

Nonparametric Modeling of Extremal Dependence

Michał Warchol

A thesis submitted in partial fulfillment of the requirements for
the degree of Docteur en Sciences

Université catholique de Louvain
Institute of Statistics, Biostatistics and Actuarial Sciences

August 2016

Thesis committee:

Johan Segers, Supervisor (Université catholique de Louvain)
Pierre Devolder, President (Université catholique de Louvain)
Richard A. Davis (Columbia University)
Holger Drees (University of Hamburg)
Jan Johannes (Heidelberg University)
Rainer von Sachs (Université catholique de Louvain)
David Veredas (Vlerick Business School)

Acknowledgements

Setting goals in life is important as it guides us on the journey. However, If we focus solely on the destination then we might miss other experiences on the way. Initially, this dissertation was a goal but it turned out to be a by-product of my scientific journey. During the last four years I had the privilege to meet and work with amazing people. Having them as mentors, friends and companions has become more valuable to me than any publication or degree. I am grateful for their support and for giving me the opportunity to learn from their experience.

I am grateful to Johan Segers for the trust he has put in me, for giving me valuable directions when I needed them and letting me follow my own path when the time was right, for keeping me constantly amazed how complex concepts can be visualized and put in simple words.

Thanks to Richard A. Davis for his hospitality and many inspiring meetings during my visit at Columbia University. With his great sense of humor he shows not to take things too seriously.

My sincere thanks goes to Holger Drees who, by his precision and attention to details, helped me to become a better researcher. I appreciate many interesting conversations we had, especially the ones during my visit in Hamburg.

Thanks to the members of my comité d'accompagnement, Rainer von Sachs, Jan Johannes, David Veredas for giving me helpful comments and suggestions and to Pierre Devolder for chairing the committee. I appreciate the support of Anthony C. Davison and Miguel de Carvalho who inspired me and encouraged to do research when I was working on my master thesis at Ecole Polytechnique Fédérale de Lausanne. I am thankful to Nancy Guillaume for her kindness and for helping me out with all the administrative issues.

I gratefully acknowledge “Fonds de la Recherche Scientifique - FNRS” for funding my work through “mandat d’aspirant”. Thanks to Université catholique de Louvain for the travel grant within “Fonds Spécial Recherche” which financially supported my New York adventure.

Thanks to Anna Kiriliouk for a being a great companion along the phd journey. Thanks to Łukasz Kidziski for showing me that “done is better than perfect” and that “exposure is better than knowledge”. Many thanks goes to Paweł Polak with whom I have shared a cozy office at Columbia University. Paweł, like no other, challenged the usefulness of my research which resulted in testing many different trading strategies. Thanks to Aleksandar Sujica for showing me the importance of commitment and putting one’s skin in the game. I am grateful to Phyllis Wan for being a great host during my research visit at Columbia University. I am thankful to Ola and Michał for their support while living abroad and for being there for me in case of any emergency.

I am grateful to my friends at ISBA: Joris, Josefine, Maïlis, Mickaël, Nicolas, Aurélie, Nathan, Adrien, Benjamin, Gildas, François, Sylvie, Yuwei, Marco, Vincent, and other colleagues for

many inspiring discussions. Thanks to Holsuk and Seong for giving me tough lessons in badminton on the rainy Belgian days.

Having an older brother makes life much easier. Usually the older brother needs to figure things out first and the younger brother can learn from his mistakes and successes. I am grateful to my brother to whom I can always turn to when I need help. Last but not least, I thank my parents for their unconditional love and support. Thank you for being with me through all the ups and downs.

Contents

Acknowledgements	iii
1 Introduction	1
1.1 Basic motivation	1
1.2 Regular variation	4
1.3 Extremal dependence in time	8
1.4 Objectives and structure of the thesis	14
2 Statistics for Tail Processes of Markov Chains	19
2.1 Introduction	20
2.2 Tail processes of Markov chains	21
2.3 Estimating Markov spectral tail processes	25
2.4 Large sample theory	28
2.4.1 Known index of regular variation	29
2.4.2 Unknown index of regular variation	31
2.5 Examples	33
2.5.1 Copula-based Markov processes	33
2.5.2 Stochastic recurrence equations	35
2.6 Monte Carlo simulations	36
2.7 Case study	39
2.8 Appendix A - Proofs and examples	42
2.8.1 Examples	42
2.8.2 Proofs	44
2.9 Appendix B - Finite levels approximation assessment	54
2.9.1 Approximation assessment for the simulation study	54
2.9.2 Stability plots for the case study	56
3 Modeling Serial Extremal Dependence	63
3.1 Introduction	64
3.2 Methodology	67
3.2.1 Estimators	67
3.2.2 Resampling	69
3.2.3 Testing for dependence of extreme observations	70
3.3 Large-sample theory	70
3.3.1 Asymptotic normality of the estimators	71
3.3.2 Consistency of the multiplier block bootstrap	73

3.4	Finite-sample performance	75
3.4.1	Forward and backward estimators	76
3.4.2	Bootstrapped spectral tail process	77
3.5	Application	79
3.6	Proofs	81
3.7	Appendix A - Additional simulation study	88
4	A Euclidean Likelihood Estimator for Bivariate Tail Dependence	93
4.1	Introduction	94
4.2	Background	95
4.3	Maximum Euclidean Likelihood Estimator	96
4.4	Large-Sample Theory	98
4.5	Monte Carlo Simulations	100
4.6	Extreme Temperature Data Analysis	102
4.6.1	Data Description and Preliminary Considerations	102
4.6.2	Extremal Dependence of Open Air and Forest Cover Temperatures	104
4.7	Appendix A	108
4.7.1	Proofs	108
4.7.2	Beta-Kernel Smoothing of Discrete Spectral Measures	109
5	Discussion	113

Chapter 1

Introduction

Abstract

There is an increasing interest to understand the interplay of extreme values over time and across coordinates. Extreme-value theory provides techniques for modeling temporal and cross-sectional extremal dependence by modeling the marginal distributions and the dependence structure separately. The aim of this chapter is to briefly introduce the probabilistic concepts that are exploited in the following chapters of the thesis. Regular variation is the key assumption in this context. Inference about serial extremal dependence within a time series can be made via the spectral tail process. For estimating dependence at high levels within a multivariate random vector, the spectral measure is particularly convenient. Throughout the theory is illustrated on financial data.

1.1 Basic motivation

The overly simplistic mathematical models used by major financial institutions underestimated the risk of large simultaneous and consecutive losses that occurred during the 2007–2008 financial crisis. By the year 2007, it was common that those stochastic models were based on the normality assumption. Even the Nobel Prize winning model of Black-Scholes-Merton assumes, in particular, that the log-returns of stock prices behave like geometric Brownian motion. For a financial time series consisting of log-returns of some asset, the key features include heavy-tailed marginal distributions and clustering of extremes. Both characteristics violate the assumption of the underlying process being Gaussian.

Not only in finance but also in insurance, telecommunication and other areas the data may contain large values that tend to occur in clusters. In this thesis we are considering time series

models with marginal distribution having infinite moment of a certain order. Clearly, such models are non-Gaussian and are exhibiting heavy tails.

Consider financial time series of three stocks: JP Morgan, Citibank and IBM. We download daily stock prices between January 4, 2000, and December 31, 2013 from <http://finance.yahoo.com>, and convert them to daily log-returns: if $P_0, P_1, \dots, P_n$ is a series of stock prices, the log-returns are

$$X_i = \log(P_i/P_{i-1}), \quad i = 1, \dots, n.$$

These three series of log-returns will be denoted by the vectors $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3})$ for $i = 1, \dots, 3520$.

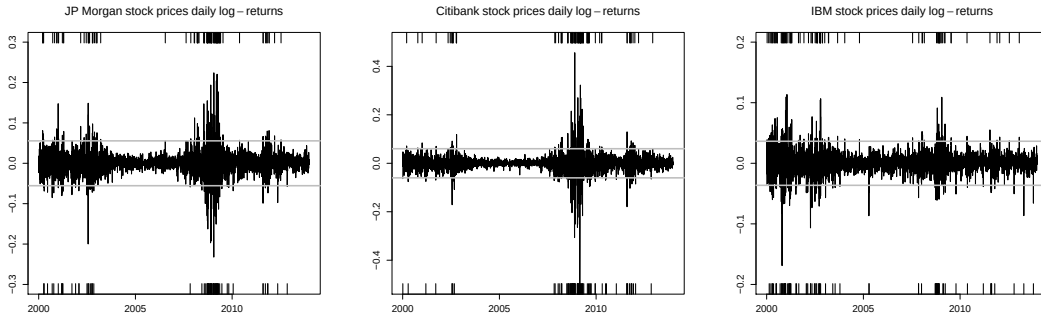


FIGURE 1.1: Daily log-returns of JP Morgan (left), Citibank (middle) and IBM (right) stock prices.

In Figure 1.1 we plot the daily log-returns of the selected stock prices. We treat an observation to be extreme, when its absolute value is above a high threshold. Such thresholds, here corresponding to the 95% percentile of the absolute values, are indicated by the grey horizontal lines. The black ticks at the bottom and at the top mark respectively the days on which a negative or positive extreme observation occurred. We observe that the spikes tend to appear in clusters. We say that the time series exhibit *serial extremal dependence*. For JP Morgan and Citibank this effect is especially visible during the financial crisis of 2007-2008 and in particular around the day of the collapse of Lehman Brothers, i.e., September 15th, 2008. The main focus of this thesis is on modeling of such spikes, their size, sign (gain and losses) and interactions (consecutive occurrence).

Extremes not only can occur consecutively in time but also simultaneously across coordinates. As an example, consider a financial portfolio containing the three stocks. It is of interest, for instance in quantitative risk management, to study simultaneous occurrence of extreme losses.

Figure 1.2 shows the scatterplots of the three possible pairs of negative log-returns on the exponential scale. We restrict the data to the observations that occurred during the financial crisis of 2007-2008. By taking negative log-returns we force (extreme) losses to be in the upper tails of

the distribution functions. To make the comparison between variables meaningful we transform the data to a common scale. Specifically, we transform the negative log-returns to unit-Pareto margins through

$$\hat{X}_{ij}^* := \frac{1}{1 - \hat{F}_j(X_{ij})}, \quad i \in \{1, \dots, n\}, j \in \{1, \dots, d\}, \quad (1.1)$$

(with $d = 3$ and $n = 756$) and plot them on the logarithmic scale; the empirical distribution functions $\hat{F}_j$ evaluated at the data are defined as

$$\hat{F}_j(X_{ij}) := \frac{R_{ij,n} - 1}{n}, \quad i \in \{1, \dots, n\}, j \in \{1, \dots, d\}, \quad (1.2)$$

where $R_{ij,n}$ is the rank of X_{ij} among $X_{1j}, \dots, X_{nj}$, i.e., $R_{ij,n} = \sum_{l=1}^n \mathbf{1}\{X_{lj} \leq X_{ij}\}$.

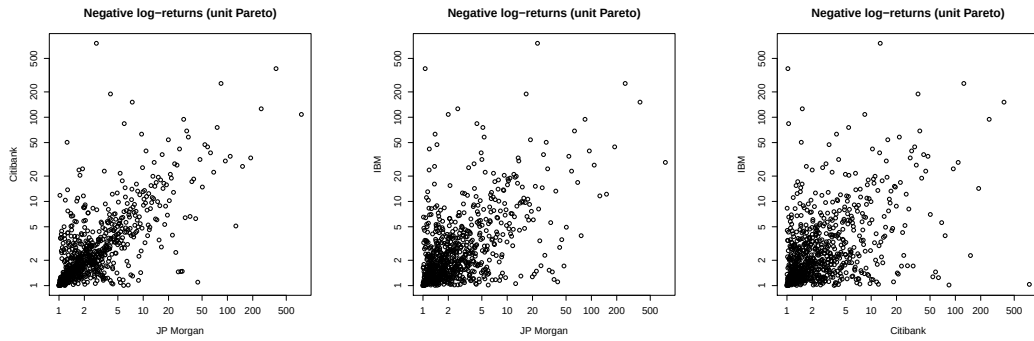


FIGURE 1.2: Scatterplots of daily negative log-returns of JP Morgan, Citibank and IBM stock prices between 03-01-2007 and 31-12-2008, plotted on the exponential scale.

It might not yet be clear, looking at Figure 1.2, for which pair do the extreme losses tend to occur together more often. However, by applying certain techniques, explained later in the thesis (see Figure 1.4 below), it will be evident that joint occurrences of large losses are more frequent for JP Morgan and Citibank (left) than for JP Morgan and IBM (middle) or for Citibank and IBM (right). Given the fact that JP Morgan and Citibank are financial institutions while IBM is an IT company, this is no surprise. In statistical parlance, the pair JP Morgan versus Citibank exhibits the strongest degree of upper *tail dependence*.

A classical way of capturing features present in the data is to focus on the (auto)covariance structure. However, when second moments are infinite, the theory does not apply and the inference drawn upon the sample estimates may be wrong. Moreover, the covariance structure, which mainly refers to the bulk of the distribution, is not informative when it comes to the dependence among extremes. Hence, specialized tools are necessary to properly capture various extremal dependence features. For modeling serial extremal dependence the *spectral tail process* introduced by Basrak and Segers (2009) is of main interest in this thesis. When dealing with

extremal dependence across coordinates we focus on the *spectral measure*. Both concepts are closely related and are based on the assumption of *regular variation* which in particular allows for heavy tails.

The remainder of this introductory chapter is organized as follows. In Section 1.2 we briefly present the fundamental assumption of the regular variation and the emerging spectral measure. Modeling extremes in time is discussed in Section 1.3, with special focus on the tail process. The objectives and structure of the thesis are described in Section 1.4. Throughout, the theory is illustrated on financial data.

Some notational conventions used in the thesis: the law of a random object is denoted by $\mathcal{L}(\cdot)$. Weak convergence is denoted by the arrow $\rightsquigarrow$. The indicator variable of the event E is denoted by $\mathbf{1}(E)$. The set of integers is denoted by $\mathbb{Z}$, while $\mathbb{N} = \{h \in \mathbb{Z} : h \geq 1\}$.

1.2 Regular variation

We say that a d -dimensional random vector $\mathbf{X}$ is *regularly varying* with index $\alpha > 0$ and *spectral measure* $\mathbf{P}[\Theta \in \cdot]$ on the Borel σ -field of the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ if the following weak limit exist for some $\alpha > 0$ and for every $y > 0$:

$$\frac{\mathbf{P}[\|\mathbf{X}\| > yu, \mathbf{X}/\|\mathbf{X}\| \in \cdot]}{\mathbf{P}[\|\mathbf{X}\| > u]} \rightsquigarrow y^{-\alpha} \mathbf{P}[\Theta \in \cdot], \quad u \rightarrow \infty. \quad (1.3)$$

The definition of regular variation does not depend on the particular choice of the norm. It is however relevant for defining the corresponding unit sphere and the spectral measure.

Loosely speaking, (1.3) means that the angular component $\mathbf{X}/\|\mathbf{X}\|$ becomes independent of the radial component $\|\mathbf{X}\|$ as the radius increases. Moreover, the conditional distribution of the rescaled radial component can be approximated by the Pareto(α) law, i.e.,

$$\mathbf{P}[\|\mathbf{X}\| > yu \mid \|\mathbf{X}\| > u] \rightarrow y^{-\alpha}, \quad y \geq 1, \quad u \rightarrow \infty.$$

We refer to α as the index of regular variation. Clearly, α controls the heaviness of the tails. Equivalently, (1.3) can be written as the following pair of conditions.

1. The norm $\|\mathbf{X}\|$ is regularly varying at infinity, i.e., there exists $\alpha > 0$ such that

$$\lim_{u \rightarrow \infty} \frac{\mathbf{P}[\|\mathbf{X}\| > uy]}{\mathbf{P}[\|\mathbf{X}\| > u]} = y^{-\alpha}, \quad y \in (0, \infty). \quad (1.4)$$

2. The angular component $\mathbf{X}/\|\mathbf{X}\|$ is independent of $\|\mathbf{X}\|$ for large values of $\|\mathbf{X}\|$ in the sense that

$$\mathbf{P}[\mathbf{X}/\|\mathbf{X}\| \in \cdot \mid \|\mathbf{X}\| > u] \rightsquigarrow \mathbf{P}[\Theta \in \cdot], \quad u \rightarrow \infty. \quad (1.5)$$

For more in depth theory on multivariate regular variation see [Bingham et al. \(1989\)](#) or [Resnick \(2007\)](#).

The spectral measure

Classical extreme value theory describes the behavior of component-wise maxima through the multivariate version of the Fisher–Tippett–Gnedenko theorem ([Fisher and Tippett, 1928](#); [Gnedenko, 1943](#)). Regular variation with index $\alpha > 0$ of the d -dimensional iid random vectors $\{\mathbf{X}_i\}_{i \geq 1}$, with values in $(0, \infty)^d$, is necessary and sufficient condition for the fact that a suitably normalized sequence of component-wise maxima $(a_n^{-1} \max_{i \leq n} X_{ij})_{j=1, \dots, d}$, converges in distribution to a d -dimensional extreme value distribution G on $(0, \infty)^d$ whose marginal distributions are Fréchet distributed with shape parameter equal to α . In particular, when the marginal distributions are in the domain of attraction of the unit-Fréchet distribution, e.g. are unit-Pareto distributed, the multivariate extreme value distribution G has the form

$$G(\mathbf{x}) = \exp \left\{ -d \int_{\Delta_{d-1}} \max(\mathbf{w}/\mathbf{x}) H(d\mathbf{w}) \right\}, \quad \mathbf{x} > 0, \quad (1.6)$$

where $\Delta_{d-1} = \{\mathbf{w} \in [0, \infty)^d : \sum_{j=1}^d w_j = 1\}$ is the unit simplex and H is the spectral measure arising as the weak limit in

$$\mathbb{P} \left[\mathbf{x} / \sum_{j=1}^d x_j \in \cdot \mid \sum_{j=1}^d x_j > u \right] \rightsquigarrow H(\cdot), \quad u \rightarrow \infty. \quad (1.7)$$

Note that (1.7) corresponds to (1.5) with the L_1 norm and restricted to the first quadrant, i.e., left tails being negligible. The spectral measure is then defined on the unit simplex Δ_{d-1} which, when $d = 2$, becomes the unit interval $[0, 1]$.

Consider the bivariate case. In Figure 1.3 we show pseudo-random samples from a distribution with Gumbel copula and unit-Pareto margins at two different values of the dependence parameter φ : strong dependence ($\varphi = 2.5$) at the top and middle rows versus weak dependence ($\varphi = 1.25$) in the bottom row. At the top and bottom rows the sample size is set to 1000, while the middle plots correspond to sample size 10000. For convenience, the axes in the scatter plots on the left are on logarithmic scale. The points for which the sum of the coordinates belong to the top-100 are shown in black. In the middle column we plot the ratio of the first component to the sum ("angle, horizontal axis) against the sum itself ("radius, vertical axis). On the right, finally, histograms are given of these ratios for those points for which the sum exceeds the threshold in the plots on the left and in the middle.

We make two observations. First, from the top and the middle plots we see that increasing the sample size and the threshold does not influence the shape of the histograms on the right.

Second, strong extremal dependence leads to a distribution of the ratios being concentrated around 0.5 (top), whereas low extremal dependence leads to the ratios being close to either 0 or 1 (bottom).

We learn that the distribution of the ratios of the components to their sum, given that the sum is large, stabilizes at high levels and carries information about dependence between large values. This distribution is the spectral measure H in (1.7). The reference lines in Figure 1.3 represent the density of the spectral measure associated with the Gumbel copula with parameter $\varphi > 1$, that is,

$$\frac{dH_\varphi}{dw}(w) = \frac{\varphi - 1}{2} \{w(1 - w)\}^{-1-\varphi} \{w^{-\varphi} + (1 - w)^{-\varphi}\}^{1/\varphi-2}, \quad 0 < w < 1.$$

In Figure 1.4 we show the same plots as in Figure 1.3 but now for the pairs of negative log-returns of the stock prices of JP Morgan versus Citibank (top) and JP Morgan versus IBM (bottom). For each of the two pairs we do the following. Let (X_{i1}, X_{i2}) , $i = 1, \dots, n$, denote a bivariate sample. First, we transform the data to the unit-Pareto scale using and (1.1) and (1.2), i.e.,

$$\widehat{X}_{ij}^* = n/(n + 1 - R_{ij,n}), \quad i = 1, \dots, n, \quad j \in \{1, 2\}, \quad (1.8)$$

where $R_{ij,n}$ is the rank of X_{ij} among $X_{1j}, \dots, X_{nj}$. Next, we construct an approximate sample from the spectral measure H by setting

$$\widehat{S}_i = \widehat{X}_{i1}^* + \widehat{X}_{i2}^* \quad \text{and} \quad \widehat{W}_i = \widehat{X}_{i1}^*/(\widehat{X}_{i1}^* + \widehat{X}_{i2}^*), \quad i = 1, \dots, n. \quad (1.9)$$

The data in the pseudo-polar representation (1.9) are shown in the middle plots. The histograms of the ratios $\widehat{W}_i$ for those points for which the sum $\widehat{S}_i$ is above the 95% sample quantile are displayed on the right-hand side. The solid black lines show the smoothed spectral density estimator defined in Chapter 4 below, see (4.11).

