.2.7). Since the index of regular variation is determined by (4.2.3) it is the same for both RDEs. □

In the following, we will often assume that both the assumptions of Theorem 4.2.1 and of Corollary 4.2.3 are satisfied. The reason for this is as follows: Kesten's proofs which we will pick up eventually use multiplication from the left when multiplying a (row) vector with a matrix, thus looking at $\mathbf{x}^t\mathbf{A}_1$ for $\mathbf{x} \in \mathbb{R}^d$. Since it seems to be confusing to multiply from the right in the definition of the RDE but to multiply from the left for the rest of the analysis, we decided to unify the approach and always multiply from the right in the following. Since

$$\mathbf{A}_1\mathbf{x} = (\mathbf{x}^t\mathbf{A}_1^t)^t$$

for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{A}_1 \in \mathbb{R}^{d\times d}$, everything that Kesten derives for $\mathbf{A}_1^t$ by multiplying from the left holds for $\mathbf{A}_1$ and multiplication from the right. Therefore, the assumptions of Theorem 4.2.1 and Corollary 4.2.3, taken together, ensure that Kesten's methods to derive κ^* from the law of $\mathbf{A}_1$ also work for $\mathbf{A}_1^t$ and therefore both multiplication from the left and from the right are feasible in our proofs. For details, see the proof of Proposition 4.3.1.

Although Theorem 4.2.1 guarantees the existence of a κ^* which equals the index of regular variation of the stationary solution to (4.1.1), Equation (4.2.3) gets quite cumbersome once $d > 1$. For the case $d = 1$ and $\mathbf{A}_i = A_i \in \mathbb{R}^+, i \in \mathbb{N}$, Equation (4.2.3) simplifies to

$$E\left(A_1^{\kappa^*}\right) = 1.$$

For $d > 1$, simpler results for the determination of κ^* can be derived under additional assumptions about the form of $\mathbf{A}_1$ in special cases. One of these examples has been analyzed in [31] and deals with the GARCH$(1, 1)$ process. Before we state the explicit

formula in this case we shall explain how GARCH(p,q) models fit in the setting of RDEs. Recall from Chapter 3 that a GARCH(p,q) process satisfies the equations

$$X_t = \sigma_t \epsilon_{t+1}, \quad t \in \mathbb{Z}, \tag{4.2.8}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \tag{4.2.9}$$

with $\alpha_i, \beta_j \geq 0$ (cf. the beginning of Section 3.3 for further details). Let now

$$\mathbf{X}_t = \left(\sigma_{t+1}^2, \ldots, \sigma_{t-\tilde{q}+2}^2, X_t^2, \ldots, X_{t-\tilde{p}+2}^2\right)^t \in \mathbb{R}^{\tilde{p}+\tilde{q}-1}, \quad t \in \mathbb{Z}, \tag{4.2.10}$$

with $\tilde{p} := \max(p,2)$ and $\tilde{q} := \max(q,2)$. Set

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1\epsilon_{t+1}^2 + \beta_1 & \beta_2 & \ldots & \beta_{\tilde{q}-1} & \beta_{\tilde{q}} & \alpha_2 & \alpha_3 & \ldots & \alpha_{\tilde{p}-1} & \alpha_{\tilde{p}} \\ 1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \epsilon_{t+1}^2 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{(\tilde{p}+\tilde{q}-1)\times(\tilde{p}+\tilde{q}-1)} \tag{4.2.11}$$

for $t \in \mathbb{Z}$ with $\epsilon_t, t \in \mathbb{Z}$, i.i.d. with $E(\epsilon_0) = 0$, $\mathrm{Var}(\epsilon_0) = 1$, $\beta_{q+1} = \beta_2 = 0$ if $q \leq 1$ and $\alpha_2 = 0$ if $p = 1$ and

$$\mathbf{B}_t = (\alpha_0, 0, \ldots, 0)^t \in \mathbb{R}^{\tilde{p}+\tilde{q}-1}, \quad t \in \mathbb{Z}, \tag{4.2.12}$$

then

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{Z}.$$

Thus, if we are only interested in the absolute values of a GARCH(p,q) process, we may write it as an RDE (cf. [1], p. 49). Furthermore, one can show (cf. [2], Theorem 3.1 (B)) that an iterated version of this RDE satisfies the conditions of Theorem 4.2.1 if

1) the Lyapunov exponent of $\mathbf{A}_1$ is negative, which is for example the case if $\alpha_0 > 0$, $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ or $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$ with the further assumption that the α_i, β_j are all positive and ϵ_1 has infinite support and no atom at zero (cf. [2], Remark 3.2),

2) the distribution of ϵ_1 has a positive density on $\mathbb{R}$ such that $E|\epsilon_1|^h < \infty$ for all $h < h_0$ and $E|\epsilon_1|^{h_0} = \infty$ for some $h_0 \in (0, \infty]$,

3) Not all of the parameters α_i, β_j vanish.

Here, with iterated version we mean that there exists $m \in \mathbb{N}$ with

$$\tilde{\mathbf{A}}_1 := \mathbf{A}_m \cdot \cdots \cdot \mathbf{A}_1, \quad \tilde{\mathbf{A}}_2 := \mathbf{A}_{2m} \cdot \cdots \cdot \mathbf{A}_{m+1}, \quad \dots$$

and

$$\begin{aligned} \tilde{\mathbf{B}}_1 &:= \mathbf{B}_m + \sum_{i=1}^{m-1} \mathbf{A}_m \cdot \cdots \cdot \mathbf{A}_{m-i+1} \mathbf{B}_{m-i}, \\ \tilde{\mathbf{B}}_2 &:= \mathbf{B}_{2m} + \sum_{i=1}^{m-1} \mathbf{A}_{2m} \cdot \cdots \cdot \mathbf{A}_{2m-i+1} \mathbf{B}_{2m-i}, \quad \dots \end{aligned}$$

such that

$$\tilde{\mathbf{X}}_t := \mathbf{X}_{mt} = \tilde{\mathbf{A}}_t \tilde{\mathbf{X}}_{t-1} + \tilde{\mathbf{B}}_t, \quad t \in \mathbb{Z},$$

and $(\tilde{\mathbf{A}}_1, \tilde{\mathbf{B}}_1)$ satisfies the assumptions of Theorem 4.2.1 (cf. [2], p. 108). For a concrete choice of the parameters and the distribution of the innovations $\epsilon_t, t \in \mathbb{Z}$, one can often show that the assumptions of Theorem 4.2.1 and Corollary 4.2.3 are already satisfied for $m = 1$ (cf. the numerical example in Section 4.5).

Note that the process $(\mathbf{V}_t)_{t\in\mathbb{Z}} := (\sigma_{t+1}, \dots, \sigma_{t-\tilde{q}+2}, |X_t|, \dots, |X_{t-\tilde{p}+2}|)_{t\in\mathbb{Z}}$ is multivariate regularly varying with index 2κ if the squared process $(\mathbf{X_t})_{t\in\mathbb{Z}}$ is multivariate regularly varying with index κ (cf. [2], Corollary 3.5).

In the special case of a GARCH$(1,1)$ process which satisfies the above conditions one can show (cf. [31], Theorem 2.1) that κ^* is determined by the equation

$$E\left[\left(\alpha_1 \epsilon_0^2 + \beta_1\right)^{\kappa^*}\right] = 1. \tag{4.2.13}$$

For a given distribution of ϵ_0 Equation (4.2.13) can properly be solved by analytical or numerical methods. So far, the GARCH$(1,1)$ process and, as a special case, the ARCH(1) process are two of very few examples for common RDEs which allow for a simple calculation of κ^*.

The straightforward approach of trying to evaluate the right hand side of Equation (4.2.3) for different values of κ in order to find the κ^* which solves (4.2.3) faces two problems. First, the expression on the right hand side of Equation (4.2.3) can only be evaluated for a fixed number n of matrices and we do not know about the convergence rate of the expression and therefore if this approximation is feasible. Second, at least in some cases, the operator norm of the product of matrices shows a heavy tailed behavior, which makes Monte Carlo estimation of deduced expressions especially difficult.

4.3 Theoretical background

In this section we will derive a characterisation of κ^* which is alternative to (4.2.3). This characteristic is used in the proof of Theorem 3 in [28] but not stated explicitly in the results.

Proposition 4.3.1 *Let $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence which satisfies the conditions of Theorem 4.2.1 and Corollary 4.2.3. Then for every $\kappa \in (0, \kappa_0]$ there exists a pair $(\rho_\kappa, \nu_\kappa)$ where $\rho_\kappa > 0$ and ν_κ is a probability measure on $\mathbb{S}_+^{d-1}$ such that*

$$\int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1\mathbf{x}\|^\kappa f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}) = \rho_\kappa \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})\, \nu_\kappa(\mathrm{d}\mathbf{x}) \tag{4.3.1}$$

holds for all continuous functions f on $\mathbb{S}_+^{d-1}$. For all pairs $(\rho_\kappa, \nu_\kappa)$ which satisfy (4.3.1) for a given value of $\kappa \in (0, \kappa_0]$, the value of ρ_κ is uniquely determined by κ. Compared with Theorem 4.2.1 it holds that $\kappa = \kappa^*$ is the unique solution to $\rho_\kappa = 1$.

Proof. Let $\kappa \in (0, \kappa_0]$. The existence of a tupel $(\rho_\kappa, \nu_\kappa)$ which satisfies (4.3.1) is shown in step 1 of the proof of Theorem 3 in [28], (cf. Equation (2.65)) by using a fixed point argument. While this fixed point argument only shows the existence, it is shown later in the proof (cf. step 4) that for all $\kappa \in (0, \kappa_0]$ every fixed point satisfies

$$\log(\rho_\kappa) = \lim_{n\to\infty} \frac{1}{n} \log\left(E\|\mathbf{A}_n \cdot \dots \cdot \mathbf{A}_1\|_{\mathrm{op}}^\kappa\right) \tag{4.3.2}$$

and the value is therefore unique. Moreover, it is shown in step 4 of the proof that the function ρ_κ is log-convex and continuous on $(0, \kappa_0]$ and that $\log \rho_\kappa < 0$ on $(0, \delta)$ for some $\delta > 0$. Therefore, there exists only one $\kappa > 0$ with $\rho_\kappa = 1$. It is shown in step 4 of Kesten's proof that this κ is the same as κ_1 in Kesten's Theorem 3 and therefore as κ^* in our setting. Note that Kesten uses multiplication from the left when dealing with matrices so Equation (2.65) in [28] is not equal to our Equation (4.3.1) but stated as

$$\int_{\mathbb{S}_+^{d-1}} E\left[\left\|\mathbf{x}^t\mathbf{A}_1\right\|^\kappa f\left(\frac{\mathbf{x}^t\mathbf{A}_1}{\|\mathbf{x}^t\mathbf{A}_1\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}) = \rho_\kappa \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})\, \nu_\kappa(\mathrm{d}\mathbf{x}).$$

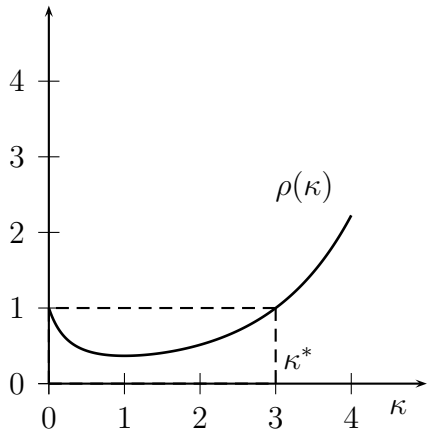
Note, however, that we may write the left hand side of our Equation (4.3.1) as

$$\begin{aligned}
&\int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1\mathbf{x}\|^\kappa f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}) \\
= &\int_{\mathbb{S}_+^{d-1}} E\left[\left\|(\mathbf{x}^t\mathbf{A}_1^t)^t\right\|^\kappa f\left(\frac{(\mathbf{x}^t\mathbf{A}_1^t)^t}{\|(\mathbf{x}^t\mathbf{A}_1^t)^t\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}) \\
= &\int_{\mathbb{S}_+^{d-1}} E\left[\left\|\mathbf{x}^t\mathbf{A}_1^t\right\|^\kappa f\left(\frac{(\mathbf{x}^t\mathbf{A}_1^t)^t}{\|\mathbf{x}^t\mathbf{A}_1^t\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}).
\end{aligned}$$

which corresponds to Kesten's approach applied for the RDE (4.2.7). Since we assume that both $\mathbf{A}_1$ and $\mathbf{A}_1^t$ satisfy the assumptions of Theorem 4.2.1, we may use Kesten's conclusions for the RDE (4.2.7). Since ρ_κ satisfies Equation (4.3.2) which is left unchanged by transposing the matrices, its value is the same whether we analyze the RDE (4.1.1) or (4.2.7). □

Remark 4.3.2.

(a) Since the function $\rho(\kappa) := \rho_\kappa$ is log-convex, continuous on $(0, \kappa_0]$ and $\log \rho_\kappa < 0$ on $(0, \delta)$ for some $\delta > 0$, a typical representative of it can be sketched as follows:



(b) The assumption (4.2.1) is only needed to ensure that $\rho_{\kappa_0} \geq 1$ (cf. [28], Equation (2.66)) which, together with the abovementioned properties of $\rho(\kappa)$, guarantees that κ^* exists. If all assumptions of Theorem 4.2.1 but Inequality (4.2.1) are satisfied, Proposition 4.3.1 still holds. If furthermore there exists a $\kappa \in (0, \kappa_0]$ such that $\rho_\kappa = 1$, the statement of Theorem 4.2.1 follows, cf. the proof of Theorem 3 in [28].