For JP Morgan versus Citibank, the ratios are spread equally across the unit interval, suggesting extremal dependence. In contrast, for JP Morgan versus IBM, the spectral density has more weight at the boundaries 0 and 1, suggesting weak extremal dependence¹

The characterization of multivariate extreme-value distributions (1.6) via spectral measures goes back to de Haan and Resnick (1977). Their version of the spectral measure is different from the one presented here in that they used ordinary polar coordinates to decompose the vector into a radial and an angular component. This boils down to using the L_2 norm rather than the L_1 norm. For the radial component one can take any L_p norm, $p \in [1, +\infty]$.

¹Similar graphical analysis of the tail dependence within various financial data can be done at https://extreme.shinyapps.io/bivariate_tail_dependence.

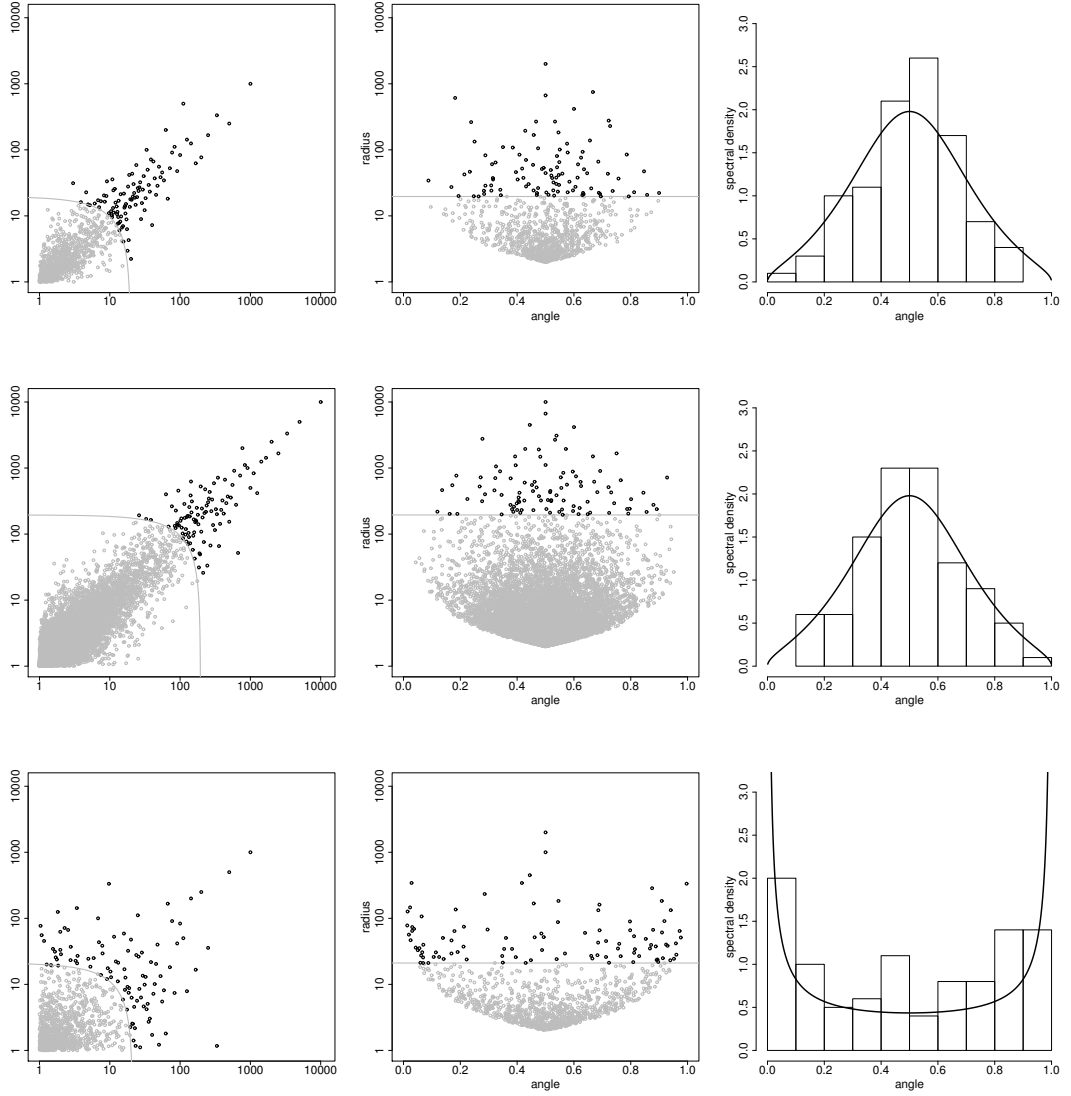


FIGURE 1.3: Pseudo-random samples from the Gumbel copula; $\varphi = 2.5$ (top and middle) and $\varphi = 1.25$ (bottom). Left: the data transformed to unit-Pareto margins (log scale); middle: further transformation into pseudo-polar coordinates; right: histogram of the angles of the extreme pairs and the true density of the spectral measure (solid line).

The fact that margins of the multivariate extreme value distribution G (1.6) are unit-Fréchet distributed implies the following moment constraints on the corresponding spectral measure

$$\int_{\Delta_{d-1}} w_j H(dw) = \frac{1}{d}, \quad j = 1, \dots, d. \quad (1.10)$$

Taking these constraints into account helps to improve the efficiency of nonparametric estimators of H .

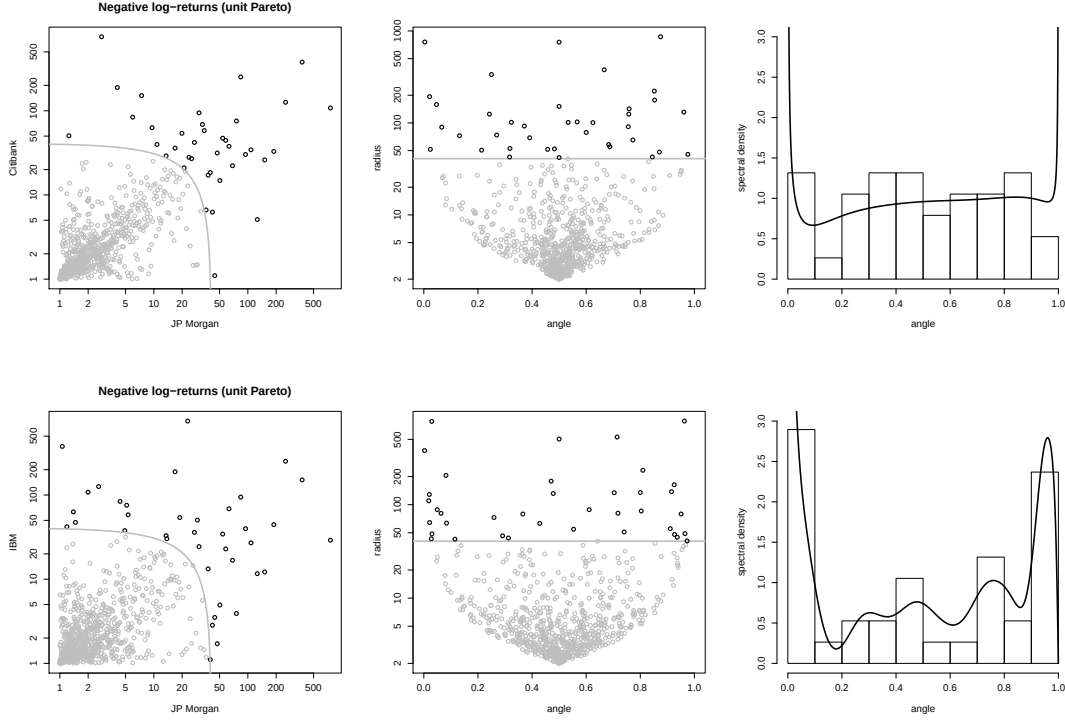


FIGURE 1.4: Log-returns of stock prices of JP Morgan versus Citibank (top) and JP Morgan versus IBM (bottom). Left: the data transformed to unit-Pareto margins (log scale); middle: further transformation into pseudo-polar coordinates; right: histogram of the angles of the extreme pairs and kernel density estimator (solid line).

The first nonparametric estimator of the spectral measure was proposed by [Einmahl et al. \(2001\)](#). The estimation procedure was then improved in [Einmahl and Segers \(2009\)](#) where the moment constraints on the spectral measure (1.10) are taken into account. Another estimator that satisfies the moment constraints but is free of Lagrange multipliers is the maximum *Euclidean likelihood estimator* of [de Carvalho et al. \(2013\)](#), which is described in detail in Chapter 4. For a summary on existing nonparametric techniques for estimating spectral measure in a bivariate setting, see Chapter 17 by Warchol et al. in [Dey and Yan \(2015\)](#).

1.3 Extremal dependence in time

The concept of multivariate regular variation can be extended to strictly stationary univariate time series. The notion of regular variation for stationary time series was first coined by [Davis and Hsing \(1995\)](#). We say that $(X_t)_{t \in \mathbb{Z}}$ is jointly regularly varying with index $\alpha > 0$, if for all integers $k \leq l$ the random vector $(X_k, \dots, X_l)$ is multivariate regularly varying with index α . In many respects, this condition can be viewed as the heavy-tailed analogue of a condition that a

process is Gaussian in the sense that all the finite-dimensional distributions are specified to be of a certain type. Note that by strict stationarity it is sufficient to take $k = 0 \leq l$. It follows that for every $l \geq 0$ there exists a spectral measure $P[\Theta \in \cdot]$ defined on the unit sphere $\mathbb{S}^l$ in $\mathbb{R}^{l+1}$. Clearly, all random variables X_t are then regularly varying with the same index α .

Many common time series models are jointly regularly varying. Examples include linear processes with heavy-tailed innovations, solutions to stochastic recurrence equations, models of the ARCH and GARCH families as well as the stochastic volatility model with heavy-tailed innovations. Sufficient conditions for such models to be regularly varying can be found in [Davis et al. \(2013\)](#).

Throughout, let $(X_t)_{t \in \mathbb{Z}}$ be a univariate, strictly stationary and jointly regularly varying time series. All the objects characterizing serial extremal dependence described below can be extended to a multivariate setting. Nevertheless, we continue with univariate time series for simplicity.

Serial extremal dependence is typically characterized by the presence of clusters of exceedances of high thresholds in the time series. The following objects are designed to capture serial extremal dependence properties.

- Aggregating positive and negative extremes, a straightforward summary of the extremal dependence within $(X_t)_{t \in \mathbb{Z}}$ is given by the sequence of *tail dependence coefficients*

$$\lambda_t := \lim_{u \rightarrow \infty} P[|X_t| > u \mid |X_0| > u], \quad t \in \mathbb{Z}, \quad (1.11)$$

provided the limits exist. By stationarity, $\lambda_t = \lambda_{-t}$. These pairwise tail dependence coefficients attracted some attention in the literature on quantitative risk management ([McNeil et al., 2005](#))

- Another description is provided by the *extremal index* ([Leadbetter, 1983](#)) which has become popular in extreme value literature. It can be viewed as the reciprocal of the expected extremal cluster size. It is defined as

$$\theta := \lim_{r \rightarrow \infty} \lim_{u \rightarrow \infty} P \left[\max_{t=1, \dots, r} |X_t| \leq u \mid |X_0| > u \right]. \quad (1.12)$$

For time series models that do not exhibit extremal clustering, like Gaussian stationary sequences under very weak conditions or stochastic volatility processes, the extremal index θ is equal to 1.

- An alternative tool for analysing extremal dependence, used for instance in data network models, is the *extremal dependence measure* ([Larsson and Resnick, 2012](#)). Let $\|\cdot\|$ be a norm on $\mathbb{R}^2$ and for $t \in \mathbb{Z}$ let $H_{0,t}$ be the spectral tail measure of (X_0, X_t) that arises as the weak limit in

$$\mathcal{L}((X_0, X_t)/\|(X_0, X_t)\| \mid \|(X_0, X_t)\| > u) \rightsquigarrow H_{0,t}, \quad u \rightarrow \infty.$$

Note that $H_{0,t} \neq H$. The extremal dependence measure is given by

$$\text{EDM}[X_0, X_t] := \int_{\mathbb{S}^1} a_0 a_t dH_{0,t}(d\mathbf{a}). \quad (1.13)$$

- The *extremogram* by [Davis and Mikosch \(2009\)](#) gives a more detailed description of serial extremal dependence. For suitable sets $A, B \subset \{x : |x| \geq 1\}$ and $t \geq 1$, the extremogram is defined as

$$\gamma_{AB}(t) := \lim_{n \rightarrow \infty} n \text{cov}(\mathbf{1}(u_n^{-1}X_0 \in A), \mathbf{1}(u_n^{-1}X_t \in B)), \quad (1.14)$$

where $\mathbf{1}(\cdot)$ is the indicator variable and u_n is such that $n \mathbb{P}[|X_0| > u_n] \rightarrow 1$ as $n \rightarrow \infty$. The quantity $\gamma_{AB}(t)$ measures the influence of the extremal event $\{X_0 \in u_n A\}$ at time 0 on the extremal event $\{X_h \in u_n B\}$, h lags apart. The extremogram can be viewed as an analog of the autocorrelation function which depends only on extreme values in the sequence. The statistical theory for the extremogram is already quite developed with a bootstrap technique for constructing confidence bands ([Davis et al., 2012](#)) and an extension to a spatial domain ([Cho et al., 2016](#)).

Further developments are also done in the frequency domain of time series analysis where the spectral representation is used to fully characterize the second order properties of a stationary time series. The extremogram has its frequency domain analog, the so-called extremal periodogram, which shares many of the classical properties of the periodogram. Its smoothed version ([Mikosch and Zhao, 2014](#)) is a consistent estimator of the spectral density (in the Fourier analysis sense) at a given frequency. Large peaks in the smoothed extremal periodogram can indicate significant extremal dependence and point at certain lengths of extremal cycles. Extended work in [Mikosch and Zhao \(2015\)](#) introduces integrated extremal periodogram which can be applied for constructing goodness-of-fit tests for a time series model based on its extremes.

The tail process

In order to get a more detailed picture of serial extremal dependence, we might want to distinguish between positive and negative extremes, for instance to allow for separate modelling of large positive and negative returns on a financial asset. Moreover, one may be interested in the exact value of the exceedance over u . These considerations lead us to study the full conditional distribution of arbitrary long stretches of the time series given that at a particular time instant a large value occurs.

Formally, a strictly stationary time series $(X_t)_{t \in \mathbb{Z}}$ is said to have a *tail process* $(Y_t)_{t \in \mathbb{Z}}$ if for all $s, t \in \mathbb{Z}$ such that $s \leq t$, we have

$$\mathcal{L}(u^{-1}X_s, \dots, u^{-1}X_t \mid |X_0| > u) \rightsquigarrow \mathcal{L}(Y_s, \dots, Y_t), \quad u \rightarrow \infty, \quad (1.15)$$

with the implicit understanding that the law of $|Y_0|$ is non-degenerate. The convergence is to be understood in the sense of weak convergence of finite-dimensional distributions. Existence of the tail dependence coefficients (1.11) is implied by existence of the tail process (1.15), and

$$\lambda_t = \mathbb{P}[|Y_t| > 1], \quad t \in \mathbb{Z}.$$

Here, we use the fact that the distribution of Y_t does not have any atom at $\mathbb{R} \setminus \{0\}$, a property which follows from the spectral decomposition in (1.17) below.

Specializing equation (1.15) to $t = 0$ implies that $\mathbb{P}[|X_0| > uy] / \mathbb{P}[|X_0| > u] \rightarrow \mathbb{P}[|Y_0| > y]$ as $u \rightarrow \infty$ for all continuity points y of the law of $|Y_0|$. Since the law of $|Y_0|$ was supposed to be non-degenerate, it follows that the function $|X_0|$ is regularly varying at infinity (1.4). In particular, the law of $|Y_0|$ is $\text{Pareto}(\alpha)$. More generally, the time series $(X_t)_{t \in \mathbb{Z}}$ admits a tail process $(Y_t)_{t \in \mathbb{Z}}$ with non-degenerate $|Y_0|$ if and only if $(X_t)_{t \in \mathbb{Z}}$ is jointly regularly varying with index $\alpha > 0$ (Basrak and Segers, 2009, Theorem 2.1).

In Section 1.2 we described how the radial component becomes independent of the angular part due to regular variation in the multivariate extremes setting. Such a decomposition is also possible in the context of jointly regularly varying time series. Define a normalized tail process, the *spectral tail process*, by

$$\Theta_t = Y_t / |Y_0|, \quad t \in \mathbb{Z}.$$

By (1.15) and the continuous mapping theorem, it follows that for all $s, t \in \mathbb{Z}$ such that $s \leq t$

$$\mathcal{L}(X_s / |X_0|, \dots, X_t / |X_0| \mid |X_0| > u) \rightsquigarrow \mathcal{L}(\Theta_s, \dots, \Theta_t), \quad u \rightarrow \infty. \quad (1.16)$$

The difference between (1.15) and (1.16) is that the variables X_t are normalized by the threshold u in the former and by $|X_0|$ in the latter. Such auto-normalization allows the tail process to be decomposed into two stochastically independent components, i.e.,

$$Y_t = |Y_0| \Theta_t, \quad t \in \mathbb{Z}. \quad (1.17)$$

Independence of $|Y_0|$ and $(\Theta_t)_{t \in \mathbb{Z}}$ is stated in Basrak and Segers (2009, Theorem 3.1). The random variable $|Y_0|$ characterizes the magnitudes of extremes, whereas $(\Theta_t)_{t \in \mathbb{Z}}$ captures serial dependence of extremes. The spectral tail process at time $t = 0$ yields information on the relative weights of the upper and lower tails of $|X_0|$: since $\Theta_0 = Y_0 / |Y_0| = \text{sign}(Y_0)$, we have

$$p = \mathbb{P}[\Theta_0 = +1] = \lim_{u \rightarrow \infty} \frac{\mathbb{P}[X_0 > u]}{\mathbb{P}[|X_0| > u]} \quad (1.18)$$

and likewise $1 - p = \mathbb{P}[\Theta_0 = -1]$. Note that the law of Θ_0 , also in the multivariate case, is the spectral measure of X_0 in the sense of (1.3). The spectral tail process Θ_t is not to be confused with the components of Θ in (1.3).

In equations (1.15) and (1.16), the conditioning event is $\{|X_0| > u\}$, making no distinction between positive extremes, $X_0 > u$, and negative extremes, $X_0 < -u$. Nevertheless, such a distinction is embedded in the spectral tail process itself. If $p = P[\Theta_0 = +1] > 0$, then (1.16) implies that for all $s, t \in \mathbb{Z}$, $s \leq t$, we have

$$\mathcal{L}(X_s/X_0, \dots, X_t/X_0 \mid X_0 > u) \rightsquigarrow \mathcal{L}(\Theta_s, \dots, \Theta_t \mid \Theta_0 = +1), \quad u \rightarrow \infty. \quad (1.19)$$

Similarly, if $1 - p = P[\Theta_0 = -1] > 0$, then

$$\mathcal{L}(X_s/|X_0|, \dots, X_t/|X_0| \mid X_0 < -u) \rightsquigarrow \mathcal{L}(\Theta_s, \dots, \Theta_t \mid \Theta_0 = -1), \quad u \rightarrow \infty. \quad (1.20)$$

For Markov chains, specializing (1.19) and (1.20) to the time instants $t = 1$ or $t = -1$ suffices to reconstruct the whole tail process (Janßen and Segers, 2014). Moreover, the distinction between positive and negative extremes can give more detailed characterization of the serial extremal dependence hidden in the data.

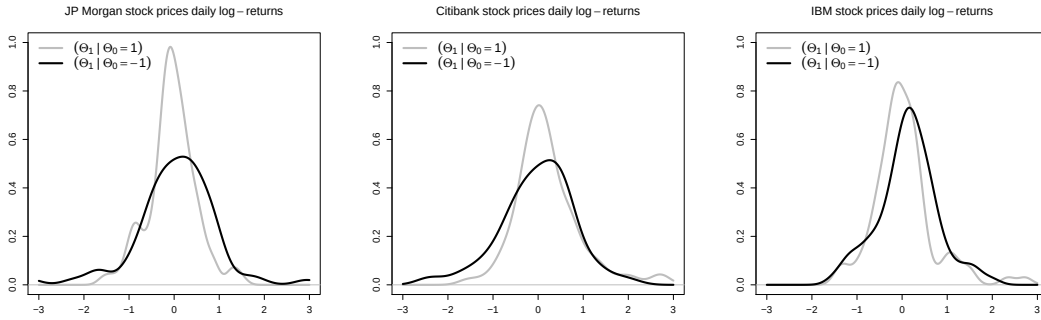


FIGURE 1.5: Estimated densities of the spectral tail process of the daily log-returns of JP Morgan (left), Citibank (middle) and IBM (right) stock prices.