As for the evaluation of ρ_κ, we get the equation

$$\int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1\mathbf{x}\|^\kappa\right] \nu_\kappa(\mathrm{d}\mathbf{x}) = \rho_\kappa \int_{\mathbb{S}_+^{d-1}} \nu_\kappa(\mathrm{d}\mathbf{x}) = \rho_\kappa, \tag{4.3.3}$$

with setting $f \equiv 1$ in (4.3.1). We will use this equation for the determination of κ^*. The idea is to try out different values of κ, simulate from the (or: one of the) respective measure(s) ν_κ and evaluate the expression on the left hand side of (4.3.3) via Monte Carlo simulation to derive a Monte Carlo estimator $\hat{\rho}_\kappa$ of ρ_κ. From the different values of κ the one with $\hat{\rho}_\kappa \approx 1$ gives an estimate for κ^*. Unfortunately, the simulation from the measure ν_κ turns out to be difficult. Nevertheless, we can make use of the following fact.

Proposition 4.3.3 *Let $(\mathbf{A}_n, \mathbf{B}_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence which satisfies the conditions of Theorem 4.2.1 and Corollary 4.2.3 and let a constant $C > 0$ exist such that $\|\mathbf{A}_1\|_{\mathrm{op}} \leq C$ almost surely. Let $\kappa \in (0, \kappa_0]$ and let $(\rho_\kappa, \nu_\kappa)$ be as in Proposition 4.3.1. Define three independent random variables $\boldsymbol{\theta}_\kappa \in \mathbb{S}_+^{d-1}$, $\mathbf{A} \in \mathbb{R}^{d\times d}$ and $U \in [0, 1]$ with*

$$\boldsymbol{\theta}_\kappa \sim \nu_\kappa, \quad \mathbf{A} \overset{d}{=} \mathbf{A}_1 \quad \textit{and} \quad U \sim \mathrm{Unif}[0, 1].$$

Then,

$$P(\boldsymbol{\theta}_\kappa \in S) = P\left(\frac{\mathbf{A}\boldsymbol{\theta}_\kappa}{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|} \in S \middle| U \leq \frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\right) \tag{4.3.4}$$

holds for all $S \in \mathbb{B}^d\big|_{\mathbb{S}_+^{d-1}}$.

Proof. The proof follows the ideas of [3], p. 1075. Let f be a continuous function on $\mathbb{S}_+^{d-1}$. Then,

$$\begin{aligned}
E\left(f\left(\frac{\mathbf{A}\boldsymbol{\theta}_\kappa}{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|}\right)\middle| U \leq \frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\right) &= \frac{E\left(f\left(\frac{\mathbf{A}\boldsymbol{\theta}_\kappa}{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|}\right)\mathbb{1}_{\{U\leq\frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\}}\right)}{P\left(U \leq \frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\right)} \\
&= \frac{E\left(f\left(\frac{\mathbf{A}\boldsymbol{\theta}_\kappa}{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|}\right)\frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\right)}{E\left(\frac{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa}{C^\kappa}\right)} \\
&= \frac{E\left(f\left(\frac{\mathbf{A}\boldsymbol{\theta}_\kappa}{\|\mathbf{A}\boldsymbol{\theta}_\kappa\|}\right)\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa\right)}{E\left(\|\mathbf{A}\boldsymbol{\theta}_\kappa\|^\kappa\right)} \\
&= \frac{\rho_\kappa E(f(\boldsymbol{\theta}_\kappa))}{\rho_\kappa} = E(f(\boldsymbol{\theta}_\kappa)),
\end{aligned}$$

where the penultimate equality follows from (4.3.3) and (4.3.1). □

Remark 4.3.4. Proposition 4.3.3 is a modification of an idea found in [3], p. 1075, where an algorithm has been proposed to sample from ν_{κ^*}. It has been shown there (cf. [3], Proposition 5.1) that ν_{κ^*} is the spectral measure of the stationary solution to (4.1.1), that means the probability distribution of $\boldsymbol{\theta}$ in (4.1.2). The idea of Basrak and Seger's algorithm is to use Markov chain Monte Carlo to simulate from ν_{κ^*}. Unfortunately, there is a mistake in their proof regarding the stationary distribution of the underlying Markov chain. As a consequence, their algorithm does, in most cases, not allow us to sample from the measure ν_{κ^*}.

To be more precise, the Markov chain defined in [3] is given in terms of a simulation algorithm of the following form:

```
Start with x_0 ∈ S_+^{d-1},
repeat
{
sample A from the law of A_1,
set Y = Ax_i/‖Ax_i‖,
with probability ‖Ax_i‖^{κ*}/C^{κ*}, accept Y as x_{i+1}
}
until Y is accepted
```

However, sticking to the same value of $\mathbf{x}_i$ at each proposal for the algorithm prevents it in most cases from having ν_{κ^*} as stationary distribution, in contrast to the statement in

[3]. The algorithm still works in special cases, for example if $\mathbf{A}_1$ is a random orthogonal matrix (B. Basrak, J. Segers, private communication), but for a continuous function f on $\mathbb{S}_+^{d-1}$ it holds in general that

$$E(f(\mathbf{x}_{i+1})) = \int_{\mathbb{S}_+^{d-1}} \frac{E\left(f\left(\frac{\mathbf{A}\mathbf{x}}{\|\mathbf{A}\mathbf{x}\|}\right)\|\mathbf{A}\mathbf{x}\|^\kappa\right)}{E\left(\|\mathbf{A}\mathbf{x}\|^\kappa\right)} P^{\mathbf{x}_i}(\mathrm{d}\mathbf{x}) \neq \frac{E\left(f\left(\frac{\mathbf{A}\mathbf{x}_i}{\|\mathbf{A}\mathbf{x}_i\|}\right)\|\mathbf{A}\mathbf{x}_i\|^\kappa\right)}{E\left(\|\mathbf{A}\mathbf{x}_i\|^\kappa\right)}.$$

Therefore, if $\mathbf{x}_i$ has distribution ν_κ, $\kappa \in (0, \kappa_0]$, the right hand side equals $E(f(\mathbf{x}_i))$, but the left hand side does not equal this expression.

In the next section, we propose an alternative algorithm which makes use of Proposition 4.3.3 in a different way.

4.4 Simulation algorithm

For $\kappa \in (0, \kappa_0]$, consider the slightly modified algorithm

```
repeat
{
sample A from the law of A_1,
sample θ_κ from the law ν_κ
set Y = Aθ_κ/‖Aθ_κ‖,
with probability ‖Aθ_κ‖^κ/C^κ accept Y as x_i
}
until Y is accepted
```

Then, by Proposition 4.3.3 $\mathbf{x}_i$ also has distribution ν_κ. Unfortunately, in order for this algorithm to work properly we would need a way to sample independent copies of $\boldsymbol{\theta}_\kappa$ from ν_κ which is exactly what we need the algorithm for. We propose the following workaround.

The idea is again based on the proof of Theorem 3 in [28]. Let certain operators T_κ on the space of probability measures on $\mathbb{S}_+^{d-1}$ be defined with the help of $\mathbf{A}_1$ and $\kappa \in (0, \kappa_0]$. Therefore, for a given probability measure ν on $\mathbb{S}_+^{d-1}$ let $T_\kappa(\nu)$ be the probability measure on $\mathbb{S}_+^{d-1}$ which satisfies

$$\int_{\mathbb{S}_+^{d-1}} f(\mathbf{x}) T_\kappa(\nu)(\mathrm{d}\mathbf{x}) = \frac{\int_{\mathbb{S}_+^{d-1}} E\left(f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|}\right)\|\mathbf{A}_1\mathbf{x}\|^\kappa\right)\nu(\mathrm{d}\mathbf{x})}{\int_{\mathbb{S}_+^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu(\mathrm{d}\mathbf{x})} \tag{4.4.1}$$

for all continuous functions f on $\mathbb{S}_+^{d-1}$. In [28], similar operators are used (cf. the operators T_κ^* in step 1 of Kesten's proof of his Theorem 3), but without the normalization which we have introduced by dividing by $\int_{\mathbb{S}_+^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu(\mathrm{d}\mathbf{x})$. Therefore, in [28] the operators are defined on the whole space of signed measures on $\mathbb{S}_+^{d-1}$ while we will restrict the operator to the space of probability measures on $\mathbb{S}_+^{d-1}$.

Proposition 4.4.1 *Let the conditions of Proposition 4.3.1 be satisfied and let T_κ for $\kappa \in (0, \kappa_0]$ be defined as above. Then, there exists at least one fixed point of T_κ and every fixed point satisfies Equation (4.3.1) with a suitable constant ρ_κ.*

Proof. Let f be a continuous function on $\mathbb{S}_+^{d-1}$ and let ν be a fixed point of T_κ. Now,

$$\begin{aligned}
& \int_{\mathbb{S}_+^{d-1}} E\left(f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|}\right)\|\mathbf{A}_1\mathbf{x}\|^\kappa\right)\nu(\mathrm{d}\mathbf{x}) \\
\overset{(4.4.1)}{=} & \int_{\mathbb{S}_+^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu(\mathrm{d}\mathbf{x}) \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})T_\kappa(\nu)(\mathrm{d}\mathbf{x}) \\
= & \int_{\mathbb{S}_+^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu(\mathrm{d}\mathbf{x}) \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})\nu(\mathrm{d}\mathbf{x}) \\
= & \rho_\kappa \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})\nu(\mathrm{d}\mathbf{x})
\end{aligned}$$

with $\rho_\kappa = \int_{\mathbb{S}_+^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu_\kappa(\mathrm{d}\mathbf{x})$. Since ρ_κ is independent of f, Equation (4.3.1) follows. Because the assumptions of Proposition 4.3.1 are satisfied there exists at least one probability law ν_κ which satisfies Equation (4.3.1). Calculations similar to the ones above then show that ν_κ is a fixed point of T_κ. □

However, it is not stated in [28] nor can it be shown by simple methods that the recursive application of the operator T_κ converges to a fixed point for an arbitrary starting measure ν on $\mathbb{S}_+^{d-1}$. Nevertheless, *if* the recursion converges *then* the fixed point ν'_κ fulfills (4.3.1). Thus, an iterated application of the operator T_κ can be used as the basic idea for an algorithm which allows us to simulate from ν_κ.

Therefore, we consider the following modification of the above algorithm:

```
Start with a sample (x_1, ..., x_n)
repeat {
repeat {
sample A from the law of A_1,
draw x randomly from the sample (x_1, ..., x_n)
set Y = Ax/||Ax||,
with probability ||Ax||^κ / C^κ accept Y as x'_i
}
until Y is accepted
}
until we have generated a new sample (x'_1, ..., x'_n)
replace (x_1, ..., x_n) by (x'_1, ..., x'_n)
```

Here, instead of sampling from a certain distribution, the algorithm draws randomly from a given population $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, which corresponds to the sampling from an empirical distribution function. Note that for $n \to \infty$ the empirical distribution

function of an i.i.d. sample converges to the underlying distribution function by the Glivenko-Cantelli theorem. We then generate $\mathbf{x}'_1, \mathbf{x}'_2, \dots$ according to the rejection sampling algorithm stated at the beginning of this section with the difference that we do not sample from ν_κ but from the empirical distribution $\hat{\nu}_{(\mathbf{x}_1,\dots,\mathbf{x}_n)}$ of the sample $(\mathbf{x}_1, \dots, \mathbf{x}_n)$. A modification of the proof of Proposition 4.3.3 shows that the sample $(\mathbf{x}'_1, \dots, \mathbf{x}'_n)$ generated in this way has the same distribution as an independent sample from the law $T_\kappa(\hat{\nu}_{(\mathbf{x}_1,\dots,\mathbf{x}_n)})$. Again by Glivenko-Cantelli, for large values of n the empirical distribution $\hat{\nu}_{(\mathbf{x}'_1,\dots,\mathbf{x}'_n)}$ of the new sample approximately equals the distribution $T_\kappa(\hat{\nu}_{(\mathbf{x}_1,\dots,\mathbf{x}_n)})$.

Now, in the next step we repeat the algorithm and draw from $(\mathbf{x}'_1, \dots, \mathbf{x}'_n)$ to generate a sample $(\mathbf{x}''_1, \dots, \mathbf{x}''_n)$. Approximately, the empirical distribution of this sample equals $T_\kappa(\hat{\nu}_{(\mathbf{x}'_1,\dots,\mathbf{x}'_n)})$. We proceed this way to iterate the operator T_κ on our samples. Doing this with a statistical software like R allows us to keep an eye on the convergence behavior of the algorithm while creating the samples one after another. In the two-dimensional case we may project $\mathbb{S}^1_+ \subset \mathbb{R}^2$ to $[0, \pi/2]$ and describe the samples by their one-dimensional empirical distribution functions. It turns out that for the cases which we have studied (see below) the empirical distribution functions show convergence behavior, as measured in the test-statistic of the Kolmogorov-Smirnov test for two successive samples. This implies that our algorithm succeeds in finding a fixed point of T_κ, although further research needs to be done in order to strengthen its theoretical fundament.

Now, putting the pieces together, a general procedure for determining κ^* for a given RDE is the following. For different values of κ we evaluate $\hat{\rho}_\kappa$ from (4.3.3) in order to find the unique κ^* for which $\hat{\rho}_{\kappa^*} \approx 1$. We simulate vectors $\mathbf{x}_1, \mathbf{x}_2, \dots$ from ν_κ (see above) and are left to evaluate $E(\|\mathbf{A}_1\mathbf{x}_i\|^\kappa)$ for $i = 1, 2, \dots$ for a Monte Carlo estimation of

$$\int_{\mathbb{S}^{d-1}} E(\|\mathbf{A}_1\mathbf{x}\|^\kappa)\nu_\kappa(d\mathbf{x}).$$

Depending on both the dimension of the matrix $\mathbf{A}_1$ and its underlying distribution, this can be done either analytically (probably with the help of numerical integration) or again by doing Monte Carlo simulations for each realization of $\mathbf{x}_i$.