As an example, in Figure 1.5 we plot the kernel smoothed estimates of the spectral density corresponding to the empirical versions of the left hand sides of (1.19) and (1.20) for lag $t = 1$. For all three time series of log-returns, the estimated spectral tail process at lag $t = 1$ puts more mass around zero when conditioning on a positive extreme value than when conditioning on a negative one. This suggests stronger serial extremal dependence at lag $t = 1$ after negative shocks than after positive ones. For more examples of applications see the case studies in Sections 2.7 and 3.5.²

²Extended graphical analysis of serial extremal dependence on other financial time series can be done at https://extreme.shinyapps.io/serial_extremal_dependence.

The time-change formula

Even though the underlying process $(X_t)_{t \in \mathbb{Z}}$ is assumed to be stationary, the (spectral) tail process is itself not stationary. Still, stationarity of $(X_t)_{t \in \mathbb{Z}}$ has consequences for the tail process $(Y_t)_{t \in \mathbb{Z}}$. For instance, for $x > 0$ and $h \in \mathbb{Z} \setminus \{0\}$, we have, by stationarity and regular variation,

$$\begin{aligned} \mathbb{P}[|Y_t| > y] &= \lim_{u \rightarrow \infty} \mathbb{P}[|X_t| > yu \mid |X_0| > u] \\ &= \lim_{u \rightarrow \infty} \mathbb{P}[|X_0| > u \mid |X_t| > yu] \frac{\mathbb{P}[|X_t| > yu]}{\mathbb{P}[|X_0| > u]} \\ &= y^{-\alpha} \mathbb{P}[|Y_{-t}| > y^{-1}]. \end{aligned}$$

Setting $y = 1$ yields $\lambda_t = \mathbb{P}[|Y_t| > 1] = \mathbb{P}[|Y_{-t}| > 1] = \lambda_{-t}$, as we already knew. For general $y > 0$, the equation in the display provides a way to retrieve the distribution of $|Y_t|$ from the one of $|Y_{-t}|$ and vice versa. More generally, it turns out that the distributions of the *forward* tail process $(Y_t)_{t \geq 0}$ and the *backward* tail process $(Y_t)_{t \leq 0}$ mutually determine each other.

The precise connection between the forward and backward spectral tail processes is as follows (Basrak and Segers, 2009, Theorem 3.1). For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all measurable functions $f : \mathbb{R}^{t-s+1} \rightarrow \mathbb{R}$ satisfying $f(y_s, \dots, y_t) = 0$ whenever $y_0 = 0$, we have, provided the expectations exist,

$$\mathbb{E}[f(\Theta_{s-i}, \dots, \Theta_{t-i})] = \mathbb{E}\left[f\left(\frac{\Theta_s}{|\Theta_i|}, \dots, \frac{\Theta_t}{|\Theta_i|}\right) |\Theta_i|^\alpha \mathbf{1}(\Theta_i \neq 0)\right]. \quad (1.21)$$

In particular, setting $s = 0$, $t = 1$, and $i = 1$ yields, for every measurable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $f(0, \cdot) = 0$ and for which the expectations exist,

$$\mathbb{E}[f(\Theta_{-1}, \Theta_0)] = \mathbb{E}\left[f\left(\frac{\Theta_0}{|\Theta_1|}, \frac{\Theta_1}{|\Theta_1|}\right) |\Theta_1|^\alpha \mathbf{1}(\Theta_1 \neq 0)\right]. \quad (1.22)$$

Setting $s = -1$, $t = 0$ and $i = -1$ yields the same formula, but with the roles of Θ_{-1} and Θ_1 interchanged. The condition that $f(0, \cdot) = 0$ cannot be omitted; to treat a general function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, consider the function $f(\theta_0, \theta_1) = g(\theta_0, \theta_1) - g(0, \theta_1)$.

We refer to (1.21) as the *time-change formula*. The time-change formula plays an important role in estimating the distribution of the spectral tail process. In particular, exploiting additional information streaming from the backward representation of the forward tail process can lead to more efficient inference, see Chapters 2 and 3 for details.

Related objects

The tail process does not only determine the tail dependence coefficients but it is also related to other insightful extreme value characteristics of time series.

- Under weak conditions, the extremal index (1.12) can be expressed through the tail process (Basrak and Segers, 2009, Remark 4.7) as

$$\theta = \mathbb{P} \left[\sup_{t \geq 1} |Y_t| \leq 1 \right] = \mathbb{E} \left[\sup_{t \geq 0} |\Theta_t|^\alpha - \sup_{t \geq 1} |\Theta_t|^\alpha \right].$$

- Assuming that all variables X_t are nonnegative, the extremal dependence measure (1.13) is written as

$$\text{EDM}[X_0, X_t] := \int_{\mathbb{S}_+^1} a_0 a_t dH_{0,t}(d\mathbf{a}) = \frac{\mathbb{E}[(1/\Gamma_t) (\Theta_t/\Gamma_t) \Gamma_t^\alpha]}{\mathbb{E}[\Gamma_t^\alpha]},$$

where $\mathbb{S}_+^1 = \{\mathbf{a} = (a_0, a_t) \in [0, \infty)^2 : \|\mathbf{a}\| = 1\}$ and $\Gamma_t = \|(1, \Theta_t)\|$.

- For suitable sets $A, B \subset \{x : |x| \geq 1\}$ and $t \geq 1$, the extremogram (1.14) can be written in terms of the tail process (Davis et al., 2013) as

$$\gamma_{AB}(t) = \mathbb{P}[Y_0 \in A, Y_t \in B].$$

Conversely, the tail process probabilities can be written as the extremogram with associated sets A and B . In this sense, the two objects are closely related. In contrast to the extremogram, the tail process is decomposed into two stochastically independent components: the radial component $|Y_0|$ modeling the magnitudes of extremes, and the spectral tail process Θ_t characterizing the dependence structure. Another difference is that the limit theory of the spectral tail process is established for nonparametric estimates of the distribution (see Chapter 2 and 3) whereas asymptotic theory for the extremogram specializes to particular sets A and B .

1.4 Objectives and structure of the thesis

The remainder of this thesis is organized in four chapters. Each of the following three chapters constitutes a reprint of a paper, published or submitted for publication. Respectively, Chapters 2, 3 and 4 correspond to Drees et al. (2015), Davis et al. (2016), and de Carvalho et al. (2013). The main objective shared by the three papers is to extend the statistical toolbox for modeling extremal dependence in space or time. In particular, the focus is on nonparametric techniques for estimating the spectral tail process and the spectral measure.

In Chapter 2 we consider serial extremal dependence of a stationary and regularly varying univariate Markov chain. We propose estimators for the spectral tail process exploiting additional information that comes from the time-change formula. The asymptotic distributions of the estimators are derived via functional central limit theorems building on the empirical process theory

developed by [Drees and Rootzén \(2010\)](#). The asymptotic results as well as the simulation study confirm that exploiting the time-change formula can lead to more efficient inference. Moreover, the estimators are applied to analyze daily log-returns on stock prices, providing insight into the dependence structure of consecutive extremes, covering both lower and upper tails.

In Chapter 3 we extend the methodology developed in Chapter 2 to the more general setting of a stationary and jointly regularly varying time series. The large-sample distributions of the estimators are derived with similar techniques. The finite-sample performance is evaluated via Monte Carlo simulations confirming that, also in the extended setting, exploiting the time-change formula can lead to more efficient inference. Moreover, two different bootstrap schemes are employed which yield confidence intervals for the pre-asymptotic spectral tail process: the stationary bootstrap used in [Davis et al. \(2012\)](#), and the multiplier block bootstrap described in [Drees \(2015\)](#). We confirm the validity of the proposed methodology and find that the multiplier block bootstrap outperforms the stationary bootstrap in terms of coverage probability. The estimators are applied to stock price data pointing out the asymmetric behavior in the persistence of positive and negative shocks.

In Chapter 4 we study extremal behavior of independent and identically distributed bivariate random vectors. A Euclidean likelihood-based estimator for the spectral measure is proposed. In contrary to the maximum empirical likelihood estimator of [Einmahl and Segers \(2009\)](#), the Euclidean likelihood-based estimator is free of Lagrange multipliers, making it thus easier to compute. It is shown that the empirical processes associated with both estimators have the same limiting distribution. The theoretical results are supported by numerical simulations. We also provide an illustration with extreme temperature data.

In Chapter 5 we present a brief discussion and propose directions for future research.

Bibliography

- Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* 119(4), 1055–1080.
- Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular variation*, Volume 27. Cambridge university press.
- Cho, Y. B., R. A. Davis, and S. Ghosh (2016). Asymptotic properties of the empirical spatial extremogram. *Scandinavian Journal of Statistics*.
- Davis, R., H. Drees, J. Segers, and M. Warchol (2016). Modeling serial extremal dependence. *arXiv preprint arXiv:1604.00954*.
- Davis, R. and T. Mikosch (2009). The extremogram: a correlogram for extreme events. *Bernoulli* 15(4), 977–1009.
- Davis, R. A. and T. Hsing (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability* 23(2), pp. 879–917.
- Davis, R. A., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* 170(1), 142 – 152.
- Davis, R. A., T. Mikosch, and Y. Zhao (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications* 123(7), 2575–2602.
- de Carvalho, M., B. Oumow, J. Segers, and M. Warchol (2013). A euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics - Theory and Methods* 42(7), 1176–1192.
- de Haan, L. and S. I. Resnick (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 40(4), 317–337.
- Dey, D. K. and J. Yan (2015). *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman and Hall/CRC.
- Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. *arXiv preprint arXiv:1511.00420*.

- Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics* 38(4), 2145–2186.
- Drees, H., J. Segers, and M. Warchoř (2015). Statistics for tail processes of markov chains. *Extremes* 18(3), 369–402.
- Einmahl, J. H., L. de Haan, and V. I. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics* 29(5), 1401–1423.
- Einmahl, J. H. J. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics* 37(5B), 2953–2989.
- Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 24, pp. 180–190. Cambridge Univ Press.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, 423–453.
- Janßen, A. and J. Segers (2014). Markov tail chains. *Advances in Applied Probability* 51(4), 1133–1153.
- Larsson, M. and S. I. Resnick (2012). Extremal dependence measure and extremogram: the regularly varying case. *Extremes* 15(2), 231–256.
- Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 65(2), 291–306.
- McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management*. Princeton, New Jersey: Princeton University Press.
- Mikosch, T. and Y. Zhao (2014). A Fourier analysis of extreme events. *Bernoulli* 20(2), 803–845.
- Mikosch, T. and Y. Zhao (2015). The integrated periodogram of a dependent extremal event sequence. *Stochastic Processes and their Applications* 125(8), 3126 – 3169.
- Resnick, S. I. (2007). *Extreme values, regular variation, and point processes*. Springer.

Chapter 2

Statistics for Tail Processes of Markov Chains

Holger Drees¹, Johan Segers², and Michał Warchol²

Manuscript published in *Extremes* (2015) 18(3), 369–402³.

Abstract

At high levels, the asymptotic distribution of a stationary, regularly varying Markov chain is conveniently given by its tail process. The latter takes the form of a geometric random walk, the increment distribution depending on the sign of the process at the current state and on the flow of time, either forward or backward. Estimation of the tail process provides a nonparametric approach to analyze extreme values. A duality between the distributions of the forward and backward increments provides additional information that can be exploited in the construction of more efficient estimators. The large-sample distribution of such estimators is derived via empirical process theory for cluster functionals. Their finite-sample performance is evaluated via Monte Carlo simulations involving copula-based Markov models and solutions to stochastic recurrence equations. The estimators are applied to stock price data to study the absence or presence of symmetries in the succession of large gains and losses.

¹University of Hamburg, Department of Mathematics, Hamburg, Germany

²Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Louvain-la-Neuve, Belgium

³Appendix B (Section 2.9) has been published as a supplement to the paper.

2.1 Introduction

If serial dependence at high levels is sufficiently strong, extreme values of a stationary time series may arrive in clusters rather than in isolation. This is the case, for instance, for linear time series with heavy-tailed innovations and for solutions of stochastic recurrence equations. If a particular time series model is to be used for prediction at such high levels, it is important to model these clusters well. Think of tail-related risk measures in finance or of return levels in hydrology: a rapid succession of particularly rainy days may be especially dangerous if the capacity of the system to absorb the water is limited.

To judge the quality of fit of a time series model at extreme levels, it is useful to have a benchmark relying on as little model assumptions as possible. A purely nonparametric approach, however, has the drawback that there may be too few data that are sufficiently large. For the purpose of extrapolation, the empirical measure is inadequate.

A solution is to rely on asymptotic theory describing possible limit distributions for the extremes of a time series. If this family of distributions is not too large, one may hope to be able to fit it to actual data.

For extremes of stationary time series, there are several asymptotic frameworks available, all of them more or less equivalent. For the study of short-range extremal dependence, the tail process ([Basrak and Segers, 2009](#)) is a convenient choice. It captures the collection of finite-dimensional limit distributions of the series conditionally on the event that, at a particular time instant, the series is far from the origin. For instance, the tail process determines the tail dependence coefficients and the extremal index ([Leadbetter, 1983](#)). It is also related to other tail-related objects such as the extremogram ([Davis and Mikosch, 2009](#)) and the extremal dependence measure ([Larsson and Resnick, 2012](#)).

The family of tail processes of regularly varying time series is still too large to permit accurate nonparametric estimation. Additional assumptions serve to render the inference problem more manageable. The choice made in this paper is to focus on stationary univariate Markov chains. The joint distribution of such a chain is determined by its bivariate margins, yielding considerable simplifications. Its tail process takes the form of a geometric random walk, the increments depending both on the sign of the process at the current state and on the direction of time, forward or backward. The random walk representation goes back to [Smith \(1992\)](#) and was developed further in [Perfekt \(1997\)](#), [Bortot and Coles \(2000\)](#), and [Yun \(2000\)](#). The formulation in terms of the tail process stems from [Segers \(2007\)](#) and [Janßen and Segers \(2014\)](#). By a marginal standardization procedure, the tail process may also be used for time series with light-tailed margins. Such time series arise for instance in environmental applications.

The tail process of a stationary time series is itself not stationary because of the special role played by the time instant figuring in the conditioning event. Still, its finite-dimensional distributions satisfy a collection of identities regarding the effect of a time shift. These equations can be summarized into the so-called time-change formula; see equation (2.6) below. Apart from being a probabilistic nicety, the time-change formula is useful from a statistical perspective because it provides additional information on the distribution of the tail process. Exploiting this information can lead to more efficient inference.

Our contribution is to propose and study nonparametric estimators for the tail process of a stationary univariate Markov chain. Large-sample theory and Monte Carlo simulations both confirm that efficiency gains are possible when the time-change formula is incorporated into the estimation procedure. The asymptotic distributions of the estimators are described via functional central limit theorems building on the empirical process theory developed in [Drees and Rootzén \(2010\)](#). The finite-sample performance is investigated for solutions of stochastic recurrence equations and for copula-based Markov models ([Chen and Fan, 2006](#)). We focus on the estimation of cumulative distribution functions. Following [Bortot and Coles \(2000\)](#), however, one could also use kernel methods to estimate their densities.

The structure of the paper is as follows. Tail processes are reviewed in Section 2.2, with special attention to those of Markov chains. The estimators of the tail process of a regularly Markov chain are described in Section 2.3. Their asymptotic properties are worked out in Sections 2.4, whereas their finite-sample performance is evaluated via Monte Carlo simulations in Section 2.6 involving models presented in the Section 2.5. In Section 2.7, the estimators are applied to analyze time series of daily log returns of Google and UBS stock prices, revealing interesting patterns regarding the succession of large losses and gains. Proofs and calculations are deferred to Section 2.8.

2.2 Tail processes of Markov chains

A strictly stationary time series $(X_t)_{t \in \mathbb{Z}}$ is said to have a *tail process* $(Y_t)_{t \in \mathbb{Z}}$ if, for all $s, t \in \mathbb{Z}$ such that $s \leq t$,

$$\mathcal{L}(u^{-1}X_s, \dots, u^{-1}X_t \mid |X_0| > u) \rightsquigarrow \mathcal{L}(Y_s, \dots, Y_t), \quad u \rightarrow \infty, \quad (2.1)$$

with the implicit understanding that the law of $|Y_0|$ is non-degenerate.

Specializing equation (2.1) to $t = 0$ implies that $\mathbb{P}[|X_0| > uy] / \mathbb{P}[|X_0| > u] \rightarrow \mathbb{P}[|Y_0| > y]$ as $u \rightarrow \infty$ for all continuity points y of the law of $|Y_0|$. Since the law of $|Y_0|$ was supposed to be non-degenerate, it follows that the function $u \mapsto \mathbb{P}[|X_0| > u]$ is regularly varying at infinity:

there exists $\alpha > 0$ such that

$$\lim_{u \rightarrow \infty} \frac{\mathbb{P}[|X_0| > uy]}{\mathbb{P}[|X_0| > u]} = y^{-\alpha}, \quad y \in (0, \infty). \quad (2.2)$$

The law of $|Y_0|$ is thus $\text{Pareto}(\alpha)$, i.e., $\mathbb{P}[|Y_0| > y] = y^{-\alpha}$ for all $y \geq 1$. More generally, by [Basrak and Segers \(2009, Theorem 2.1\)](#), the time series $(X_t)_{t \in \mathbb{Z}}$ admits a tail process $(Y_t)_{t \in \mathbb{Z}}$ with non-degenerate $|Y_0|$ if and only if $(X_t)_{t \in \mathbb{Z}}$ is jointly regularly varying with some index $\alpha > 0$, i.e., if for all integers $k \leq l$ the random vector $(X_k, \dots, X_l)$ is multivariate regularly varying with index α .

Many time series models are jointly regularly varying and hence admit a tail process. Examples include linear processes with heavy-tailed innovations, solutions to stochastic recurrence equations, and models of the ARCH and GARCH families. Sufficient conditions for such models to be regularly varying can be found in [Davis et al. \(2013\)](#).

The *spectral tail process* is defined by

$$\Theta_t = Y_t/|Y_0|, \quad t \in \mathbb{Z}.$$

By [\(2.1\)](#) and the continuous mapping theorem, it follows that for all $s, t \in \mathbb{Z}$ such that $s \leq t$

$$\mathcal{L}(X_s/|X_0|, \dots, X_t/|X_0| \mid |X_0| > u) \rightsquigarrow \mathcal{L}(\Theta_s, \dots, \Theta_t), \quad u \rightarrow \infty. \quad (2.3)$$

The difference between [\(2.1\)](#) and [\(2.3\)](#) is that in the latter equation, the variables X_t are normalized by $|X_0|$ rather than by the threshold u . Such auto-normalization allows the tail process to be decomposed into two stochastically independent components, i.e.,

$$Y_t = |Y_0| \Theta_t, \quad t \in \mathbb{Z}. \quad (2.4)$$

Independence of $|Y_0|$ and $(\Theta_t)_{t \in \mathbb{Z}}$ is stated in [Basrak and Segers \(2009, Theorem 3.1\)](#). The random variable $|Y_0|$ characterizes the magnitudes of extremes, whereas $(\Theta_t)_{t \in \mathbb{Z}}$ captures serial dependence. The spectral tail process at time $t = 0$ yields information on the relative weights of the upper and lower tails of $|X_0|$: since $\Theta_0 = Y_0/|Y_0| = \text{sign}(Y_0)$, we have

$$p = \mathbb{P}[\Theta_0 = +1] = \lim_{u \rightarrow \infty} \frac{\mathbb{P}[X_0 > u]}{\mathbb{P}[|X_0| > u]}, \quad 1 - p = \mathbb{P}[\Theta_0 = -1]. \quad (2.5)$$

The distributions of the *forward* tail process $(Y_t)_{t \geq 0}$ and the *backward* tail process $(Y_t)_{t \leq 0}$ mutually determine each other. The precise connection between the forward and backward (spectral) tail processes is captured by Theorem 3.1 in [Basrak and Segers \(2009\)](#). For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all measurable functions $f : \mathbb{R}^{t-s+1} \rightarrow \mathbb{R}$ satisfying $f(y_s, \dots, y_t) = 0$ whenever

$y_0 = 0$, we have, provided the expectations exist,

$$\mathbb{E}[f(\Theta_{s-i}, \dots, \Theta_{t-i})] = \mathbb{E}\left[f\left(\frac{\Theta_s}{|\Theta_i|}, \dots, \frac{\Theta_t}{|\Theta_i|}\right) |\Theta_i|^\alpha \mathbf{1}\{\Theta_i \neq 0\}\right]. \quad (2.6)$$

We will refer to (2.6) as the *time-change formula*. By exploiting the time-change formula, we will be able to improve upon the efficiency of estimators of the tail process.

A common procedure in multivariate extreme value theory is to standardize the margins. For jointly regularly varying time series, such a standardization is possible too, although some care is needed because of the possible presence of both positive and negative extremes.