We close this section with a remark on the assumption of the finite upper bound for the operator norm of $\mathbf{A}_1$ which is needed for our algorithm. Of course, there are many applications in which an upper bound of the operator norm of $\mathbf{A}_1$ does not exist. For example, consider the matrix given in (4.2.11) when the $\epsilon_t, t \in \mathbb{Z}$, are normally distributed, which is a frequent assumption for this model. An ad hoc solution to this problem would be to use a truncated version $\mathbf{A}_1^{(C)}$ instead of $\mathbf{A}_1$ for the simulations, with

$$\mathcal{L}(\mathbf{A}_1^{(C)}) = \mathcal{L}\left(\mathbf{A}_1 \middle| \|\mathbf{A}_1\|_{\text{op}} \le C\right). \tag{4.4.2}$$

The following lemma shows that this approach is feasible.

Lemma 4.4.2 *Let the assumptions of Proposition 4.3.3 be satisfied and for a sequence*

$C_n \to \infty, n \in \mathbb{N}$, let $\mathbf{A}_1^{(C_n)}$ be random matrices related to $\mathbf{A}_1$ and C_n by (4.4.2) such that the $\mathbf{A}_1^{(C_n)}$ also satisfy the assumptions of Theorem 4.3.3. For $\kappa \in (0, \kappa_0]$ denote the value of ρ_κ associated with $\mathbf{A}_1^{(C_n)}$ by $\rho_\kappa^{(C_n)}$. Then

$$\rho_\kappa^{(C_n)} \to \rho_\kappa$$

for $n \to \infty$ and all $\kappa \in (0, \kappa_0]$.

Proof. Note that

$$E\left(\|\mathbf{A}_1^{(C_n)}\|_{\text{op}}^{\kappa_0} \ln^+ \|\mathbf{A}_1^{(C_n)}\|_{\text{op}}\right) < \infty, \quad \forall\, n \in \mathbb{N},$$

and that

$$E\left(\min_{i=1,\dots,d} \sum_{j=1}^{d} A_{ij}^{(C_n)}\right)^{\kappa_0} \to E\left(\min_{i=1,\dots,d} \sum_{j=1}^{d} A_{ij}\right)^{\kappa_0}, \quad n \to \infty,$$

as well as

$$E\left(\min_{j=1,\dots,d} \sum_{i=1}^{d} A_{ij}^{(C_n)}\right)^{\kappa_0} \to E\left(\min_{j=1,\dots,d} \sum_{i=1}^{d} A_{ij}\right)^{\kappa_0}, \quad n \to \infty,$$

therefore the value of κ_0 associated with $\mathbf{A}_1^{C_n}$ can be chosen as $\kappa_0^{(C_n)} \in [\kappa_0, \kappa_0 + \epsilon]$ for $\epsilon > 0$ and n large enough. Now, for every $\kappa \in (0, \kappa_0]$ there exists a sequence of probability measures $\nu_\kappa^{(C_n)}, n \in \mathbb{N}$, on $\mathbb{S}_+^{d-1}$ and a sequence of constants $\rho_\kappa^{(C_n)}, n \in \mathbb{N}$, such that

$$\int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1^{(C_n)}\mathbf{x}\|^\kappa f\left(\frac{\mathbf{A}_1^{(C_n)}\mathbf{x}}{\|\mathbf{A}_1^{(C_n)}\mathbf{x}\|}\right)\right] \nu_\kappa^{(C_n)}(\mathrm{d}\mathbf{x}) = \rho_\kappa^{(C_n)} \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x}) \nu_\kappa^{(C_n)}(\mathrm{d}\mathbf{x}) \tag{4.4.3}$$

holds for all continuous functions f on $\mathbb{S}_+^{d-1}$ and all $n \in \mathbb{N}$. In the following, let $f \equiv 1$. Because $\mathbb{S}_+^{d-1}$ is a compact space, every sequence of probability measures on it has a convergent subsequence. Let now $n_k, k \in \mathbb{N}$, and ν_κ^* such that $\nu_\kappa^{(C_{n_k})} \xrightarrow{w} \nu_\kappa^*$ for $k \to \infty$. We will apply the continuous mapping theorem (cf. Theorem 4.27, [27]) to show that the expression

$$\left| \int_{\mathbb{S}_+^{d-1}} E\left(\|\mathbf{A}_1^{(C_{n_k})}\mathbf{x}\|^\kappa\right) \nu_\kappa^{(C_{n_k})}(\mathrm{d}\mathbf{x}) - \int_{\mathbb{S}_+^{d-1}} E\left(\|\mathbf{A}_1\mathbf{x}\|^\kappa\right) \nu_\kappa^*(\mathrm{d}\mathbf{x}) \right| \tag{4.4.4}$$

converges to 0. Therefore, we will derive first that

$$E\|\mathbf{A}_1^{(C_{n_k})}\mathbf{x}_k\|^\kappa \to E\|\mathbf{A}_1\mathbf{x}\|^\kappa \tag{4.4.5}$$

as $k \to \infty$ for all sequences $\mathbf{x}_k \to \mathbf{x} \in \mathbb{S}_+^{d-1}$. This is because

$$\left|E\|\mathbf{A}_1^{(C_{n_k})}\mathbf{x}_k\|^\kappa - E\|\mathbf{A}_1\mathbf{x}\|^\kappa\right| \leq \int_{\mathbb{S}_+^{d-1}} \left| \frac{\mathbb{1}_{\{\|\mathbf{a}\|_{\text{op}} \leq C_{n_k}\}} \|\mathbf{a}\mathbf{x}_k\|^\kappa}{P(\|\mathbf{A}_1\|_{\text{op}} \leq C_{n_k})} - \|\mathbf{a}\mathbf{x}\|^\kappa \right| P^{\mathbf{A}_1}(\mathrm{d}\mathbf{a}).$$

Now, with $M = P(\|\mathbf{A}_1\|_{\text{op}} \leq C_{n_k})^{-1} + 1$ the integrand is bounded by $M\|\mathbf{a}\|_{\text{op}}^{\kappa}$ and for k large enough we may for example set $M \leq 3$. Since $E\|\mathbf{A}_1\|_{\text{op}}^{\kappa} < \infty$ (cf. Equation (4.2.2)), dominated convergence allows us to interchange limit and integral and therefore (4.4.5) holds. Now, Equation (4.4.5) allows us to apply the continuous mapping theorem to (4.4.4) which shows that the expression converges to 0 as $k \to \infty$. With (4.4.3) we then get that

$$\left| \rho_{\kappa}^{(C_{n_k})} - \int_{\mathbb{S}_+^{d-1}} E\left(\|\mathbf{A}_1\mathbf{x}\|^{\kappa}\right) \nu_{\kappa}^*(\mathrm{d}\mathbf{x}) \right| \to 0, \quad k \to \infty,$$

and therefore $\rho_{\kappa}^{(C_{n_k})}$ converges to a constant ρ_{κ}^*. But then