Lemma 2.1. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary time series, jointly regularly varying with index $\alpha > 0$, and having spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$. Put $\bar{F}_{|X_0|}(u) = \mathbb{P}[|X_0| > u]$ for $u \geq 0$. Define a stationary time series $(X_t^*)_{t \in \mathbb{Z}}$ by*

$$X_t^* = \frac{\text{sign}(X_t)}{\bar{F}_{|X_0|}(|X_t|)}, \quad t \in \mathbb{Z}. \quad (2.7)$$

Then $(X_t^*)_{t \in \mathbb{Z}}$ is jointly regularly varying with index 1. Its spectral tail process $(\Theta_t^*)_{t \in \mathbb{Z}}$ is given by

$$\Theta_t^* = \text{sign}(\Theta_t) |\Theta_t|^\alpha, \quad t \in \mathbb{Z}. \quad (2.8)$$

In (2.8), note that the map $y \mapsto \text{sign}(y) |y|^\alpha$ is monotone and symmetric. The standardized series $(X_t^*)_{t \in \mathbb{Z}}$ may be regularly varying even if the original series $(X_t)_{t \in \mathbb{Z}}$ is not. In that sense, the standardization procedure in (2.7) widens the field of possible applications of tail processes. For instance, the marginal distributions of environmental variables are often light-tailed rather than regularly varying. After standardization as in Lemma 2.1, the serial dependence of extremes of such time series may still be modelled via tail processes.

Some time series models exhibit asymptotic independence of consecutive observations, that is, $\mathbb{P}[|X_k| > u \mid |X_0| > u] \rightarrow 0$ as $\mathbb{P}[|X_0| > u] \rightarrow 0$ for all $k \in \mathbb{Z} \setminus \{0\}$. Well-known examples are non-degenerate Gaussian time series and classical stochastic volatility models. In such cases, the spectral tail process is noninformative in the sense that $\Theta_k = 0$ almost surely for all $k \neq 0$. More refined approaches to handle tail independence were developed in [Ledford and Tawn \(1996, 2003\)](#) and, more recently, in [Janßen and Drees \(2013\)](#) and [Kulik and Soulier \(2013\)](#).

Regularly varying Markov chains

For the purpose of statistical inference, the class of spectral tail processes is too large to be really useful: without additional modelling assumptions, it is impossible to estimate all limiting finite-dimensional distributions that appear in (2.1) or (2.3). Therefore, it is reasonable to consider

families of spectral tail processes arising under additional constraints on the underlying time series.

One such family was identified in Segers (2007) and Janßen and Segers (2014) in the context of first-order Markov chains. Let $(\Theta_t)_{t \in \mathbb{Z}}$ be the spectral tail process of an α -regularly varying, stationary time series $(X_t)_{t \in \mathbb{Z}}$, arising as the limit process in (2.3). Put $p = \mathbb{P}[\Theta_0 = 1]$ as in (2.5). Introduce random variables A_1, B_1, A_{-1}, B_{-1} (or rather their laws) as follows: if $p > 0$, then, as $u \rightarrow \infty$,

$$\mathcal{L}(X_1/X_0 \mid X_0 > u) \rightsquigarrow \mathcal{L}(A_1) = \mathcal{L}(\Theta_1 \mid \Theta_0 = +1), \quad (2.9)$$

$$\mathcal{L}(X_{-1}/X_0 \mid X_0 > u) \rightsquigarrow \mathcal{L}(A_{-1}) = \mathcal{L}(\Theta_{-1} \mid \Theta_0 = +1), \quad (2.10)$$

and if $p < 1$, then

$$\mathcal{L}(X_1/X_0 \mid X_0 < -u) \rightsquigarrow \mathcal{L}(B_1) = \mathcal{L}(-\Theta_1 \mid \Theta_0 = -1), \quad (2.11)$$

$$\mathcal{L}(X_{-1}/X_0 \mid X_0 < -u) \rightsquigarrow \mathcal{L}(B_{-1}) = \mathcal{L}(-\Theta_{-1} \mid \Theta_0 = -1). \quad (2.12)$$

Further, let $\Theta_0, A_1, A_{-1}, A_2, A_{-2}, \dots, B_1, B_{-1}, B_2, B_{-2}$ be independent random variables such that $\mathcal{L}(A_t) = \mathcal{L}(A_1)$, $\mathcal{L}(A_{-t}) = \mathcal{L}(A_{-1})$, $\mathcal{L}(B_t) = \mathcal{L}(B_1)$, and $\mathcal{L}(B_{-t}) = \mathcal{L}(B_{-1})$ for all $t \in \mathbb{N}$. Then the spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$ is said to be a *Markov spectral tail chain* if the following holds: the forward spectral tail process is given recursively by

$$\Theta_t = \begin{cases} \Theta_{t-1} A_t & \text{if } \Theta_{t-1} > 0, \\ 0 & \text{if } \Theta_{t-1} = 0, \\ \Theta_{t-1} B_t & \text{if } \Theta_{t-1} < 0, \end{cases} \quad t \in \mathbb{N}, \quad (2.13)$$

whereas the backward spectral tail process is given by

$$\Theta_{-t} = \begin{cases} \Theta_{-t+1} A_{-t} & \text{if } \Theta_{-t+1} > 0, \\ 0 & \text{if } \Theta_{-t+1} = 0, \\ \Theta_{-t+1} B_{-t} & \text{if } \Theta_{-t+1} < 0, \end{cases} \quad t \in \mathbb{N}. \quad (2.14)$$

If $p = 1$, then $\Theta_t \geq 0$ almost surely for all $t \in \mathbb{Z}$ and thus the definition of $B_{\pm t}$ is immaterial; similarly if $p = 0$. This can be seen by applying the time-change formula (2.6).

The motivation behind the above definition is that such spectral tail processes typically arise when $(X_t)_{t \in \mathbb{Z}}$ is a stationary, first-order Markov chain; see Theorem 5.2 in Segers (2007) and Corollary 5.1 in Janßen and Segers (2014). However, they may as well arise in settings where the underlying process, $(X_t)_{t \in \mathbb{Z}}$, is non-Markovian; see Remark 5.1 in Janßen and Segers (2014). The forward and backward spectral tail processes $(\Theta_t)_{t \geq 0}$ and $(\Theta_t)_{t \leq 0}$ are Markovian themselves, and, conditionally on Θ_0 , they are independent. Their structure is that of a geometric random

walk where the distribution of the increment at time t depends on the sign of the process at time $t - 1$. The point zero acts as an absorbing state.

For Markov spectral tail chains, the distribution of the forward part $(\Theta_t)_{t \geq 0}$ is determined by p , A_1 , and B_1 . Given additionally the index of regular variation $\alpha > 0$, the distributions of A_{-1} and B_{-1} and thus of the backward part $(\Theta_t)_{t \leq 0}$ can be reconstructed from the time-change formula (2.6); see Lemma 2.2 below. It follows that the law of a Markov spectral tail process is determined by $\alpha > 0$, $p \in [0, 1]$, and the laws of A_1 and B_1 . This reduction provides a handle on the spectral tail process that can be exploited for statistical inference.

2.3 Estimating Markov spectral tail processes

In this section we propose estimators for p , A_1 and B_1 . In combination with the index of regular variation $\alpha > 0$, this triplet fully determines the law of a Markov spectral tail process as defined in equations (2.13) and (2.14), and of the tail processes $(Y_t)_{t \in \mathbb{Z}}$.

Replacing population distributions by sampling distributions in the left-hand sides of (2.9) and (2.11) yields forward estimators for the laws of A_1 and B_1 . However, exploiting the time-change formula (2.6) allows to express the laws of A_1 and B_1 in terms of A_{-1} and B_{-1} (and p and α). These expressions motivate so-called backward estimators for A_1 and B_1 . Convex combinations of forward and backward estimators finally produce mixture estimators. For an appropriate choice of the mixture weights, the mixture estimators may be more efficient than both the forward and the backward estimators separately.

In order to estimate $p = \mathbb{P}[\Theta_0 = 1]$, we simply take the empirical version of (2.5), yielding

$$\hat{p}_n = \frac{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}. \quad (2.15)$$

For $\hat{p}_n$ to be consistent and asymptotically normal, the threshold sequence u_n should tend to infinity at a certain rate described in detail in condition **(B)** in the next section.

For estimating the cdf, $F^{(A_1)}$, of A_1 we propose

$$\hat{F}_n^{(f, A_1)}(x) = \frac{\sum_{i=1}^n \mathbf{1}(X_{i+1}/X_i \leq x, X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}, \quad (2.16)$$

which we refer to as the *forward estimator* of the cdf of A_1 . Similarly, for the forward estimator of the cdf of B_1 we take

$$\hat{F}_n^{(f, B_1)}(x) = \frac{\sum_{i=1}^n \mathbf{1}(X_{i+1}/X_i \leq x, X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}(X_i < -u_n)}. \quad (2.17)$$

The forward estimators of the cdf's of A_1 and B_1 are empirical versions of the left-hand sides of (2.9) and (2.11), respectively. Note that one can expect consistency of these estimators only if the target distribution functions are continuous in x , because otherwise $\mathbb{P}[X_1/X_0 \leq x \mid X_0 > u_n]$ need not converge to $\mathbb{P}[A_1 \leq x]$, for instance.

The time-change formula (2.6) yields a different representation of A_1 and B_1 , motivating different estimators than the ones above, based on different data points. For ease of reference, we record the relevant formulas in a lemma, whose proof is given in Appendix 2.8.2.

Lemma 2.2. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary time series, jointly regularly varying with index α and spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$. Let A_1, A_{-1}, B_1, B_{-1} be given as in (2.9) to (2.12). If $p = \mathbb{P}[\Theta_0 = 1] > 0$, then*

$$\mathbb{P}[A_1 > x] = \mathbb{E}[A_{-1}^\alpha \mathbf{1}(1/A_{-1} > x)], \quad x \geq 0, \quad (2.18)$$

$$\mathbb{P}[A_1 \leq x] = \frac{1-p}{p} \mathbb{E}[(-B_{-1})^\alpha \mathbf{1}(1/B_{-1} \leq x)], \quad x < 0. \quad (2.19)$$

Similarly, if $p < 1$, then

$$\mathbb{P}[B_1 > x] = \mathbb{E}[B_{-1}^\alpha \mathbf{1}(1/B_{-1} > x)], \quad x \geq 0, \quad (2.20)$$

$$\mathbb{P}[B_1 \leq x] = \frac{p}{1-p} \mathbb{E}[(-A_{-1})^\alpha \mathbf{1}(1/A_{-1} \leq x)], \quad x < 0. \quad (2.21)$$

Formulas (2.18) to (2.21) remain valid when the time instances 1 and -1 are interchanged.

Assume for the moment that α is known. Below, we will consider the more realistic situation that α is unknown. Lemma 2.2 suggests the following *backward estimator* of the cdf of A_1 :

$$\hat{F}_n^{(b, A_1)}(x) = \begin{cases} 1 - \frac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^\alpha \mathbf{1}(X_i/X_{i-1} > x, X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(X_i > u_n)} & \text{if } x \geq 0, \\ \frac{\sum_{i=1}^n \left(\frac{-X_{i-1}}{X_i}\right)^\alpha \mathbf{1}(X_i/X_{i-1} \leq x, X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}(X_i > u_n)} & \text{if } x < 0. \end{cases} \quad (2.22)$$

Similarly, we define the backward estimator of the cdf of B_1 as

$$\hat{F}_n^{(b, B_1)}(x) = \begin{cases} 1 - \frac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i}\right)^\alpha \mathbf{1}(X_i/X_{i-1} > x, X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}(X_i < -u_n)} & \text{if } x \geq 0, \\ \frac{\sum_{i=1}^n \left(\frac{-X_{i-1}}{X_i}\right)^\alpha \mathbf{1}(X_i/X_{i-1} \leq x, X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(X_i < -u_n)} & \text{if } x < 0. \end{cases} \quad (2.23)$$

For $|x|$ large, the backward estimators usually have a smaller variance than the forward estimators. To see this, note that for negative x with large modulus only very few summands in the

numerator of (2.16) do not vanish, because X_{i+1} must be even larger in absolute value than $|x|X_i > |x|u_n$, leading to a large variance of the numerator. In contrast, usually many more non-vanishing terms will be summed up in the numerator of (2.22), while each of them gets a rather low weight $(-X_{i-1}/X_i)^\alpha \leq |x|^{-\alpha}$, leading to a smaller variance. For large positive x one may argue similarly by considering the corresponding estimators of the survival function. Indeed, we show in Remark 2.4 that, provided $x \geq 1$, the backward estimator of the cdf of A_1 at x has a smaller asymptotic variance than the forward estimator.

For well-chosen weights, convex combinations of the forward and backward estimators can achieve a lower asymptotic variance than each of the estimators individually. Unfortunately, the expression for the asymptotic covariance of the two estimators is intractable; see Theorem 2.3. It remains an open issue how to choose the mixture weights in order to minimize the asymptotic variance.

A pragmatic approach is to give more weight to the forward estimator for small $|x|$ and to give more weight to the backward estimator for large $|x|$. To this end, define weights by

$$\lambda(x) = \max(1 - |x|, 0).$$

The *mixture estimator* for the cdf of A_1 is defined as

$$\hat{F}_n^{(m, A_1)}(x) = \begin{cases} \lambda(x) \hat{F}_n^{(f, A_1)}(x) + [1 - \lambda(x)] \hat{F}_n^{(b, A_1)}(x) & \text{if } x \geq 0, \\ \lambda(x) \hat{F}_n^{(f, A_1)}(x) + [1 - \lambda(x)] \hat{F}_n^{(b, A_1)}(x) & \text{if } x < 0. \end{cases} \quad (2.24)$$

The mixture estimator for B_1 is defined by replacing A_1 in (2.24) with B_1 .

The above estimators require the selection of a high threshold u_n , which involves a bias-variance trade-off: the lower the threshold, the larger the estimation bias will usually be; the higher the threshold, the larger the variance. In practice, we take a widely-used informal graphical approach based on stability plots, i.e. we plot the estimated cdf's for several levels and choose a threshold above which the estimators look relatively stable.

The backward and the mixture estimators also require the value of the index α of regular variation, which is unknown in most applications. There are at least two approaches to deal with this issue:

1. Estimate α separately, for instance, by the Hill-type estimator

$$\hat{\alpha} = \frac{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}{\sum_{i=1}^n \log(|X_i|/u_n) \mathbf{1}(|X_i| > u_n)}, \quad (2.25)$$

and plug in the estimated value of α in (2.22), (2.23) and (2.24).

2. Employ an empirical version of the transformation in Lemma 2.1 to ensure that, after transformation, $\alpha = 1$. The transformation in (2.7) requires the tail function $\bar{F}_{|X_0|}(u) = \mathbb{P}[|X_0| > u]$. This function can be estimated, for instance, by

$$\hat{\bar{F}}_{|X_0|,n}(u) = 1 - \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(|X_j| \leq u), \quad (2.26)$$

where we divide by $n+1$ rather than by n in order to avoid division by zero later on. The transformed variable

$$\hat{X}_{n,i}^* = \text{sign}(X_i) / \hat{\bar{F}}_{|X_0|,n}(|X_i|)$$

is based on the sign of X_i and the rank of $|X_i|$ among $|X_1|, \dots, |X_n|$.

In the simulation study in Section 2.6, the mixture estimator based on the rank-transformed data performs better than the plug-in version. Note, however, that the two approaches are not directly comparable: with the second approach, what we estimate is the tail process $(\Theta_t^*)_{t \in \mathbb{Z}}$ of the transformed series $(X_t^*)_{t \in \mathbb{Z}}$. From (2.8) and (2.9), it follows that, if $(\Theta_t)_{t \in \mathbb{Z}}$ is a Markov spectral tail chain as in (2.13) and (2.14), then so is $(\Theta_t^*)_{t \in \mathbb{Z}}$, with A_t and B_t to be replaced by $A_t^* = \text{sign}(A_t) |A_t|^\alpha$ and $B_t^* = \text{sign}(B_t) |B_t|^\alpha$, respectively. Combining the above two estimation approaches, one could even recover A_t and B_t via $A_t = \text{sign}(A_t^*) |A_t^*|^{1/\alpha}$ and $B_t = \text{sign}(B_t^*) |B_t^*|^{1/\alpha}$. Finally, as $(X_t^*)_{t \in \mathbb{Z}}$ may have a tail process even when the original time series has none, the second approach is more widely applicable.

2.4 Large sample theory

Under certain conditions, the standardized estimation errors of the forward and the backward estimators converge jointly to a centered Gaussian process. In order not to overload the presentation, we focus on nonnegative Markov chains. In that case, the distribution of $\Theta_1 = A_1$ determines the distribution of the forward spectral tail process, and thus, via the time-change formula, together with α , also the one of the backward spectral tail process. We distinguish between the cases where α is known (Section 2.4.1) and unknown (Section 2.4.2). In addition, we briefly indicate how the conditions and results must be modified in the real-valued case (Remark 2.8).

2.4.1 Known index of regular variation

If the index of regular variation, α , is known, all estimators under consideration can be expressed in terms of *generalized tail array sums*, that is, statistics of the form $\sum_{i=1}^n \phi(X_{n,i})$, with

$$X_{n,i} := \frac{(X_{i-1}, X_i, X_{i+1})}{u_n} \mathbf{1}(X_i > u_n). \quad (2.27)$$

Drees and Rootzén (2010) give conditions under which, after standardization, such statistics converge to a centered Gaussian process, uniformly over appropriate families of functions ϕ . From these results we will deduce a functional central limit theorem for the processes of forward and backward estimators defined in (2.16) and (2.22), respectively.

To ensure consistency, the threshold u_n must tend to infinity such that

$$v_n := \mathbb{P}[X_0 > u_n]$$

tends to 0, but the expected number nv_n of exceedances tends to infinity. Moreover, let

$$\beta_{n,k} := \sup_{1 \leq l \leq n-k-1} \mathbb{E} \left[\sup_{B \in \mathcal{B}_{n,l+k+1}^n} |\mathbb{P}[B | \mathcal{B}_{n,1}^l] - \mathbb{P}[B]| \right]$$

denote the β -mixing coefficients. Here $\mathcal{B}_{n,i}^j$ is the σ -field generated by $(X_{n,l})_{i \leq l \leq j}$. We assume that there exist sequences $l_n, r_n \rightarrow \infty$ and some $x_0 \geq 0$ such that the following conditions hold:

(A(x_0)) The cdf, $F^{(A_1)}$, of A_1 is continuous on $[x_0, \infty)$.

(B) (i) As $n \rightarrow \infty$, we have $l_n \rightarrow \infty$, $l_n = o(r_n)$, $r_n = o((nv_n)^{1/2})$, $r_n v_n \rightarrow 0$;

(ii) $\beta_{n,l_n} n/r_n \rightarrow 0$ as $n \rightarrow \infty$ and $\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} \beta_{n,m} = 0$.

Condition (B) pose restrictions on the rate at which v_n tends to 0 and thus on the rate at which u_n tends to ∞ . Sufficient conditions to ensure that a Markov chain is β -mixing can be found in Doukhan (1995, Section 2.4). Usually, $\beta_{n,k} = O(\eta^k)$ for some $\eta \in (0, 1)$ and one may choose $l_n = O(\log n)$, and (B) is fulfilled for a suitably chosen r_n if $(\log n)^2/n = o(v_n)$ and $v_n = o(1/\log n)$.

(C) For all $k \in \{0, \dots, r_n\}$ there exists

$$s_n(k) \geq \mathbb{P}[X_k > u_n | X_0 > u_n]$$

such that $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} s_n(k) < \infty$.

Typically $s_n(k)$ will be of the form $b_n + c_k$ with $b_n = o(1/r_n)$ and $\sum_{k=1}^{\infty} c_k < \infty$. The interchangeability of the limit and the sum is then automatically fulfilled. For stochastic recurrence equations (Section 2.5.2), conditions (B) and (C) are verified in Example 2.3 below.

Under these conditions, one can prove the asymptotic normality of relevant generalized tail array sums (see Proposition 2.11 below) and thus the joint asymptotic normality of the forward and the backward estimator of $F^{(A_1)}$ centered by their expectation. However, additional conditions are needed to ensure that their bias is asymptotically negligible:

$$\sup_{x \in [x_0, \infty)} \left| \mathbb{P} \left[\frac{X_1}{X_0} \leq x \mid X_0 > u_n \right] - F^{(A_1)}(x) \right| = o((nv_n)^{-1/2}), \quad (2.28)$$

$$\sup_{y \in [y_0, \infty)} \left| \mathbb{E} \left[\left(\frac{X_{-1}}{X_0} \right)^\alpha \mathbf{1}(X_0/X_1 > y) \mid X_0 > u_n \right] - \bar{F}^{(A_1)}(y) \right| = o((nv_n)^{-1/2}). \quad (2.29)$$

Here $\bar{F}^{(A_1)} := 1 - F^{(A_1)}$ denotes the survival function of A_1 (and hence of Θ_1). These conditions are fulfilled if nv_n tends to ∞ sufficiently slowly, because by definition of the spectral tail process and by (2.18), the left-hand sides in (2.28)–(2.29) tend to 0 if $F^{(A_1)}$ is continuous on $[x_0, \infty)$.