$$\begin{aligned}
\rho_{\kappa}^* \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x}) \nu_{\kappa}^*(\mathrm{d}\mathbf{x}) &= \lim_{k\to\infty} \rho_{\kappa}^{(C_{n_k})} \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x}) \nu_{\kappa}^{(C_{n_k})}(\mathrm{d}\mathbf{x}) \\
&= \lim_{k\to\infty} \int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1^{(C_{n_k})}\mathbf{x}\|^{\kappa} f\left(\frac{\mathbf{A}_1^{(C_{n_k})}\mathbf{x}}{\|\mathbf{A}_1^{(C_{n_k})}\mathbf{x}\|} \right) \right] \nu_{\kappa}^{(C_{n_k})}(\mathrm{d}\mathbf{x}) \\
&= \int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1\mathbf{x}\|^{\kappa} f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|} \right) \right] \nu_{\kappa}^*(\mathrm{d}\mathbf{x})
\end{aligned}$$

holds for all continuous functions f on $\mathbb{S}_+^{d-1}$ again by the continuous mapping theorem with arguments similar to the ones used above where we may replace M by $M\|f\|_{\infty}$. Therefore $\rho_{\kappa}^* = \rho_{\kappa}$ regardless of the chosen subsequence, that means $\rho_{\kappa}^{(C_n)} \to \rho_{\kappa}$ for $n \to \infty$. □

4.5 Simulation results

We will finish this chapter with an example for the evaluation of κ for a GARCH$(1,2)$ model with parameters taken from [9], where different GARCH(p,q) models (with parameters (p,q) equal to $(1,1),(1,2),(2,1)$ and $(2,2)$, respectively) are fitted to DM-U.S. Dollar exchange rate data. We will analyze the index of regular variation for the estimated GARCH$(1,1)$ and GARCH$(1,2)$ model. For the first model with parameters

$$\alpha_0 = 0.0069, \quad \alpha_1 = 0.0449, \quad \beta_1 = 0.9414$$

and standard normal innovations we may evaluate $\kappa_{1,1}^*$ by numerical integration from the Equation

$$E\left[\left(\alpha_1 \epsilon_0^2 + \beta_1 \right)^{\kappa_{1,1}^*} \right] = 1$$

(cf. (4.2.13)) as

$$\kappa_{1,1}^* \approx 6.2,$$

that means the associated stationary distributions of σ_t and $|X_t|$ are regularly varying with index $2\kappa^*_{1,1} \approx 12.4$. For the GARCH$(1,2)$ case, the model parameters are given by

$$\alpha_0 = 0.0049, \quad \alpha_1 = 0.0313, \quad \beta_1 = 1.2819, \quad \beta_2 = -0.3229$$

with standard normal innovations. Note that one of the parameters is negative which is in line with a well-known extension of the model (cf. [33]), where our parameters satisfy equations (23)-(27) in [33] (note that due to a different notation, the model is named GARCH$(2,1)$ in that paper, instead of GARCH$(1,2)$ in our notation). This guarantees that the volatility process as defined by (4.2.9) stays positive if we choose the initial values σ_0^2, σ_1^2 from a certain range.

In our case, we can reduce the RDE notation for GARCH(p,q) models from Section 4.2 to the analysis of the time series $(\sigma_t^2, \sigma_{t-1}^2)_{t\in\mathbb{N}}$ which can be written as a two-dimensional RDE with

$$\begin{pmatrix} \sigma_{t+1}^2 \\ \sigma_t^2 \end{pmatrix} = \begin{pmatrix} 1.2819 + 0.0313\epsilon_{t+1}^2 & -0.3229 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \end{pmatrix} + \begin{pmatrix} 0.0049 \\ 0 \end{pmatrix}, \tag{4.5.1}$$

for $t \in \mathbb{N}$ with i.i.d. standard normal $\epsilon_t, t \in \mathbb{N}$. The matrix $\mathbf{A}_1$ in this model has negative entries but one can show that if $(\sigma_1^2, \sigma_0^2) \in S := \{\mathbf{x} \in \mathbb{R}_+^2 | \mathbf{x} = (\cos(\theta), \sin(\theta)), \theta \in [0, s]\} \subset \mathbb{S}_+^1$ with $s \approx 1.2391$ then also $\mathbf{A_n} \cdot \ldots \cdot \mathbf{A}_1(\sigma_1^2, \sigma_0^2)^t \in S$ almost surely for all $n \in \mathbb{N}$. This suffices to apply the results which have been stated for non-negative matrices, cf. Remark 1 in [28]. In order to show that the assumptions of Theorem 4.2.1 and Corollary 4.2.3 are satisfied, we evaluate the Lyapunov exponent γ by Monte Carlo simulations and get

$$\gamma \overset{a.s.}{=} \lim_{n\to\infty} \frac{1}{n} \log(\|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1 \mathbf{x}\|) \approx -0.048 < 0, \quad \forall \mathbf{x} \in S,$$

and due to the light tails and the continuity of the assumed normal distribution for the innovations $\epsilon_t, t \in \mathbb{N}$, we can easily check that the conditions of Theorem 4.2.1 and Corollary 4.2.3 are satisfied for arbitrarily large κ_0 (with exception of (4.2.1) and (4.2.6), but cf. Remark 4.3.2 (b)). For our simulations we have used a truncated version of the matrix by truncating the absolute value the innovations $\epsilon_t, t \in \mathbb{N}$, at the 99.999%-quantile of the standard normal distribution. To derive the corresponding upper bound C for $\|\mathbf{A}_1\|_{\text{op}}$ we have used that $\|\mathbf{A}_1\|_{\text{op}} = \sup_{\mathbf{x}\in S} \|\mathbf{A}_1\mathbf{x}\| = \|\mathbf{A}_1(1,0)^t\|$.

Our simulations for the GARCH$(1,2)$ model give us the estimate

$$\hat{\kappa}^*_{1,2} \approx 6.1,$$

cf. the estimated curve of ρ_κ for the GARCH$(1,2)$ model (Figure 4.1) below.

Compared to the value of $\kappa^*_{1,1} \approx 6.2$ for the GARCH$(1,1)$ model both models seem to reflect a similar structure of the extremal behavior of the time series, at least regarding the index of regular variation.

In order to visualize the function $\rho(\kappa)$ for the GARCH$(1,2)$ model we have evaluated ρ_κ one after another for a equidistantly spaced sequence of κ's in Figure 4.1. If the aim