Theorem 2.3. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary, regularly varying process with a Markov spectral tail chain. If $(A(x_0))$, (B), (C), (2.28) and (2.29) are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0, \infty) \cap (0, \infty)$, then*

$$(nv_n)^{1/2} \begin{pmatrix} (\hat{F}_n^{(f, A_1)}(x) - F^{(A_1)}(x))_{x \in [x_0, \infty)} \\ (\hat{F}_n^{(b, A_1)}(y) - F^{(A_1)}(y))_{y \in [y_0, \infty)} \end{pmatrix} \rightsquigarrow \begin{pmatrix} (Z^{(f, A_1)}(x))_{x \in [x_0, \infty)} \\ (Z^{(b, A_1)}(y))_{y \in [y_0, \infty)} \end{pmatrix}, \quad n \rightarrow \infty, \quad (2.30)$$

where the limit is a centered Gaussian process whose covariance function is given by

$$\begin{aligned} \text{cov} \left(Z^{(f, A_1)}(x), Z^{(f, A_1)}(y) \right) &= \bar{F}^{(A_1)}(\max(x, y)) - \bar{F}^{(A_1)}(x) \bar{F}^{(A_1)}(y), \\ \text{cov} \left(Z^{(b, A_1)}(x), Z^{(b, A_1)}(y) \right) &= \mathbb{E} [\Theta_1^{-1} \mathbf{1}(\Theta_1 > \max(x, y))] - \bar{F}^{(A_1)}(x) \bar{F}^{(A_1)}(y), \end{aligned}$$

and

$$\begin{aligned} \text{cov} \left(Z^{(f, A_1)}(x), Z^{(b, A_1)}(y) \right) &= \sum_{k=1}^{\infty} (\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1) \\ &\quad - \bar{F}^{(A_1)}(x) \sum_{k=1}^{\infty} \mathbb{E} [(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)] \\ &\quad - \bar{F}^{(A_1)}(y) \sum_{k=1}^{\infty} \mathbb{P} [\Theta_1 > x, Y_k > 1] \\ &\quad + \bar{F}^{(A_1)}(x) \bar{F}^{(A_1)}(y) \sum_{k=1}^{\infty} \mathbb{P} [Y_k > 1]. \end{aligned}$$

Remark 2.4. For $x \geq 1$, we have

$$\begin{aligned} \text{var}(Z^{(b,A_1)}(x)) &= \mathbb{E} [\Theta_1^{-1} \mathbf{1}(\Theta_1 > x)] - (\bar{F}^{(A_1)}(x))^2 \\ &< \mathbb{P} [\Theta_1 > x] - (\bar{F}^{(A_1)}(x))^2 \\ &= \text{var}(Z^{(f,A_1)}(x)), \end{aligned}$$

provided $\bar{F}^{(A_1)}(x) = \mathbb{P} [A_1 > x] = \mathbb{P} [\Theta_1 > x] > 0$. Hence, for such x , when the tail index α is known, the backward estimator is asymptotically more efficient than the forward estimator.

Remark 2.5. While it is not too restrictive to assume that the cdf of A_1 is continuous on $(0, \infty)$, often $\mathcal{L}(A_1)$ has positive mass at 0; see Example 2.2. In this case, one may prove a version of Theorem 2.3 where the first coordinate in (2.30) is replaced with

$$(nv_n)^{1/2} \left(w(x) (\hat{F}_n^{(f,A_1)}(x) - F^{(A_1)}(x)) \right)_{x \in [0, \infty)} \rightsquigarrow \left(w(x) Z^{(f,A_1)}(x) \right)_{x \in [0, \infty)}, \quad n \rightarrow \infty,$$

for a weight function $w(x) = h(F^{(A_1)}(x) - F^{(A_1)}(0))$ and any nondecreasing, continuous function h with $h(0) = 0$.

2.4.2 Unknown index of regular variation

In most applications, the index of regular variation, α , is unknown. In the definition of the backward estimator $\hat{F}_n^{(b,A_1)}(y)$, it must then be replaced with a suitable estimator. A popular estimator of α is the Hill-type estimator (2.25). More generally, one may consider estimators that can be written in the form

$$\hat{\alpha}_n = \frac{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}{\sum_{i=1}^n \tilde{\psi}(X_i/u_n) \mathbf{1}(X_i > u_n) + R_n} \quad (2.31)$$

with a remainder term $R_n = o_P((nv_n)^{1/2})$ and a suitable function $\tilde{\psi} : [0, \infty) \rightarrow [0, \infty)$ which is a.s. continuous w.r.t. $\mathcal{L}(Y_k)$ for all $k \in \mathbb{N} \cup \{0\}$ such that $\mathbb{E}[\tilde{\psi}(Y_0)] = 1/\alpha$ and $\tilde{\psi}(x) = 0$ for all $x \in [0, 1]$. Obviously, the Hill-type estimator is of this form with $\tilde{\psi}(x) = \log(x) \mathbf{1}(x > 1)$. Under weak dependence conditions, other well-known estimators like the maximum likelihood estimator in a generalized Pareto model examined by Smith (1987) and the moment estimator suggested by Dekkers et al. (1989) can be written in this way too; see Drees (1998a, Example 4.1) and Drees (1998b, Example 4.1) for similar results in the case of i.i.d. sequences.

Estimators of type (2.31) can be approximated by the ratio of the generalized tail array sums corresponding to the functions $\phi_1(y_{-1}, y_0, y_1) := \mathbf{1}(y_0 > 1)$ and $\psi(y_{-1}, y_0, y_1) := \tilde{\psi}(y_0) \mathbf{1}(y_0 > 1)$, respectively, and their asymptotic behavior can hence be derived from Theorem 2.3 of Drees and Rootzén (2010). To this end, we replace (C) with the following condition:

(C') For all $0 \leq k \leq r_n$ there exists

$$s_n(k) \geq \mathbb{E} \left[\max \left(\left| \tilde{\psi} \left(\frac{X_0}{u_n} \right) \right|, \mathbf{1}(X_0 > u_n) \right) \max \left(\left| \tilde{\psi} \left(\frac{X_k}{u_n} \right) \right|, \mathbf{1}(X_k > u_n) \right) \mid X_0 > u_n \right] \quad (2.32)$$

such that $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} s_n(k) < \infty$.

Moreover, there exists $\delta > 0$ such that

$$\sum_{k=1}^{r_n} \left(\mathbb{E} \left[\left| \tilde{\psi} \left(\frac{X_0}{u_n} \right) \tilde{\psi} \left(\frac{X_k}{u_n} \right) \right|^{1+\delta} \mid X_0 > u_n \right] \right)^{1/(1+\delta)} = O(1), \quad n \rightarrow \infty. \quad (2.33)$$

If $\tilde{\psi}$ is bounded, then (C') follows from condition (C), but in general it is more restrictive, though it can often be established by similar arguments. For example, condition (C') holds for the solutions to the stochastic recurrence equation studied in Example 2.3 and the Hill estimator, i.e., for $\tilde{\psi}(x) = \log(x) \mathbf{1}(x > 1)$.

The following result gives the asymptotic normality of $\hat{\alpha}_n$ centered at

$$\alpha_n := 1/\mathbb{E}[\tilde{\psi}(X_0/u_n) \mid X_0 > u_n]. \quad (2.34)$$

This quantity tends to α as $n \rightarrow \infty$ by the assumptions on the function $\tilde{\psi}$ and condition (C').

Lemma 2.6. *If $\hat{\alpha}_n$ is of the form (2.31) and if the conditions (B) and (C') hold, then*

$$(nv_n)^{1/2}(\hat{\alpha}_n - \alpha_n) = \alpha \tilde{Z}_n(\phi_1) - \alpha^2 \tilde{Z}_n(\psi) + o_P(1) \rightsquigarrow \alpha \tilde{Z}(\phi_1) - \alpha^2 \tilde{Z}(\psi), \quad n \rightarrow \infty,$$

for a centered Gaussian process $\tilde{Z}$ with covariance function given by (2.51).

Similarly as in (2.28) and (2.29), we need an extra condition to ensure that the bias of $\hat{\alpha}_n$ is asymptotically negligible:

$$|\mathbb{E}[\tilde{\psi}(X_0/u_n) \mid X_0 > u_n] - 1/\alpha| = o((nv_n)^{-1/2}), \quad n \rightarrow \infty. \quad (2.35)$$

Now we are ready to state the asymptotic normality of the backward estimator with estimated index α , i.e.,

$$\hat{F}_n^{(\hat{b}, A_1)}(x) := 1 - \frac{\sum_{i=1}^n \left(\frac{X_{i-1}}{X_i} \right)^{\hat{\alpha}_n} \mathbf{1}(X_i/X_{i-1} > x, X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}.$$

Theorem 2.7. *Suppose that the conditions of Corollary 2.3 and of Lemma 2.6 are fulfilled and that (2.35) holds. Then*

$$(nv_n)^{1/2} \begin{pmatrix} (\hat{F}_n^{(f, A_1)}(x) - F^{(A_1)}(x))_{x \in [x_0, \infty)} \\ (\hat{F}_n^{(\hat{b}, A_1)}(y) - F^{(A_1)}(y))_{y \in [y_0, \infty)} \end{pmatrix} \rightsquigarrow \begin{pmatrix} (Z^{(f, A_1)}(x))_{x \in [x_0, \infty)} \\ (Z^{(\hat{b}, A_1)}(y))_{y \in [y_0, \infty)} \end{pmatrix}, \quad n \rightarrow \infty, \quad (2.36)$$

with

$$\begin{aligned} Z^{(f,A_1)}(x) &= \tilde{Z}(\phi_{2,x}) - \bar{F}^{(A_1)}(x)\tilde{Z}(\phi_1), \\ Z^{(\hat{b},A_1)}(y) &= \tilde{Z}(\phi_{3,y}) - \bar{F}^{(A_1)}(y)\tilde{Z}(\phi_1) + (\alpha^2 \tilde{Z}(\psi) - \alpha \tilde{Z}(\phi_1)) \mathbb{E} [\log(\Theta_1) \mathbf{1}(\Theta_1 > y)] \end{aligned}$$

for a centered Gaussian process $\tilde{Z}$ with covariance function given by (2.51).

The covariance function of the limiting process can be calculated in the same way as in the proof of Theorem 2.3. In general, the resulting expressions will involve sums over all $k \in \mathbb{N}$. Moreover, it is no longer guaranteed that the backward estimator of $F^{(A_1)}(y)$ at $y > 1$ has a smaller variance than the forward estimator.

Remark 2.8. For Markovian time series which are not necessarily positive, the forward and backward estimators of $F^{(A_1)}$ and $F^{(B_1)}$ can be represented in terms of generalized tail array sums constructed from

$$X_{n,i} := \frac{(X_{i-1}, X_i, X_{i+1})}{u_n} \mathbf{1}(|X_i| > u_n).$$

When $x < 0$, for example, the backward estimator $\hat{F}_n^{(b,B_1)}(x)$ equals the ratio of the generalized tail array sums pertaining to

$$\begin{aligned} \phi_{4,x}(y_{-1}, y_0, y_1) &:= (-y_{-1}/y_0)^\alpha \mathbf{1}(y_0/y_{-1} \leq x, y_0 > 1), \\ \phi_5(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_0 < -1). \end{aligned}$$

Hence their limit processes can be obtained by the same methods as in the case $X_t > 0$ under obvious analogues to the conditions (A(x_0)), (B) and (C) with $v_n := \mathbb{P}[|X_0| > u_n]$.

2.5 Examples

We first show how, for regularly varying Markov chains, the distributions of $A_{\pm 1}$ and $B_{\pm 1}$ can be calculated from the copula of (X_0, X_1) if this copula satisfies certain regularity conditions. Next, we focus on solutions to stochastic recurrence equations, for which the Markov spectral tail chain exists and the distributions of $A_{\pm 1}$ and $B_{\pm 1}$, as well as p , admit a simple representation.

2.5.1 Copula-based Markov processes

For stationary Markov processes, the joint distribution of $(X_t)_{t \in \mathbb{Z}}$ is determined by the law of (X_0, X_1) . In view of Sklar's theorem, the latter is determined by a univariate marginal distribution function, G , and a copula, C , via

$$\mathbb{P}[X_0 \leq x_0, X_1 \leq x_1] = C(G(x_0), G(x_1)), \quad (x_0, x_1) \in \mathbb{R}^2. \quad (2.37)$$

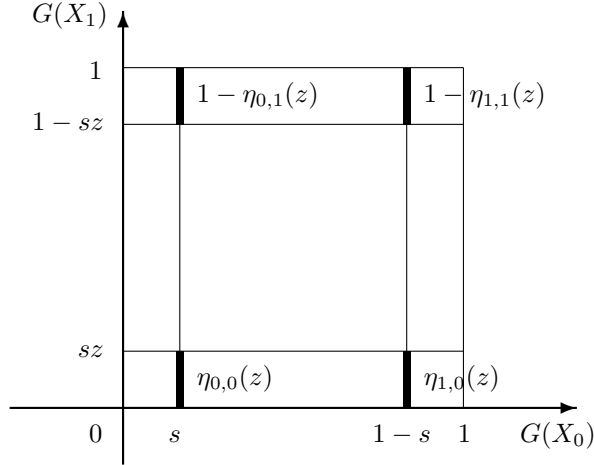


FIGURE 2.1: Viewing the functions $\eta_{0,0}$, $\eta_{0,1}$, $\eta_{1,0}$ and $\eta_{1,1}$ in terms of the conditional distribution of $G(X_1)$ conditionally on $G(X_0) = s$ or on $G(X_0) = 1 - s$.

Copula-based Markov models arise if the law of the Markov process is described by specifying G and C directly (Chen and Fan, 2006; Chen et al., 2009).

The result below links the distributions of A_1 and B_1 in (2.9) and (2.11) to the margin G and the copula C . The result is framed in terms of the limit behaviour of $\dot{C}_1(u, v) = \partial C(u, v)/\partial u$ as u tends to 0 or 1. The function $\dot{C}_1$ is related to the conditional distribution of X_1 given X_0 ; see (2.48) below. For $z \geq 0$, consider the limits (whose existence is an assumption)

$$\lim_{s \searrow 0} \dot{C}_1(1 - s, 1 - sz) = \eta_{1,1}(z), \quad (2.38)$$

$$\lim_{s \searrow 0} \dot{C}_1(1 - s, sz) = \eta_{1,0}(z), \quad (2.39)$$

and

$$\lim_{s \searrow 0} \dot{C}_1(s, 1 - sz) = \eta_{0,1}(z), \quad (2.40)$$

$$\lim_{s \searrow 0} \dot{C}_1(s, sz) = \eta_{0,0}(z), \quad (2.41)$$

covering the four corners of the unit square; see Figure 2.1.

Proposition 2.9. *Let the distribution of (X_0, X_1) be given by (2.37), where X_0 has a Lebesgue density and satisfies the regular variation condition (2.2) and the tail-balance condition (2.5). Assume that C admits a continuous first-order partial derivative $\dot{C}_1$ on $(0, 1) \times [0, 1]$. If $p > 0$*

and if the limits in (2.38) and (2.39) exist and are continuous on $[0, \infty)$, then (2.9) holds and

$$P[A_1 \leq x] = \begin{cases} \eta_{1,1}(x^{-\alpha}) & \text{if } x > 0, \\ \eta_{1,0}(\frac{1-p}{p} |x|^{-\alpha}) & \text{if } x < 0. \end{cases} \quad (2.42)$$

Similarly, if $p < 1$ and if the limits in (2.40) and (2.41) exist and are continuous on $[0, \infty)$, then (2.11) holds and

$$P[B_1 \leq x] = \begin{cases} 1 - \eta_{0,0}(x^{-\alpha}) & \text{if } x > 0, \\ 1 - \eta_{0,1}(\frac{p}{1-p} |x|^{-\alpha}) & \text{if } x < 0. \end{cases} \quad (2.43)$$

Similarly, the distributions of A_{-1} and B_{-1} in (2.10) and (2.12) can be obtained via the limit behaviour of $\dot{C}_2(u, v) = \partial C(u, v)/\partial v$ as v tends to 0 or 1. Some examples are worked out in Appendix 2.8.1, whereas the proof of Proposition 2.9 is given in Appendix 2.8.2.

2.5.2 Stochastic recurrence equations

The stochastic recurrence equation

$$X_t = C_t X_{t-1} + D_t, \quad t \in \mathbb{Z}, \quad (2.44)$$

received some attention in time series analysis and extreme value theory. We focus on the case where (C_t, D_t) , $t \in \mathbb{Z}$, is an i.i.d. $\mathbb{R}^2$ -valued sequence. Provided that $-\infty \leq E[\log |C_1|] < 0$ and $E[\log^+ |D_1|] < \infty$, where $\log^+ x = \max(\log x, 0)$, there exists a unique strictly stationary causal solution to (2.44) (Basrak et al., 2002, Corollary 2.2).

Results on regular variation of X_0 were first developed by Kesten (1973). His Theorem 5 states that if there exists $\alpha > 0$ such that $E[|C_1|^\alpha] = 1$, $E[|C_1|^\alpha \log^+ |C_1|] < \infty$ and $E[|D_1|^\alpha] < \infty$ and if some other conditions are satisfied, then X_0 is regularly varying; more specifically,

$$\left. \begin{aligned} \Pr(X_0 > x) &= c_+ x^{-\alpha} (1 + o(1)) \\ \Pr(X_0 \leq -x) &= c_- x^{-\alpha} (1 + o(1)) \end{aligned} \right\}, \quad x \rightarrow \infty, \quad (2.45)$$

for constants $c_+, c_- \geq 0$ such that $c_+ + c_- > 0$. This result was extended by Goldie (1991, Lemma 2.2 and Theorem 4.1), who gave explicit expressions for c_+ and c_- . Regular variation of X_0 and iteration of (2.44) gives joint regular variation of $(X_t)_{t \in \mathbb{Z}}$ with index α .

The forward spectral tail process $(\Theta_t)_{t \geq 0}$ of $(X_t)_{t \in \mathbb{Z}}$ admits the representation

$$\Theta_t = \Theta_0 \prod_{h=1}^t \tilde{C}_h, \quad t \in \mathbb{N},$$

where $\tilde{C}_t$, $t \in \mathbb{N}$, are i.i.d. random variables with the same distribution as C_1 which are independent of Θ_0 (Janßen and Segers, 2014, Example 6.1). Hence, $(\Theta_t)_{t \geq 0}$ becomes a Markov spectral tail chain satisfying (2.13) with $\mathcal{L}(A_1) = \mathcal{L}(B_1) = \mathcal{L}(C_1)$. Moreover, by Goldie (1991, Theorem 4.1), if $\mathbb{P}[C_1 < 0] > 0$ then necessarily $c_+ = c_-$ and therefore $p = \mathbb{P}[\Theta_0 = +1] = 1/2$. This property also follows from the following result on more general tail spectral processes; its proof is given in Appendix 2.8.2.

Lemma 2.10. *Let $(\Theta_t)_{t \in \mathbb{Z}}$ be the spectral tail process of an α -regularly varying stationary time series $(X_t)_{t \in \mathbb{Z}}$. Suppose the following two conditions hold:*

- (i) $\mathbb{E}[|\Theta_1|^\alpha] = 1$;
- (ii) Θ_1/Θ_0 and Θ_0 are independent.

Then Θ_{-1}/Θ_0 and Θ_0 are independent too. Moreover, $\mathbb{P}[\Theta_1/\Theta_0 < 0] > 0$ implies $p = \mathbb{P}[\Theta_0 = 1] = 1/2$.

Lemma 2.10 applies to the solution $(X_t)_{t \in \mathbb{Z}}$ of the stochastic recurrence equation (2.44). Above, we had already seen that $\mathcal{L}(A_1) = \mathcal{L}(B_1) = \mathcal{L}(C_1)$ and $\mathbb{E}[|C_1|^\alpha] = 1$. Hence, the spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$ satisfies the two conditions in Lemma 2.10. We obtain that Θ_{-1}/Θ_0 is independent of Θ_0 too. But since the backward spectral tail process admits the representation in (2.14), we conclude that $\mathcal{L}(A_{-1}) = \mathcal{L}(B_{-1})$. If $p = 0$ or $p = 1$, the previous statements simplify, since A_1 and A_{-1} in (2.9)–(2.10) are defined only if $p > 0$ whereas B_1 and B_{-1} in (2.11)–(2.12) are defined only if $p < 1$. For $p \in \{0, 1\}$, the law of Θ_0 is degenerate at -1 or at 1 , respectively, so that the statements on independence are trivially satisfied.

2.6 Monte Carlo simulations

To compare the finite-sample performance of the forward, backward, and mixture estimators, we conducted a Monte Carlo simulation study. We use both approaches discussed at the end of Section 2.3 for dealing with the problem that α is unknown.

Pseudo-random samples are generated from two different time series models.

- For the copula Markov model (Section 2.5.1), we choose the symmetric t-distribution with $\nu_1 = 2$ degrees of freedom as the margin G and the t-copula $C_{\nu_2, \rho}^t(u, v)$ with $\nu_2 = 2.5$ and $\rho = 0.2$ to model the temporal dependence structure. Hence, $\mathbb{P}[\Theta_0 = 1] = \mathbb{P}[\Theta_0 = -1] = 1/2$ and the index of regular variation is equal to 2. The distribution function of A_1 is calculated in Example 2.1. The simulation algorithm is described in Chen et al. (2009, Section 6.2).

- For the stochastic recurrence equation (Section 2.5.2), we let C_t and D_t be independent $N(1/3, 8/9)$ and $N(-10, 1)$ random variables, respectively. This choice ensures that $E[C_t^2] = 1$ and that the sufficient conditions of Theorem 5 in Kesten (1973) hold. As a consequence, a stationary solution $(X_t)_{t \in \mathbb{Z}}$ exists which is regularly varying with index $\alpha = 2$.

For each model, we generate time series $(X_i)_{1 \leq i \leq n}$ of length $n = 2000$ and set the threshold u_n to be the 97.5% quantile of the absolute values of the sample, i.e., we use 50 extremes for estimation. Based on 1000 Monte Carlo repetitions, we estimate the bias, the standard deviation (SD) and the root mean squared error (RMSE) of all estimators under consideration. In the first part of the supplementary material we address in more detail the bias problem that can arise when approximating the limiting distribution of the tail process by the finite level conditional distribution.

Note that both models are symmetric in the sense that $\mathcal{L}(A_1) = \mathcal{L}(B_1)$. Moreover, the estimators of the cdf's of A_1 and B_1 are identically distributed for the copula model, and they behave similarly for the solutions to the stochastic recurrence equation. Therefore, we report the results only for A_1 .

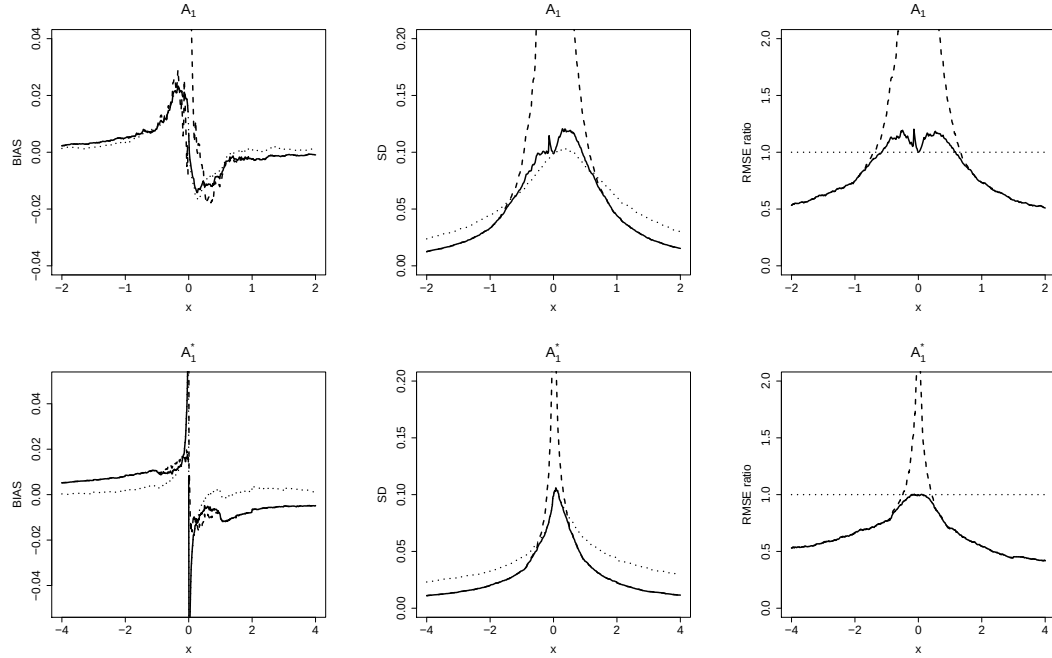


FIGURE 2.2: Simulation results for the copula Markov model. Top: Estimation of the cdf of A_1 using the plug-in estimator for α . Bottom: Estimation of the cdf of $A_1^* = \text{sign}(A_1)|A_1|^\alpha$ with the signed rank transformation. Left: bias; middle: standard deviation; right: root mean squared error ratio with respect to the forward estimator. Solid line: mixture estimator; dashed line: backward estimator; dotted line: forward estimator.

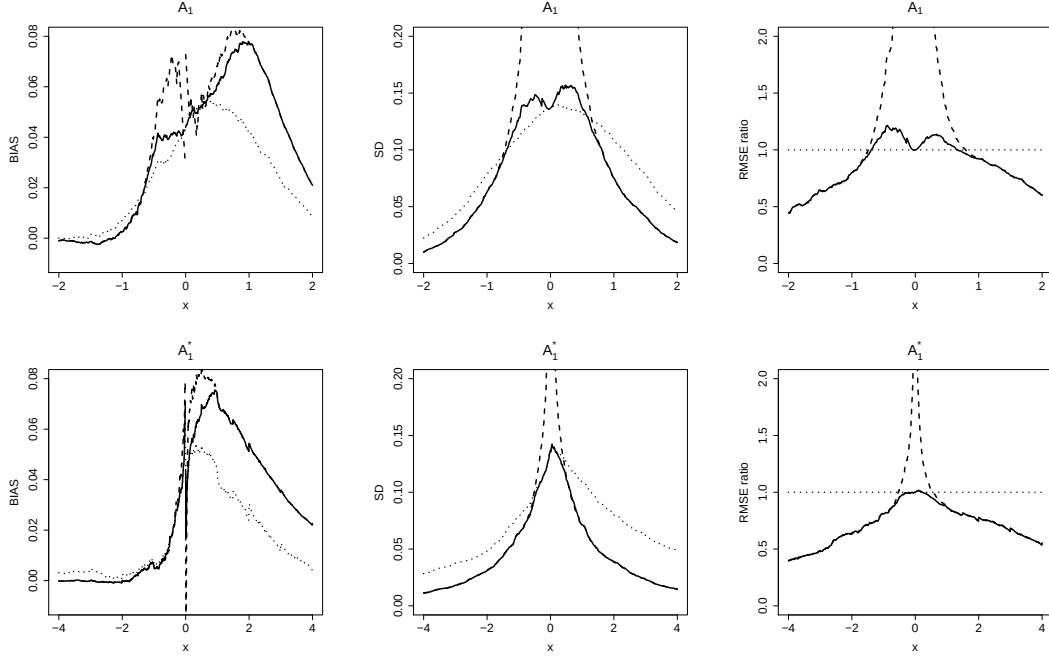


FIGURE 2.3: Simulation results for the stochastic recurrence equations. Top: Estimation of the cdf of A_1 using the plug-in estimator for α . Bottom: Estimation of the cdf of $A_1^* = \text{sign}(A_1)|A_1|^\alpha$ with the signed rank transformation. Left: bias; middle: standard deviation; right: root mean squared error ratio with respect to the forward estimator. Solid line: mixture estimator; dashed line: backward estimator; dotted line: forward estimator.

Figure 2.2 shows the results for the estimators of the cdf's of A_1 and A_1^* in the copula Markov model. For this model, $\text{Bias}(\hat{p}) = 0.001$, $\text{SD}(\hat{p}) = 0.08$, and $\text{RMSE}(\hat{p}) = 0.08$. In addition, for the simulations with estimated α , we obtain $\text{Bias}(\hat{\alpha}) = 0.077$, $\text{SD}(\hat{\alpha}) = 0.407$, and $\text{RMSE}(\hat{\alpha}) = 0.414$. In the top row, the Hill-type estimator (2.25) is used, whereas in the bottom row, we apply the signed rank transformation (2.26). The plots on the left and in the middle show the bias and the standard deviation of the forward estimator as a dotted line, of the backward estimator as a dashed line and of the mixture estimator as a solid line. The plots on the right-hand side display the relative RMSE of the backward and the mixture estimator w.r.t. the forward estimator.

As expected, the backward estimator outperforms the forward estimator for $|x| \geq 1$, whereas for arguments x close to 0 the performance of the backward estimator is poor and it is the forward estimator that has the lower RMSE. The mixture estimator performs much better than the backward estimator for small $|x|$, while for large values of $|x|$ its RMSE is similar to the one of the backward estimator. Indeed, in the approach using rank transforms, the RMSE of the mixture estimator is never much larger than that of the forward estimator, while the mixture estimator is almost twice as efficient as the forward estimator for $|x| > 2$. Hence, it clearly outperforms both the forward and the backward estimator in this approach. In contrast, the

mixture estimator performs worse than the forward estimator for small values of $|x|$ if α is replaced with the Hill-type estimator.

Figure 2.3 shows the analogous results for the solutions of the stochastic recurrence equation. We obtain $\text{Bias}(\hat{p}) = -0.209$, $\text{SD}(\hat{p}) = 0.083$, $\text{RMSE}(\hat{p}) = 0.225$, and $\text{Bias}(\hat{\alpha}) = 0.173$, $\text{SD}(\hat{\alpha}) = 0.526$, $\text{RMSE}(\hat{\alpha}) = 0.553$. By and large, the relative performance of all estimators is similar to the one in the copula Markov model. The relative efficiency of the mixture estimator with respect to the forward estimator is a bit higher for small values of $|x|$, while it is slightly worse for larger values. Overall, the absolute estimation errors of all estimators are about 20 to 30% larger.

2.7 Case study

The spectral tail process of a heavy tailed time series conveys important information on its serial extremal dependence. Such extremal dependence typically arises e.g. in financial time series which exhibit clustering of extremes. By estimating the joint distribution of (Θ_0, Θ_1) and (Θ_0, Θ_{-1}) , we gain insight into short-range dependence between extremes of consecutive observations, covering both lower and upper tails.

An interesting question is whether the distribution of a time series remains unaffected by reversing the direction of time. It is well understood how to test for such time reversibility regarding the bulk of the distribution (Beare and Seo, 2014; Chen et al., 2000). Clearly, if a time series does not look time reversible, it should not be modelled with processes having this property, such as, e.g., stationary Gaussian processes. Financial time series do not have this feature in general (Chen and Kuan, 2002). This, however, does not imply that there can be no time reversibility at extreme levels. Formally, such extremal time reversibility would mean that $\mathcal{L}(\Theta_0, \Theta_1, \dots, \Theta_t) = \mathcal{L}(\Theta_0, \Theta_{-1}, \dots, \Theta_{-t})$. In particular, it implies that $\mathcal{L}(\Theta_0, \Theta_1) = \mathcal{L}(\Theta_0, \Theta_{-1})$ and, equivalently, $\mathcal{L}(A_1) = \mathcal{L}(A_{-1})$ and $\mathcal{L}(B_1) = \mathcal{L}(B_{-1})$, or $\mathcal{L}(A_1^*) = \mathcal{L}(A_{-1}^*)$ and $\mathcal{L}(B_1^*) = \mathcal{L}(B_{-1}^*)$. Comparison of the estimated distribution functions may be used to reject the extremal time reversibility hypothesis, and thus can serve as a diagnostic tool for model selection.

Note that under Markov assumption $\mathcal{L}(\Theta_0, \Theta_1, \dots, \Theta_t) = \mathcal{L}(\Theta_0, \Theta_{-1}, \dots, \Theta_{-t})$ is equivalent to $\mathcal{L}(\Theta_0, \Theta_1) = \mathcal{L}(\Theta_0, \Theta_{-1})$, and hence inference about extremal time reversibility can be based only on the pairwise comparison, i.e., $\mathcal{L}(A_1^*) = \mathcal{L}(A_{-1}^*)$ and $\mathcal{L}(B_1^*) = \mathcal{L}(B_{-1}^*)$. However, Markovianity is rarely a realistic assumption for financial time series.

We analyze daily log-returns of Google and UBS stock prices between 2005-01-03 and 2013-12-31 (taken from www.google.com/finance), leading to 2279 observations for Google and 2280 for UBS. The thresholds are set at the 95% quantiles, giving 144 extremes. This choice is based on the stability plots which can be found in the second part of the supplementary material.

For Google log-returns we obtain 53 positive extremes and 61 negative extremes, whereas for UBS it is 50 and 64 positive and negative extremes respectively. The estimated index of regular variation is equal to 2.88 for Google and 2.51 for UBS.

We discuss the results jointly for both stocks, as their extremal dependence structures exhibit similar patterns. The two data series are presented in Figure 2.4. The three rows show the daily closing prices (top), the daily log-returns X_i (middle) and the pertaining rank transformed log-returns $X_{n,i}^*$ as in (2.26) (bottom).

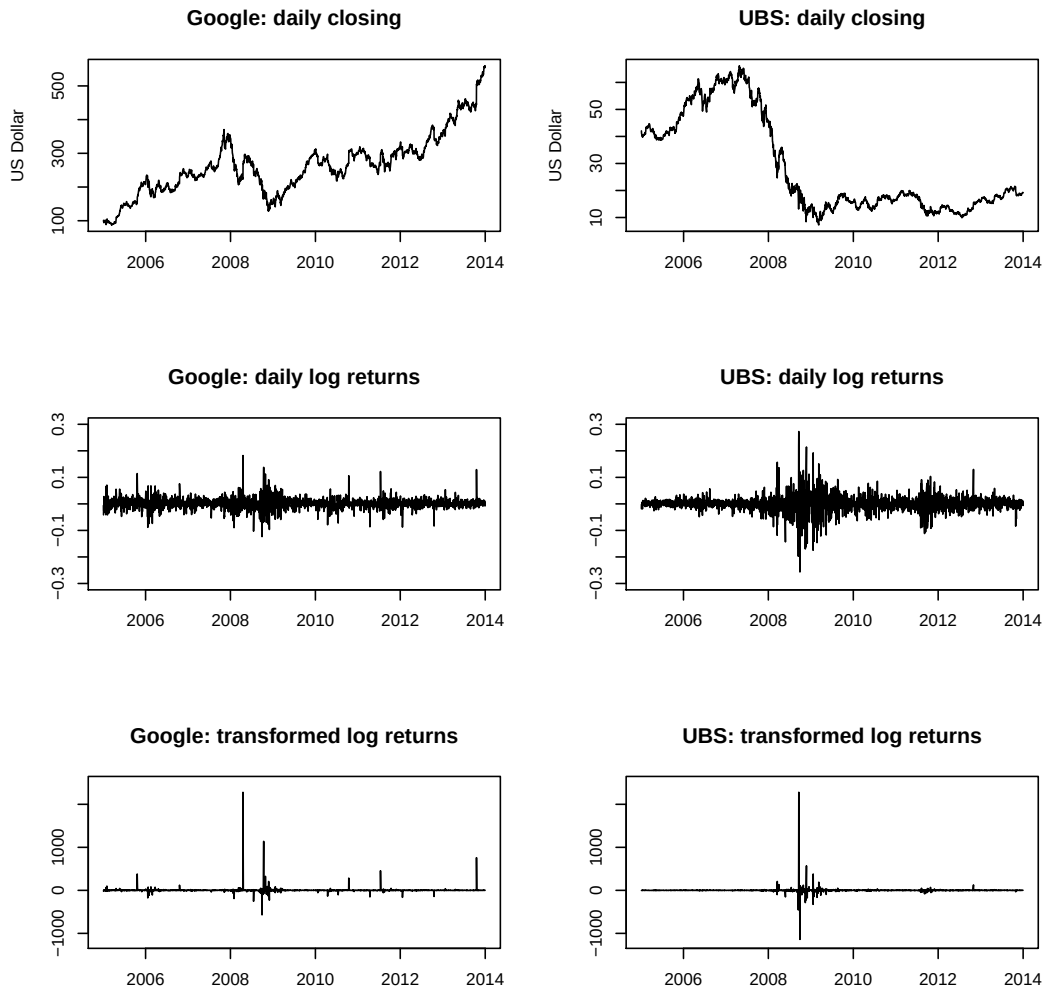


FIGURE 2.4: Google and UBS daily stock prices, daily log-returns and transformed log-returns.

The distribution functions of A_{-1}^* and B_{-1}^* were estimated using the reversed time series. For A_1^* , A_{-1}^* , B_1^* , and B_{-1}^* , we applied a monotonized version of the mixture estimator (2.24), which otherwise is not necessarily monotone. The monotonized version is defined by the smallest

increasing function larger than or equal to the mixture estimator, for $x \geq 0$, and the biggest increasing function smaller than or equal to the mixture estimator, for $x < 0$. We display the estimated distribution functions of A_1^* , A_{-1}^* , B_1^* , and B_{-1}^* for the rank transformed log-returns based on the monotonized mixture estimator in Figure 2.5.

Since the cdf's of A_1^* and A_{-1}^* as well as the cdf's of B_1^* and B_{-1}^* apparently differ on the negative real line, the extreme values of the log-returns exhibit no time reversibility. Moreover, most of the probability mass of A_1^* and B_{-1}^* is concentrated near the origin. This hints at asymptotic independence of X_t^+ and X_{t+1} , and of X_t and X_{t+1}^- . In contrast, the laws of A_{-1}^* and B_1^* put considerable mass on the negative half-line. This indicates that with asymptotically non-negligible probability a large loss is succeeded by a large gain. Such an event can be interpreted as a correction to an overreaction by the stock market. Indeed, the existence of long-term and short-term overreaction behaviour in various stock markets is well documented (De Bondt and Thaler, 1985; Bowman and Iverson, 1998; Norli, 2010).

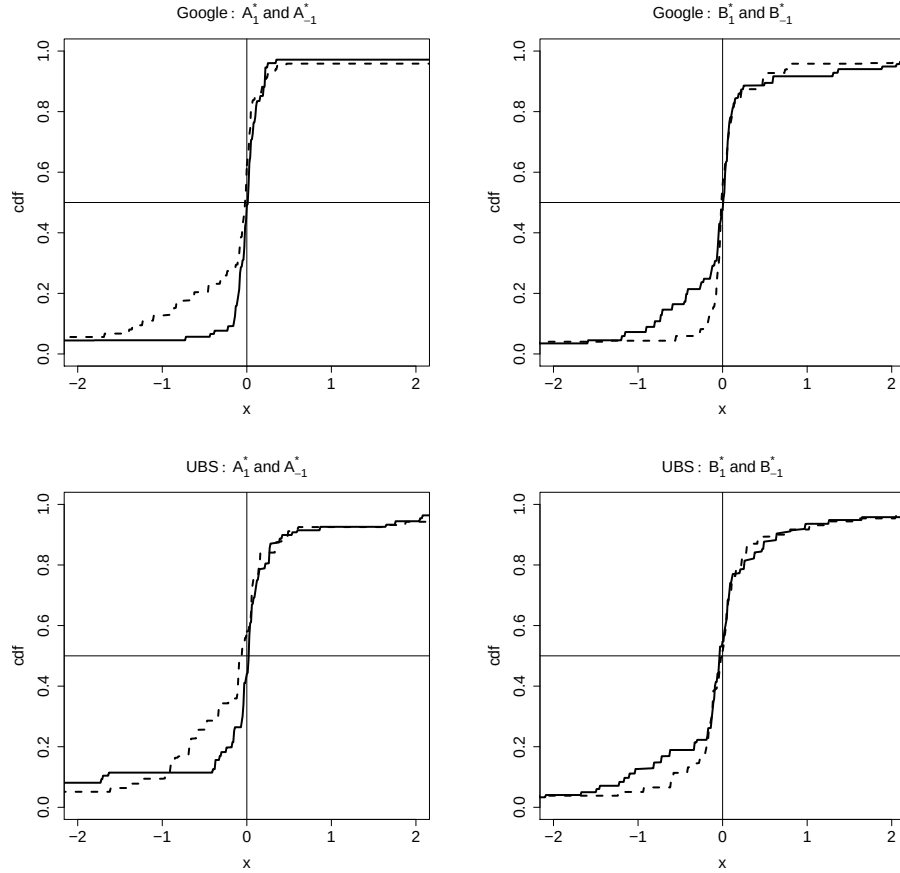


FIGURE 2.5: Estimated distribution functions of A_1^* and A_{-1}^* (left) and of B_1^* and B_{-1}^* (right) based on the monotonized mixture estimator for transformed log-returns on Google stock prices (top) and UBS stock prices (bottom). Solid lines: A_1^* and B_1^* ; dashed lines: A_{-1}^* and B_{-1}^* .

2.8 Appendix A - Proofs and examples

2.8.1 Examples

Example 2.1 (The t -copula). *In the context of the copula Markov model (Section 2.5.1), consider the bivariate t -copula defined by*

$$C_{\nu,\rho}^t(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} dx dy,$$

where $\rho \in (-1, 1)$, t_{ν} is the cdf of the univariate t -distribution with $\nu > 0$ degrees of freedom and t_{ν}^{-1} is the corresponding quantile function. Contrary to the Gaussian copula, the t -copula allows for an asymptotically non-negligible probability of joint extremes. The t -copula is exchangeable and radially symmetric. The partial derivative of $C_{\nu,\rho}^t(u, v)$ with respect to the first coordinate equals (Demarta and McNeil, 2005, proof of Proposition 4)

$$\frac{\partial}{\partial u} C_{\nu,\rho}^t(u, v) = t_{\nu+1} \left(\left\{ \frac{t_{\nu}^{-1}(v)}{t_{\nu}^{-1}(u)} - \rho \right\} \left(\frac{\nu+1}{1-\rho^2} \right)^{1/2} \left\{ 1 + \frac{\nu}{(t_{\nu}^{-1}(u))^2} \right\}^{-1/2} \text{sign}(t_{\nu}^{-1}(u)) \right). \quad (2.46)$$

The limits (2.38) and (2.39) can be calculated using symmetry of the t -distribution and the regular variation of t_{ν} with index ν . Under the assumptions of Proposition 2.9, we obtain

$$\mathbb{P}[A_1 \leq x] = \begin{cases} t_{\nu+1} \left((x^{\alpha/\nu} - \rho) \left(\frac{\nu+1}{1-\rho^2} \right)^{1/2} \right) & \text{if } x \geq 0, \\ t_{\nu+1} \left(\left(-(\frac{1-p}{p} |x|^{-\alpha})^{-1/\nu} - \rho \right) \left(\frac{\nu+1}{1-\rho^2} \right)^{1/2} \right) & \text{if } x < 0. \end{cases} \quad (2.47)$$

As the t -copula is exchangeable, we find $\mathcal{L}(A_1) = \mathcal{L}(A_{-1})$ and $\mathcal{L}(B_1) = \mathcal{L}(B_{-1})$. Moreover, by radial symmetry we have $\eta_{1,1} = 1 - \eta_{0,0}$ and $\eta_{1,0} = 1 - \eta_{0,1}$, linking up the laws of A_1 and B_1 through (2.42) and (2.43). In particular, if $p = 1/2$, then $\mathcal{L}(A_1) = \mathcal{L}(B_1)$.