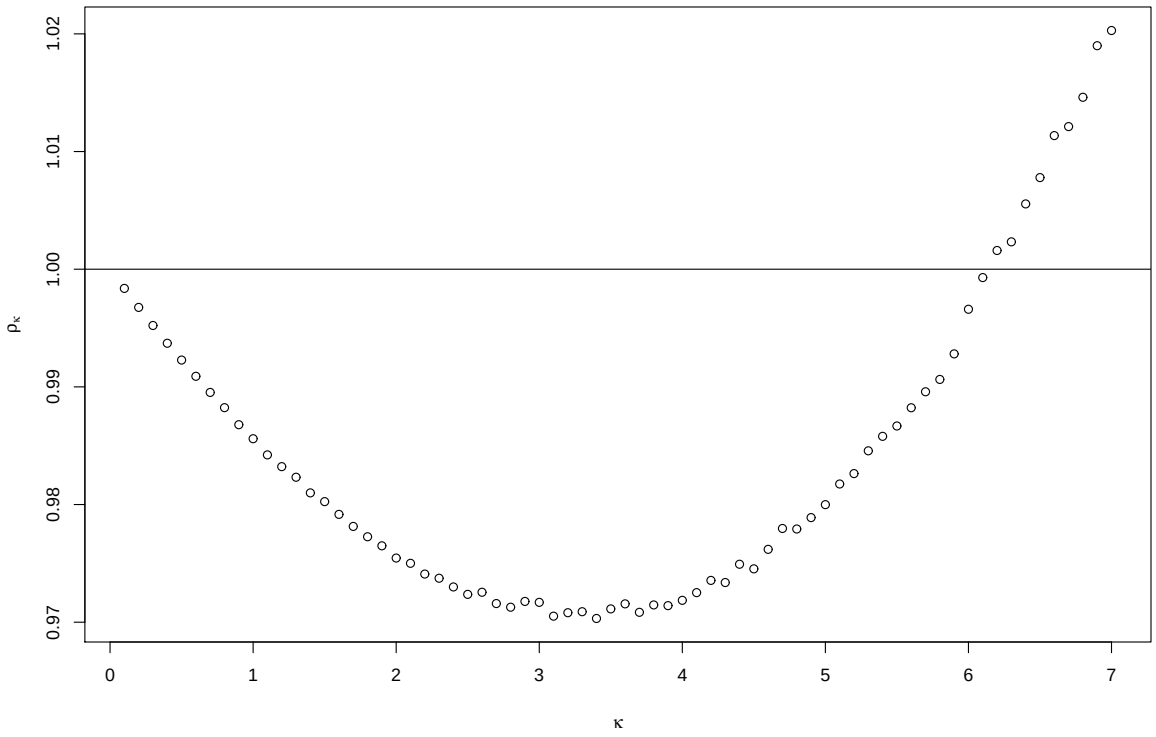


Figure 4.1: Estimated values of ρ_κ for the GARCH$(1,2)$ model in [32].

is to find κ^* in a more efficient way, the next value of κ for which ρ_κ is to be evaluated can be chosen subject to the previously found values of ρ_κ.

The example code for the evaluation of ρ_κ for $\kappa = 6$, a sample size of $n = 100.000$ and 10 iterations of the sampling procedure for ν_κ is shown below.

Listing 4.1: Example code

```
# Numerical evaluation of rho_kappa for kappa=6
# and GARCH(1,2) model
alpha1<-0.0313
beta1<-1.2819
beta2<- -0.3229
quantile<-0.99999 # Quantile used for truncation

n<-100000 # sample size
Ceps<-qnorm(quantile) # Truncation point for the innovations
Cmatrix<-sqrt((beta1+alpha1*Ceps^2)^2+1)
# Resulting bound for the operator norm
```

```
kappa<-6
# Initialize sample, uniformly distributed on S
x<-seq(0,1.2391,length.out=n)
# Ten iterations of the algorithm
for (j in 1:10)
{
  xnew<-rep(0,n)
  i<-0
  # Rejection sampling
  while(i<n)
   {
    theta<-sample(x,1,replace=FALSE)
    epsilon<-rnorm(1)
    while(abs(epsilon)>Ceps){epsilon<-rnorm(1)}
    Y<-matrix(c(beta1+alpha1*epsilon^2,1,beta2,0),2)
       %*%c(cos(theta),sin(theta))
    if (runif(1)<((Y[1]^2+Y[2]^2)^(kappa/2))/Cmatrix^kappa)
     {
      i<-i+1;xnew[i]<-atan(Y[2]/Y[1])
     }
   }
  # Print KS-Test to check convergence
  print(ks.test(x,xnew))
  # renew sample
  x<-xnew
}
# Evaluate E|Ax_i|^kappa for each x_i
# Use numerical integration
final<-rep(0,n)
for (i in 1:n)
{
  Yvector<-c(cos(x[i]),sin(x[i]))
  integrand<-function(x){(sqrt(((beta1+alpha1*x^2)*Yvector[1]+
  beta2*Yvector[2])^2+Yvector[1]^2))^kappa*dnorm(x)/
  (1-2*(1-quantile))}
  final[i]<- integrate(integrand,lower=-Ceps,upper=Ceps)$value
}
# Mean of this values is estimate for rho_kappa
print(mean(final))
```

Figure 4.2 shows the histogram of the sample $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ after ten iterations of the algorithm for $\kappa = 6$. The value of $\mathbf{x} = (x_1, x_2) \in \mathbb{S}^1_+$ is projected on $\theta = \arctan(x_2/x_1) \in [0, \pi/2]$.

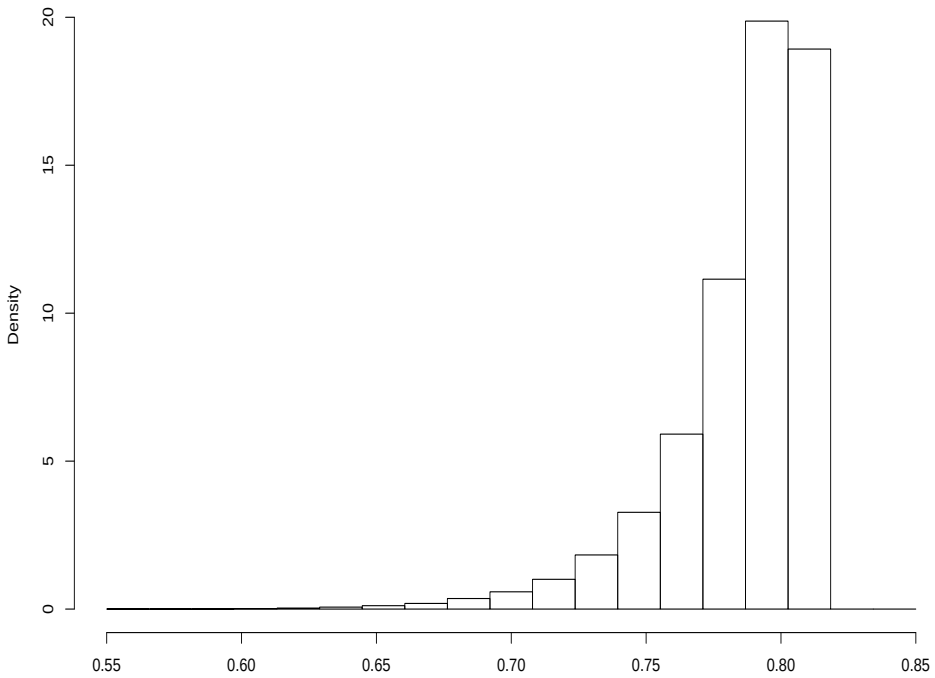


Figure 4.2: Histogram of the sample after ten iterations of the algorithm for $\kappa = 6$.