Example 2.2 (Extreme-value copulas). *A bivariate copula $C(u, v)$ is an extreme value copula if and only if it admits the representation*

$$C(u, v) = \exp \left\{ D \left(\frac{\log v}{\log(uv)} \right) \log(uv) \right\}, \quad (u, v) \in (0, 1]^2 \setminus \{(1, 1)\}.$$

The Pickands dependence function $D : [0, 1] \rightarrow [1/2, 1]$ is convex and satisfies $\max(w, 1-w) \leq f(w) \leq 1$ for all $w \in [0, 1]$. See Gudendorf and Segers (2010) for a survey of extreme-value copulas.

If D is continuously differentiable with derivative D' , then, under the conditions of Proposition 2.9,

$$\mathbb{P}[A_1 \leq x] = \left\{ D\left(\frac{1}{x^\alpha + 1}\right) - \frac{1}{x^\alpha + 1} D'\left(\frac{1}{x^\alpha + 1}\right) \right\} \mathbf{1}(x \geq 0),$$

whereas B_1 and B_{-1} are degenerate at 0. In particular, $\mathbb{P}[A_1 = 0] = 1 - D'(0)$. The law of A_{-1} has the same form as the one of A_1 upon replacing D by $w \mapsto D(1 - w)$. If $D(w) = D(1 - w)$ then C is exchangeable and $\mathcal{L}(A_{-1}) = \mathcal{L}(A_1)$.

The following two parametric families are well known:

- For the asymmetric logistic model (Tawn, 1988) with parameters $\theta \geq 1$ and $\psi_1, \psi_2 \in (0, 1]$, we have

$$\begin{aligned} D(w) &= (1 - \psi_1)w + (1 - \psi_2)(1 - w) + \{(\psi_1 w)^\theta + (\psi_2(1 - w))^\theta\}^{1/\theta}, \\ \mathbb{P}[A_1 \leq x] &= 1 - \psi_2 + \psi_2 \left\{ 1 + (\psi_1/(\psi_2 x^\alpha))^\theta \right\}^{(1-\theta)/\theta}, \quad x \geq 0. \end{aligned}$$

The special case $\psi_1 = \psi_2 = 1$ yields the Gumbel–Hougaard copula.

- For the asymmetric negative logistic model (Joe, 1990) with parameters $\theta > 0$ and $\psi_1, \psi_2 \in (0, 1]$, we have

$$\begin{aligned} D(w) &= 1 - \{(\psi_1 w)^{-\theta} + (\psi_2(1 - w))^{-\theta}\}^{-1/\theta}, \\ \mathbb{P}[A_1 \leq x] &= 1 - \psi_2 \left\{ 1 + (\psi_2 x^\alpha / \psi_1)^\theta \right\}^{-(1+\theta)/\theta}, \quad x \geq 0. \end{aligned}$$

The special case $\psi_1 = \psi_2 = 1$ yields the Galambos copula.

Example 2.3 (Stochastic recurrence equations). We will show that stationary solutions to the stochastic recurrence equation (2.44) with $(C_t, D_t) \in [0, \infty)^2$ satisfy the conditions (B) and (C) in Section 2.4.1. Asymptotic normality of the forward and backward estimators (Theorem 2.3) follows if the cdf of $\mathcal{L}(A_1) = \mathcal{L}(C_1)$ is continuous on $[x_0, \infty)$.

We assume that the conditions in Kesten (1973) are fulfilled (see Section 2.2). Then $(X_t)_{t \in \mathbb{Z}}$ is geometrically β -mixing, i.e., there exist constants $\eta \in (0, 1)$ and $\tau > 0$ such that $\beta_{n,k} \leq \tau \eta^k$ (Doukhan, 1995, Corollary 2.4.1). Therefore, condition (B) is satisfied with $l_n = 2 \log n / |\log \eta|$ and suitably chosen $r_n = o(\min\{(nv_n)^{1/2}, v_n^{-1}\})$, provided $v_n = o(1/\log n)$ and $(\log n)^2/n = o(v_n)$.

To establish condition (C), let $\Pi_{i+1,j} := \prod_{k=i+1}^j C_k$ and $V_k := \sum_{j=1}^k \Pi_{j+1,k} D_j$. Iterating (2.44) yields $X_k = V_k + \Pi_{1,k} X_0$. By independence of $(V_k, \Pi_{1,k})$ and X_0 , one has

$$\begin{aligned} \mathbb{P}[X_k > u_n \mid X_0 > u_n] &\leq v_n^{-1} \mathbb{P}[X_0 > u_n, V_k > u_n/2] + v_n^{-1} \mathbb{P}[X_0 > u_n, \Pi_{1,k} X_0 > u_n/2] \\ &= \mathbb{P}[V_k > u_n/2] + v_n^{-1} \int_{u_n}^{\infty} \mathbb{P}[\Pi_{1,k} > u_n/(2t)] \mathcal{L}(X_0)(dt). \end{aligned}$$

There exists $\xi \in (0, \alpha)$ such that $\rho := E[C_1^\xi] < 1$. Thus $P[\Pi_{1,k} > u_n/(2t)] \leq E[\Pi_{1,k}^\xi](2t/u_n)^\xi = \rho^k(2t/u_n)^\xi$, which in turn implies

$$v_n^{-1} \int_{u_n}^{\infty} P[\Pi_{1,k} > u_n/(2t)] \mathcal{L}(X_0)(dt) \leq \rho^k E[(2X_0/u_n)^\xi | X_0 > u_n] \leq 2^{\xi+1} E[Y_0^\xi] \rho^k$$

for all $k \in \mathbb{N}$ and sufficiently large n . Because $P[V_k > u_n/2] \leq P[X_k > u_n/2] \leq 2^{1-\alpha} v_n$ for all $k \in \mathbb{N}$ and sufficiently large n , one may conclude that, for some constant $c > 0$,

$$P[X_k > u_n | X_0 > u_n] \leq c(v_n + \rho^k) =: s_n(k).$$

Condition (C) then follows from the fact that, as $n \rightarrow \infty$,

$$\sum_{k=1}^{r_n} s_n(k) = cr_n v_n + c \sum_{k=1}^{r_n} \rho^k \rightarrow c \sum_{k=1}^{\infty} \rho^k = \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} s_n(k) < \infty.$$

2.8.2 Proofs

Proof of Lemma 2.1. Note that $|X_t^*| \geq u$ if and only if $F_{|X_0|}(|X_t|) \geq 1 - 1/u$ if and only if $|X_t| \geq U(u) := F_{|X_0|}^{-1}(1 - 1/u)$ with $F_{|X_0|}^{-1}$ denoting the quantile function and $F_{|X_0|}$ the distribution function of $|X_t|$. Moreover, regular variation of $|X_0|$ implies regular variation of U as well as $P[|X_0| > u] / P[|X_0| \geq u] \rightarrow 1$ and $u \bar{F}_{|X_0|}(U(u)) \rightarrow 1$ as $u \rightarrow \infty$. Thus, from weak convergence (2.1) we may conclude

$$\begin{aligned} & \mathcal{L}\left(\left(\frac{X_i^*}{u}\right)_{s \leq i \leq t} \mid |X_0^*| \geq u\right) \\ &= \mathcal{L}\left(\left(\frac{X_i/|X_i|}{u \bar{F}_{|X_0|}(|X_i|)} \cdot \mathbf{1}(X_i \neq 0)\right)_{s \leq i \leq t} \mid |X_0| \geq U(u)\right) \\ &= \mathcal{L}\left(\left(\frac{X_i/U(u)}{|X_i/U(u)|} \cdot \mathbf{1}\left(\frac{X_i}{U(u)} \neq 0\right) \cdot \frac{1}{u \bar{F}_{|X_0|}(U(u))} \cdot \frac{\bar{F}_{|X_0|}(U(u))}{\bar{F}_{|X_0|}(U(u) \frac{|X_i|}{U(u)})}\right)_{s \leq i \leq t} \mid |X_0| \geq U(u)\right) \\ &\rightsquigarrow \mathcal{L}\left(\left(\frac{Y_i}{|Y_i|} \cdot \mathbf{1}(Y_i \neq 0) \cdot \mathbf{1} \cdot \frac{1}{|Y_i|^{-\alpha}}\right)_{s \leq i \leq t}\right), \quad u \rightarrow \infty, \end{aligned}$$

where in the last step we applied the extended continuous mapping theorem 1.11.1 of [van der Vaart and Wellner \(1996\)](#). Therefore, $(X_t^*)_{t \in \mathbb{Z}}$ has a tail process $(Y_t^*)_{t \in \mathbb{Z}}$ with $Y_t^* = \text{sign}(Y_t)|Y_t|^\alpha$ and is thus regular varying with index $\alpha^* = 1$. Moreover, $\Theta_t^* = Y_t^*/|Y_0^*| = \text{sign}(\Theta_t)|\Theta_t|^\alpha$. $\square$

Proof of Proposition 2.9. Let g be the density function of G . For real x and u ,

$$P[X_1/X_0 \leq x | X_0 > u] = \int_u^\infty P[X_1/X_0 \leq x | X_0 = y] \frac{g(y)}{1 - G(u)} dy.$$

As a consequence, $\lim_{u \rightarrow \infty} \mathbb{P}[X_1/X_0 \leq x \mid X_0 > u] = \lim_{y \rightarrow \infty} \mathbb{P}[X_1/X_0 \leq x \mid X_0 = y]$, provided the latter limit exists.

We compute the conditional distribution of X_1 given $X_0 = y$. For $(x_0, x_1) \in \mathbb{R}^2$, as $G(-\infty) = 0$ and $C(0, \cdot) = 0$, we have

$$\begin{aligned} \mathbb{P}[X_0 \leq x_0, X_1 \leq x_1] &= C(G(x_0), G(x_1)) - C(G(-\infty), G(x_1)) \\ &= \int_{-\infty}^{x_0} \dot{C}_1(G(y), G(x_1)) g(y) dy \\ &= \mathbb{E} \left[\mathbf{1}\{X_0 \leq x_0\} \dot{C}_1(G(X_0), G(x_1)) \right]. \end{aligned}$$

We find that a version of the conditional distribution of X_1 given $X_0 = y$ is given by

$$\mathbb{P}[X_1 \leq x_1 \mid X_0 = y] = \dot{C}_1(G(y), G(x_1)), \quad x_1 \in \mathbb{R}. \quad (2.48)$$

Assume $p > 0$. As $s \searrow 0$, the functions $[0, \infty) \rightarrow \mathbb{R} : z \mapsto \dot{C}_1(1-s, 1-sz)$ are assumed to converge pointwise to a continuous limit. Since these functions are monotone, the convergence holds locally uniformly, i.e., if $0 \leq z(s) \rightarrow z$ as $s \searrow 0$, then $\lim_{s \searrow 0} \dot{C}_1(1-s, 1-sz(s)) = \lim_{s \searrow 0} \dot{C}_1(1-s, 1-sz) = \eta_{1,1}(z)$.

Moreover, the function $\bar{G} = 1 - G$ is regularly varying at infinity with index $-\alpha$, too. Indeed, as $u \rightarrow \infty$ we have $\mathbb{P}[X_0 > u] / \mathbb{P}[|X_0| > u] = p > 0$, whereas the function $u \mapsto \mathbb{P}[|X_0| > u]$ is regularly varying at infinity with index $-\alpha$.

For $x \in (0, \infty)$, we find

$$\begin{aligned} \mathbb{P}[X_1/X_0 \leq x \mid X_0 = y] &= \dot{C}_1(G(y), G(xy)) \\ &= \dot{C}_1 \left(1 - \bar{G}(y), 1 - \bar{G}(y) \frac{\bar{G}(xy)}{\bar{G}(y)} \right) \\ &\rightarrow \lim_{s \searrow 0} \dot{C}_1(1-s, 1-sx^{-\alpha}) = \eta_{1,1}(x^{-\alpha}), \quad y \rightarrow \infty. \end{aligned}$$

Similarly, $\lim_{y \rightarrow \infty} G(-y)/\bar{G}(y) = (1-p)/p$ implies, for $x \in (-\infty, 0)$,

$$\begin{aligned} \mathbb{P}[X_1/X_0 \leq x \mid X_0 = y] &= \dot{C}_1(G(y), G(-|x|y)) \\ &= \dot{C}_1 \left(1 - \bar{G}(y), \bar{G}(y) \frac{G(-|x|y)}{\bar{G}(|x|y)} \frac{\bar{G}(|x|y)}{\bar{G}(y)} \right) \\ &\rightarrow \lim_{s \searrow 0} \dot{C}_1(1-s, s[(1-p)/p] |x|^{-\alpha}) = \eta_{1,0}([(1-p)/p] |x|^{-\alpha}), \end{aligned}$$

as $y \rightarrow \infty$. We conclude that $\mathcal{L}(X_1/X_0 \mid X_0 = y)$ converges weakly, as $y \rightarrow \infty$, to the distribution $\mathcal{L}(A_1)$ given by (2.42). By the argument at the beginning of the proof, the same then holds true for $\mathcal{L}(X_1/X_0 \mid X_0 > u)$ as $u \rightarrow \infty$.

The proof of (2.43) is entirely similar. $\square$

Proof of Lemma 2.10. For real x , write $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. Recall the time-change formula, (2.6). Setting $s = 0$, $t = 1$, and $i = 1$ yields, for every measurable function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $f(0, \cdot) = 0$ and for which the expectations below exist,

$$\mathbb{E}[f(\Theta_{-1}, \Theta_0)] = \mathbb{E}\left[f\left(\frac{\Theta_0}{|\Theta_1|}, \frac{\Theta_1}{|\Theta_1|}\right) |\Theta_1|^\alpha \mathbf{1}(\Theta_1 \neq 0)\right]. \quad (2.49)$$

Setting $s = -1$, $t = 0$ and $i = -1$ yields the same formula, but with the roles of Θ_{-1} and Θ_1 interchanged. By the time-change formula (2.49),

$$\mathbb{E}[(\Theta_1^+)^{\alpha}] = \mathbb{P}[\Theta_0 = +1, \Theta_{-1} \neq 0], \quad \mathbb{E}[(\Theta_1^-)^{\alpha}] = \mathbb{P}[\Theta_0 = -1, \Theta_{-1} \neq 0].$$

[The above equations even hold without conditions (i) and (ii).] Adding both identities yields, in view of (i),

$$\mathbb{P}[\Theta_{-1} \neq 0] = \mathbb{E}[|\Theta_1|^{\alpha}] = 1$$

and thus $\mathbb{E}[(\Theta_1^+)^{\alpha}] = \mathbb{P}[\Theta_0 = +1] = p$.

For $M := \Theta_1/\Theta_0$, let $\mu_+ = \mathbb{E}[(M^+)^{\alpha}]$ and $\mu_- = \mathbb{E}[(M^-)^{\alpha}]$. Then

$$\mu_+ + \mu_- = \mathbb{E}[|M|^{\alpha}] = \mathbb{E}[|\Theta_1|^{\alpha}] = 1.$$

By (ii), we have

$$\begin{aligned} p = \mathbb{E}[(\Theta_1^+)^{\alpha}] &= \mathbb{E}[(M\Theta_0)^+)^{\alpha}] = p \mathbb{E}[(M^+)^{\alpha}] + (1-p) \mathbb{E}[(M^-)^{\alpha}] \\ &= p(1 - \mu_-) + (1-p) \mu_-. \end{aligned}$$

After simplification, we find

$$(1 - 2p) \mu_- = 0,$$

so that either $p = 1/2$ or $\mu_- = 0$, i.e., $\mathbb{P}[M < 0] = 0$. This yields the first statement.

Second, we show that Θ_{-1}/Θ_0 is independent of Θ_0 , too. If $p \in \{0, 1\}$, then Θ_0 is degenerate; without loss of generality, assume that $0 < p < 1$. We need to show that, for bounded measurable functions $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}[g(\Theta_{-1}/\Theta_0) \mid \Theta_0 = +1] = \mathbb{E}[g(\Theta_{-1}/\Theta_0) \mid \Theta_0 = -1].$$

Equivalently, we need to show that, for such g ,

$$(1 - p) \mathbb{E}[g(\Theta_{-1}/\Theta_0) \mathbf{1}(\Theta_0 = +1)] = p \mathbb{E}[g(\Theta_{-1}/\Theta_0) \mathbf{1}(\Theta_0 = -1)]. \quad (2.50)$$

The above formula clearly holds for constant functions g . Hence, it holds for a function g as soon as it holds for the function $y \mapsto g(y) - g(0)$. Without loss of generality, we may therefore assume that $g(0) = 0$. But then, by the time-change formula (2.49) applied to the two functions $f_{\pm}(\theta_0, \theta_1) = g(\theta_0/\theta_1) \mathbf{1}(\theta_1 = \pm 1)$ and using independence of $M = \Theta_1/\Theta_0$ and Θ_0 ,

$$\begin{aligned} \mathbb{E}[g(\Theta_{-1}/\Theta_0) \mathbf{1}(\Theta_0 = +1)] &= \mathbb{E}[g(\Theta_0/\Theta_1) (\Theta_1^+)^{\alpha}] \\ &= \mathbb{E}[g(\Theta_0/\Theta_1) ((\Theta_1/\Theta_0)^+)^{\alpha} \mathbf{1}(\Theta_0 = +1)] \\ &\quad + \mathbb{E}[g(\Theta_0/\Theta_1) (-\Theta_1/\Theta_0)^+)^{\alpha} \mathbf{1}(\Theta_0 = -1)] \\ &= p \mathbb{E}[g(1/M) (M^+)^{\alpha}] + (1-p) \mathbb{E}[g(1/M) (M^-)^{\alpha}], \\ \mathbb{E}[g(\Theta_{-1}/\Theta_0) \mathbf{1}(\Theta_0 = -1)] &= p \mathbb{E}[g(1/M) (M^-)^{\alpha}] + (1-p) \mathbb{E}[g(1/M) (M^+)^{\alpha}]. \end{aligned}$$

Since either $p = 1/2$ or $\mathbb{P}[M < 0] = 0$, equation (2.50) follows. $\square$

Proof of Lemma 2.2. To show (2.18), apply the time-change formula (2.6) with $s = -1$, $t = 0$, $i = -1$, and $f(y_{-1}, y_0) = \mathbf{1}(y_0/y_{-1} > x, y_{-1} = 1)$, where $x \geq 0$. It follows that

$$\begin{aligned} \mathbb{P}[A_1 > x] &= \mathbb{E}\left[\mathbf{1}\left(\frac{\Theta_1}{\Theta_0} > x\right) \mathbf{1}(\Theta_0 = 1)\right] / \mathbb{P}[\Theta_0 = 1] \\ &= \mathbb{E}\left[\mathbf{1}\left(\frac{\Theta_0}{\Theta_{-1}} > x\right) \mathbf{1}(\Theta_{-1} > 0) |\Theta_{-1}|^{\alpha}\right] / \mathbb{P}[\Theta_0 = 1] \\ &= \mathbb{E}[A_{-1}^{\alpha} \mathbf{1}(1/A_{-1} > x)]. \end{aligned}$$

To show (2.19), apply the time-change formula to the function $f(y_{-1}, y_0) = \mathbf{1}(y_0/y_{-1} \leq x, y_{-1} = 1)$. The proofs of (2.20) and (2.21) are similar. $\square$

Next we establish the asymptotic normality of the forward and backward estimator, using the limit theory for empirical processes of cluster functionals developed by Drees and Rootzén (2010). To this end, for $x \geq 0$, define functions $\phi_1, \phi_{2,x}, \phi_{3,x} : [0, \infty)^3 \rightarrow [0, \infty)$ by

$$\begin{aligned} \phi_1(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_0 > 1) \\ \phi_{2,x}(y_{-1}, y_0, y_1) &:= \mathbf{1}(y_1/y_0 > x, y_0 > 1) \\ \phi_{3,x}(y_{-1}, y_0, y_1) &:= (y_{-1}/y_0)^{\alpha} \mathbf{1}(y_0/y_{-1} > x, y_{-1} > 0, y_0 > 1). \end{aligned}$$

The forward and backward estimators of the cdf of A_1 can be written as

$$\hat{F}_n^{(f, A_1)}(x) = 1 - \frac{\sum_{i=1}^n \phi_{2,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})}, \quad \hat{F}_n^{(b, A_1)}(x) = 1 - \frac{\sum_{i=1}^n \phi_{3,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})}$$

with $X_{n,i}$ defined in (2.27).