The efficiency of this algorithm depends heavily on the values of both κ and C, the bound for the operator norm of $\mathbf{A}_1$. Since the rejection algorithm is fastest for low values of C^κ, it works best for relatively low values of κ and C (keep in mind that $C > 1$, since otherwise $\rho_\kappa < 1$ for all κ). For the GARCH$(2,1)$ model estimated in [32] the value of C for a reasonably truncated normal distribution of the innovations $\epsilon_t, t \in \mathbb{Z}$, does not allow for the evaluation of ρ_κ in the region of interest. For the GARCH$(1,2)$ model analyzed above, the successive evaluation of ρ_κ of $\kappa = 0.1, 0.2, \ldots, 6.9, 7$ with result visualized in Figure 4.1 takes about two days and the analysis of GARCH$(1,2)$ models for financial time series should in general be feasible with our new method. Our method can be used to explore whether different models which are fitted to the same data have similar extremal characteristics or to analyze the influence of the parameters of a certain model (for example the α_i, β_j in the GARCH(p,q) model but also the distribution of the $\epsilon_t, t \in \mathbb{Z}$) on its extremal properties.

The idea of simulating with this new method has a lot in common with the idea of simulating from the tail chain which was used in Chapter 3. In both cases we are able to simulate directly from the tail of the distribution of interest, either by sampling from the tail chain or from the spectral measure. Compared to the alternative of first

simulating realizations of the time series (for example of a GARCH(p, q) process) and then using only the extremal observations of it, these new methods are much more efficient and stable. Furthermore, they illuminate how the extremal behavior of a time series is influenced by its often light-tailed innovations which are the main ingredients for our simulations.

Bibliography

[1] Andersen, T. G., Davis, R. A., Kreiss, J.-P. and Mikosch, T., editors. *Handbook of Financial Time Series*. Springer, New York, 2009

[2] Basrak, B., Davis, R. A. and Mikosch, T. Regular variation of GARCH processes. *Stoch. Proc. Appl.*, **99** 95–115, 2002

[3] Basrak, B. and Segers, J. Regularly varying multivariate time series. *Stoch. Proc. Appl.*, **119** 1055–1080, 2009

[4] Ben Arous, G., Bogachev, L. V. and Molchanov, S. A. Limit theorems for sums of random exponentials. *Probab. Theory Related Fields*, **132** 579–612, 2005

[5] Billingsley, P. *Probability and Measure*. Wiley, New York, 1995

[6] Bingham, N. H., Goldie, C. M. and Teugels, J. L. *Regular Variation*. University Press, Cambridge, 1987

[7] Bogachev, L. Limit laws for norms of IID samples with Weibull tails. *J. Theoret. Probab.*, **19** 849–873, 2006

[8] Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3) 307–327, 1986

[9] Bollerslev, T. and Wright, J. H. High-frequency data, frequency domain inference, and volatility forecasting. *Rev. Econ. Statist.*, **83**(4) 596–602, 2001

[10] Boman, J. and Lindskog, F. Support theorems for the Radon transform and Cramér-Wold theorems. *J. Theoret. Probab.*, **22**(3) 683–710, 2009

[11] Bovier, A., Kurkova, I. and Löwe, M. Fluctuations of the free energy in the REM and the p-spin SK-model. *Ann. Probab.*, **30** 605–651, 2002

[12] Breiman, L. On some limit theorems similar to the arc-sin law. *Theory Probab. Appl.*, **10** 323–331, 1965

[13] Darling, D. A. The influence of the maximum term in the addition of independent random variables. *Trans. Amer. Math. Soc.*, **73** 95–107, 1952

[14] de Haan, L., Resnick, S., Rootzén, H. and de Vries, C. G. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stoch. Proc. Appl.*, **32** 213–224, 1989

[15] Ehlert, A. *Characteristics for Dependence in Time Series of Extreme Values.* Ph.D. thesis, University of Göttingen, 2010

[16] El'sgol'ts, L. E. and Norkin, S. B. *Introduction to the Theory and Application of Differential Equations with Deviating Arguments.* Academic Press, New York, 1973

[17] Embrechts, P., Klüppelberg, C. and Mikosch, T. *Modelling Extremal Events.* Springer, Berlin, 1997

[18] Fasen, V., Klüppelberg, C. and Schlather, S. High-level dependence in time series models. *Extremes*, **13** 1–33, 2010

[19] Feller, W. Über den zentralen Grenzwertsatz in der Wahrscheinlichkeitsrechnung. *Math. Z.*, **40** 521–559, 1936

[20] Feller, W. *An Introduction to Probability Theory and its Applications*, volume II. Wiley, New York, 1971

[21] Gnedenko, B. V. and Kolmogorov, A. N. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, MA, 1968

[22] Goldie, M., C. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Prob.*, **1** 126–166, 1991

[23] Gradshteyn, I. S. and Ryzhik, I. M. *Table of Integrals, Series, and Products, 6th ed.* Academic Press, San Diego, 2000

[24] Grimmett, G. R. and Stirzaker, D. R. *Probability and Random Processes.* Oxford University Press, Oxford, 2001

[25] Hult, H. and Lindskog, F. Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Prob.*, **34** 587–608, 2002

[26] Janssen, A. Limit laws for power sums and norms of i.i.d. samples. *Probab. Theory Related Fields*, **146** 515–533, 2010

[27] Kallenberg, O. *Foundations of Modern Probability.* Springer, New York, 2002

[28] Kesten, H. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, **131** 207–248, 1973

[29] Loève, M. *Probability Theory.* Van Nostrand, Toronto, New York, London, 1955

[30] Maulik, K. and Resnick, S. I. Characterizations and examples of hidden regular variation. *Extremes*, **7** 31–67, 2004

[31] Mikosch, T. and Stărică, C. Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *Ann. Statist.*, **28** 1427–1451, 2000

[32] Nelson, D. B. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6** 318–334, 1990

[33] Nelson, D. B. and Cao, C. Q. Inequality constraints in the univariate GARCH model. *J. Bus. Econ. Statist.*, **10**(2) 229–235, 1992

[34] Perfekt, R. Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Prob.*, **4** 529–548, 1994

[35] Perfekt, R. Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Adv. Appl. Probab.*, **29** 138–164, 1997

[36] Resnick, S. I. *Heavy-Tail Phenomena.* Springer, New York, 2007

[37] Resnick, S. I. *Extreme Values, Regular Variation and Point Processes.* Springer, New York, 2008

[38] Schlather, M. Limit distributions of norms of vectors of positive i.i.d. random variables. *Ann. Probab.*, **29** 862–881, 2001

[39] Segers, J. Multivariate regular variation of heavy-tailed Markov chains. Technical report, Institut de statistique, available on arxiv.org as math.PR/0701411, 2007

[40] Smith, R. L. The extremal index for a Markov chain. *J. Appl. Prob.*, **29** 37–45, 1992

[41] Taylor, S. *Modelling Financial Time Series.* Wiley, Chichester, 1986

[42] Vervaat, W. On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Prob.*, **11** 750–783, 1979

[43] Wuertz, D., with contribution from Michal Miklovic, Y. C., Boudt, C., Chausse, P. et al. *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*, 2009. URL http://CRAN.R-project.org/package=fGarch. R package version 2110.80