Taking up the notation of [Drees and Rootzén \(2010\)](#), we consider the empirical process $\tilde{Z}_n$ defined by

$$\tilde{Z}_n(\psi) := (nv_n)^{-1/2} \sum_{i=1}^n (\psi(X_{n,i}) - \mathbb{E}[\psi(X_{n,i})]),$$

where ψ is one of ϕ_1 , $\phi_{2,x}$ or $\phi_{3,y}$.

Proposition 2.11. *Suppose that $(X_t)_{t \in \mathbb{Z}}$ is a stationary, regularly varying time series and that the conditions $(A(x_0))$, (B) , and (C) are fulfilled for some $x_0 \geq 0$. Then, for all $y_0 \in [x_0, \infty) \cap (0, \infty)$, the process $(\tilde{Z}_n(\phi_1), (\tilde{Z}_n(\phi_{2,x}))_{x \in [x_0, \infty)}, (\tilde{Z}_n(\phi_{3,y}))_{y \in [y_0, \infty)})$ converges weakly to a centered Gaussian process $\tilde{Z}$ with covariance function given by*

$$\begin{aligned} & \text{cov}(\tilde{Z}(\psi_1), \tilde{Z}(\psi_2)) \\ &= \mathbb{E}[\psi_1(Y_{-1}, Y_0, Y_1) \psi_2(Y_{-1}, Y_0, Y_1)] + \sum_{k=1}^{\infty} \left(\mathbb{E}[\psi_1(Y_{-1}, Y_0, Y_1) \psi_2(Y_{k-1}, Y_k, Y_{k+1})] \right. \\ & \quad \left. + \mathbb{E}[\psi_2(Y_{-1}, Y_0, Y_1) \psi_1(Y_{k-1}, Y_k, Y_{k+1})] \right) \end{aligned} \quad (2.51)$$

for all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \geq x_0, y \geq y_0\}$.

Proof. We argue similarly as in the proof of Corollary 3.6 and Remark 3.7 of [Drees and Rootzén \(2010\)](#): we first establish weak convergence of all finite dimensional distributions using Theorem 2.3 of that paper, and then the asymptotic equicontinuity of the processes $(\tilde{Z}_n(\phi_{2,x}))_{x \geq x_0}$ and $(\tilde{Z}_n(\phi_{3,y}))_{y \geq y_0}$ by applying Theorem 2.10, from which the assertion follows.

First we verify that conditions (C1)–(C3) of [Drees and Rootzén \(2010\)](#) are fulfilled so that Theorem 2.3 on the convergence of finite-dimensional distributions applies. As in the proof of Corollary 3.6, (C1) can be derived from condition (3.5) of that paper, which in turn is an easy consequence of condition (C), because by the stationarity of $(X_t)_{t \in \mathbb{Z}}$, we have

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{i=1}^{r_n} \mathbf{1}(X_{n,i} \neq 0) \right)^2 \right] &= \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \mathbb{P}[X_i > u_n, X_j > u_n] \\ &\leq 2r_n v_n \sum_{k=0}^{r_n-1} \left(1 - \frac{k}{r_n} \right) \mathbb{P}[X_k > u_n \mid X_0 > u_n] \\ &\leq 2r_n v_n \sum_{k=0}^{r_n-1} s_n(k) = O(r_n v_n), \quad n \rightarrow \infty. \end{aligned}$$

Since all functions ϕ_1 , $\phi_{2,x}$ and $\phi_{3,y}$ are bounded and since $r_n = o((nv_n)^{1/2})$, condition (C2) is obviously fulfilled.

For the convergence of all finite-dimensional marginal distributions it remains to establish condition (C3) of [Drees and Rootzén \(2010\)](#), i.e.,

$$\frac{1}{r_n v_n} \text{cov} \left(\sum_{i=1}^{r_n} \psi_1(X_{n,i}), \sum_{j=1}^{r_n} \psi_2(X_{n,j}) \right) \rightarrow \text{cov} \left(\tilde{Z}(\psi_1), \tilde{Z}(\psi_2) \right), \quad n \rightarrow \infty, \quad (2.52)$$

for all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \geq x_0, y \geq y_0\}$. Similarly as above, by the stationarity of $(X_t)_{t \in \mathbb{Z}}$, the left-hand side equals

$$\begin{aligned} & \frac{1}{r_n v_n} \mathbb{E} \left[\sum_{i=1}^{r_n} \sum_{j=1}^{r_n} \psi_1(X_{n,i}) \psi_2(X_{n,j}) \right] + O(r_n v_n) \\ &= \mathbb{E} [\psi_1(X_{n,0}) \psi_2(X_{n,0}) \mid X_0 > u_n] \\ & \quad + \sum_{k=1}^{r_n-1} \left(1 - \frac{k}{r_n} \right) \left(\mathbb{E} [\psi_1(X_{n,0}) \psi_2(X_{n,k}) \mid X_0 > u_n] + \mathbb{E} [\psi_1(X_{n,k}) \psi_2(X_{n,0}) \mid X_0 > u_n] \right). \end{aligned}$$

By assumption (A(x_0)), all functions under consideration are a.s. continuous and bounded. The definition of the tail process then yields

$$\mathbb{E} [\psi_1(X_{n,0}) \psi_2(X_{n,k}) \mid X_0 > u_n] \rightarrow \mathbb{E} [\psi_1(Y_{-1}, Y_0, Y_1) \psi_2(Y_{k-1}, Y_k, Y_{k+1})], \quad n \rightarrow \infty,$$

for all $k \geq 0$ and all $\psi_1, \psi_2 \in \{\phi_1, \phi_{2,x}, \phi_{3,y} : x \geq x_0, y \geq y_0\}$. Thus (2.52) follows by Pratt's lemma ([Pratt, 1960](#)) and condition (C), since

$$\begin{aligned} \left(1 - \frac{k}{r_n} \right) \mathbb{E} [\psi_1(X_{n,0}) \psi_2(X_{n,k}) \mid X_0 > u_n] &\leq \max(1, \tilde{x}_0^{-2\alpha}) \mathbb{P} [X_k > u_n \mid X_0 > u_n] \\ &\leq \max(1, \tilde{x}_0^{-2\alpha}) s_n(k). \end{aligned}$$

In the second step, the asymptotic equicontinuity of $(\tilde{Z}_n(\phi_{2,x}))_{x \in [x_0, \infty)}$ and $(\tilde{Z}_n(\phi_{3,y}))_{y \in [y_0, \infty)}$ follows from Theorem 2.10 of [Drees and Rootzén \(2010\)](#) if their conditions (D1), (D2'), (D3), (D5) and (D6) are verified. Note that (D1) is obvious, that (D5) is an immediate consequence of the separability of the processes, and that (D2') follows from $r_n = o((nv_n)^{1/2})$ and the boundedness of all functionals $\phi_{2,x}$ and $\phi_{3,y}$. Moreover, because the maps $x \mapsto \phi_{2,x}(y_{-1}, y_0, y_1)$ and $x \mapsto \phi_{3,x}(y_{-1}, y_0, y_1)$ are decreasing, condition (D6) can be concluded in the same way as in the case $d = 1$ of Example 3.8 in [Drees and Rootzén \(2010\)](#).

It remains to establish the continuity condition (D3) for the semi-norm generated by the cdf $F^{(A_1)}$, i.e.,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{y > x \geq x_0 \\ F^{(A_1)}(y) - F^{(A_1)}(x) \leq \delta}} \frac{1}{r_n v_n} \mathbb{E} \left[\left(\sum_{i=1}^{r_n} (\phi_{2,y}(X_{n,i}) - \phi_{2,x}(X_{n,i})) \right)^2 \right] = 0 \quad (2.53)$$

and an analogous condition for $\phi_{3,y}$. By the usual stationarity argument, the expectation on the left-hand side can be bounded by a multiple of

$$r_n v_n \sum_{k=0}^{r_n-1} \mathbb{P} [X_1/X_0 \in (x, y], X_{k+1}/X_k \in (x, y], X_k > u_n \mid X_0 > u_n].$$

Now, for all fixed $M > 0$, as $n \rightarrow \infty$,

$$\begin{aligned} \sum_{k=0}^M \mathbb{P} \left[\frac{X_1}{X_0} \in (x, y], \frac{X_{k+1}}{X_k} \in (x, y], X_k > u_n \mid X_0 > u_n \right] &\leq (M+1) \mathbb{P} \left[\frac{X_1}{X_0} \in (x, y] \mid X_0 > u_n \right] \\ &\rightarrow (M+1)(F^{(A_1)}(y) - F^{(A_1)}(x)) \leq (M+1)\delta, \end{aligned} \quad (2.54)$$

uniformly for all $y > x \geq x_0$ by condition (A(x_0)). Moreover, condition (C) implies

$$\begin{aligned} \sum_{k=M+1}^{r_n} \mathbb{P} \left[\frac{X_1}{X_0} \in (x, y], \frac{X_{k+1}}{X_k} \in (x, y], X_k > u_n \mid X_0 > u_n \right] \\ \leq \sum_{k=M+1}^{r_n} s_n(k) \rightarrow \sum_{k=M+1}^{\infty} \lim_{n \rightarrow \infty} s_n(k) < \infty. \end{aligned} \quad (2.55)$$

By choosing M sufficiently large, the right-hand side of (2.55) can be made arbitrarily small. Given such M , by choosing δ small, the right-hand side of (2.54) can be made arbitrarily small too. Equation (2.53) follows. The convergence statement (2.53) involving the functions $(\phi_{3,y})_{y \geq y_0}$ can be concluded in a similar way. $\square$

Proof of Theorem 2.3. For notational simplicity, we write $\bar{F}$ for $\bar{F}^{(A_1)} = 1 - F^{(A_1)}$ and use the notations $\hat{\bar{F}}_n^{(\mathbf{f}, A_1)}(x) := 1 - \hat{F}_n^{(\mathbf{f}, A_1)}(x)$ and $\hat{\bar{F}}_n^{(\mathbf{b}, A_1)}(y) := 1 - \hat{F}_n^{(\mathbf{b}, A_1)}(y)$ for the estimators of the survival function. Check that

$$\begin{aligned} &(nv_n)^{1/2} \left(\hat{\bar{F}}_n^{(\mathbf{f}, A_1)}(x) - \mathbb{P} [X_1/X_0 > x \mid X_0 > u_n] \right) \\ &= (nv_n)^{1/2} \left(\frac{\sum_{i=1}^n \phi_{2,x}(X_{n,i})}{\sum_{i=1}^n \phi_1(X_{n,i})} - \frac{\mathbb{P} [X_1/X_0 > x, X_0 > u_n]}{v_n} \right) \\ &= (nv_n)^{1/2} \left(\frac{n \mathbb{E} [\phi_{2,x}(X_{n,0})] + (nv_n)^{1/2} \tilde{Z}_n(\phi_{2,x})}{nv_n + (nv_n)^{1/2} \tilde{Z}_n(\phi_1)} - \frac{n \mathbb{E} [\phi_{2,x}(X_{n,0})]}{nv_n} \right) \\ &= \frac{\tilde{Z}_n(\phi_{2,x}) - \mathbb{P} [X_1/X_0 > x \mid X_0 > u_n] \tilde{Z}_n(\phi_1)}{1 + (nv_n)^{-1/2} \tilde{Z}_n(\phi_1)}, \end{aligned}$$

and likewise

$$\begin{aligned} (nv_n)^{1/2} \left(\hat{\bar{F}}_n^{(b, A_1)}(y) - \mathbb{E}[(X_{-1}/X_0)^\alpha \mathbf{1}(X_0/X_{-1} > y) \mid X_0 > u_n] \right) \\ = \frac{\tilde{Z}_n(\phi_{3,y}) - \mathbb{E}[(X_{-1}/X_0)^\alpha \mathbf{1}(X_0/X_{-1} > y) \mid X_0 > u_n] \tilde{Z}_n(\phi_1)}{1 + (nv_n)^{-1/2} \tilde{Z}_n(\phi_1)}. \end{aligned}$$

Hence, in view of (2.28) and (2.29), Proposition 2.11 implies (2.30) with

$$Z^{(f, A_1)}(x) = \bar{F}(x) \tilde{Z}(\phi_1) - \tilde{Z}(\phi_{2,x}) \quad \text{and} \quad Z^{(b, A_1)}(y) = \bar{F}(y) \tilde{Z}(\phi_1) - \tilde{Z}(\phi_{3,y}).$$

The covariance structure of the limiting process follows by direct calculations. Nonnegativity of X_0 implies $Y_0 > 1$ a.s. Moreover, as $(\Theta_t)_{t \in \mathbb{Z}}$ is a Markov spectral tail chain, the random variables $Y_k/Y_{k-1} = \Theta_k/\Theta_{k-1}$, $k \in \mathbb{Z}$, are independent; for $k \in \mathbb{N}$, their common survival function is $\bar{F}$. From Proposition 2.11, we obtain

$$\begin{aligned} \text{var}(\tilde{Z}(\phi_1)) &= 1 + 2 \sum_{k=1}^{\infty} \mathbb{P}[Y_k > 1] \\ \text{cov}(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{2,x})) &= \mathbb{P}[\Theta_1 > x] + \sum_{k=1}^{\infty} (\mathbb{P}[\Theta_{k+1}/\Theta_k > x, Y_k > 1] + \mathbb{P}[\Theta_1 > x, Y_k > 1]) \\ &= \bar{F}(x) \sum_{k=0}^{\infty} \mathbb{P}[Y_k > 1] + \sum_{k=1}^{\infty} \mathbb{P}[\Theta_1 > x, Y_k > 1], \\ \text{cov}(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{2,y})) &= \mathbb{P}[\Theta_1 > \max(x, y)] + \sum_{k=1}^{\infty} (\mathbb{P}[\Theta_1 > x, \Theta_{k+1}/\Theta_k > y, Y_k > 1] \\ &\quad + \mathbb{P}[\Theta_1 > y, \Theta_{k+1}/\Theta_k > x, Y_k > 1]) \\ &= \bar{F}(\max(x, y)) + \sum_{k=1}^{\infty} (\mathbb{P}[\Theta_1 > x, Y_k > 1] \bar{F}(y) \\ &\quad + \mathbb{P}[\Theta_1 > y, Y_k > 1] \bar{F}(x)). \end{aligned}$$

Similarly,

$$\begin{aligned}
\text{cov} \left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{3,x}) \right) &= \bar{F}(x) \sum_{k=0}^{\infty} \mathbb{P}[Y_k > 1] + \sum_{k=1}^{\infty} \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_k/\Theta_{k-1} > x, Y_k > 1)] \\
\text{cov} \left(\tilde{Z}(\phi_{3,x}), \tilde{Z}(\phi_{3,y}) \right) &= \bar{F}(\max(x, y)) + \sum_{k=1}^{\infty} \left(\bar{F}(x) \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)] \right. \\
&\quad \left. + \bar{F}(y) \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_k/\Theta_{k-1} > x, Y_k > 1)] \right) \\
\text{cov} \left(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{3,y}) \right) &= \bar{F}(x) \bar{F}(y) \sum_{k=0}^{\infty} \mathbb{P}[Y_k > 1] \\
&\quad + \sum_{k=1}^{\infty} \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1)].
\end{aligned}$$

The asymptotic covariance functions of the forward and the backward estimators can then be derived as follows:

$$\begin{aligned}
&\text{cov} \left(Z^{(f, A_1)}(x), Z^{(b, A_1)}(y) \right) \\
&= \text{cov} \left(\tilde{Z}(\phi_{2,x}), \tilde{Z}(\phi_{3,y}) \right) - \bar{F}(x) \text{cov} \left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{3,y}) \right) - \bar{F}(y) \text{cov} \left(\tilde{Z}(\phi_1), \tilde{Z}(\phi_{2,x}) \right) \\
&\quad + \bar{F}(x) \bar{F}(y) \text{var}(\tilde{Z}(\phi_1)) \\
&= \bar{F}(x) \bar{F}(y) \sum_{k=1}^{\infty} \mathbb{P}[Y_k > 1] - \bar{F}(x) \sum_{k=1}^{\infty} \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_k/\Theta_{k-1} > y, Y_k > 1)] \\
&\quad - \bar{F}(y) \sum_{k=1}^{\infty} \mathbb{P}[\Theta_1 > x, Y_k > 1] + \sum_{k=1}^{\infty} \mathbb{E}[(\Theta_{k-1}/\Theta_k)^\alpha \mathbf{1}(\Theta_1 > x, \Theta_k/\Theta_{k-1} > y, Y_k > 1)].
\end{aligned}$$

The other covariances can be calculated in a similar way. $\square$

Proof of Lemma 2.6. By similar arguments as used in the proof of Proposition 2.11, one can show that under the present conditions the conclusion of Proposition 2.11 remain valid if the family of functions is extended to $\{\phi_1, \phi_{2,x}, \phi_{3,y}, \psi : x \geq x_0, y \geq y_0\}$. Hence the assertion follows from

$$\begin{aligned}
\hat{\alpha}_n - \alpha_n &= \frac{nv_n + (nv_n)^{1/2} \tilde{Z}_n(\phi_1)}{n \mathbb{E}[\tilde{\psi}(X_0/u_n) \mathbf{1}(X_0 > u_n)] + (nv_n)^{1/2} \tilde{Z}_n(\psi) + R_n} - \alpha_n \\
&= \alpha_n \frac{1 + (nv_n)^{-1/2} \tilde{Z}_n(\phi_1)}{1 + \alpha_n((nv_n)^{-1/2} \tilde{Z}_n(\psi) + (nv_n)^{-1} R_n)} - \alpha_n \\
&= \alpha_n (nv_n)^{-1/2} \frac{\tilde{Z}_n(\phi_1) - \alpha_n \{Z_n(\psi) + (nv_n)^{-1/2} R_n\}}{1 + \alpha_n \{(nv_n)^{-1/2} \tilde{Z}_n(\psi) + (nv_n)^{-1} R_n\}} \\
&= (nv_n)^{-1/2} (\alpha \tilde{Z}_n(\phi_1) - \alpha^2 Z_n(\psi) + o_P(1)) (1 + o_P(1)), \quad n \rightarrow \infty.
\end{aligned}$$

In the last step we have used stochastic boundedness of $\tilde{Z}_n(\psi)$ and $\tilde{Z}_n(\phi_1)$, which follows from their weak convergence, and the assumption that $R_n = o_P((nv_n)^{1/2})$ as $n \rightarrow \infty$. $\square$

Proof of Theorem 2.7. A Taylor expansion of the function $t \mapsto z^t$ yields

$$z^{\hat{\alpha}_n} - z^\alpha = z^\alpha \log(z) (\hat{\alpha}_n - \alpha) + \frac{1}{2} z^{\alpha + \lambda(\hat{\alpha}_n - \alpha)} (\log z)^2 (\hat{\alpha}_n - \alpha)^2$$

for some (random) $\lambda = \lambda_{z, \alpha} \in (0, 1)$. Because $z^{\tilde{\alpha}} (\log z)^2$ is bounded for all $\tilde{\alpha}$ in a neighborhood of α and all $z \leq 1/y_0$, it follows that on the event $\{X_i/X_{i-1} \geq y_0\}$, we have

$$\left| \left(\frac{X_{i-1}}{X_i} \right)^{\hat{\alpha}_n} - \left(\frac{X_{i-1}}{X_i} \right)^\alpha - (nv_n)^{-1/2} \left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) (\alpha \tilde{Z}_n(\phi_1) - \alpha^2 \tilde{Z}_n(\psi)) \right| \leq C(\hat{\alpha}_n - \alpha)^2$$

for some constant C depending only on α and y_0 (but not on i or n). Hence, by Lemma 2.6, as $n \rightarrow \infty$,

$$\begin{aligned} & (nv_n)^{-1/2} \sum_{i=1}^n \left(\left(\frac{X_{i-1}}{X_i} \right)^{\hat{\alpha}_n} - \left(\frac{X_{i-1}}{X_i} \right)^\alpha \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \\ &= \frac{\alpha \tilde{Z}_n(\phi_1) - \alpha^2 \tilde{Z}_n(\psi)}{nv_n} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \\ & \quad + o_P \left((nv_n)^{-1} \sum_{i=1}^n \mathbf{1}(X_i > u_n) \right) \\ &= \frac{\alpha \tilde{Z}_n(\phi_1) - \alpha^2 \tilde{Z}_n(\psi)}{nv_n} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) + o_P(1). \end{aligned}$$

The last equality follows from the weak convergence of $\tilde{Z}_n(\phi_1)$.

As in the proof of Proposition 2.11, one may establish weak convergence of

$$\begin{aligned} & \left((nv_n)^{-1/2} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \right. \\ & \quad \left. - \mathbb{E} \left[\left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \right] \right)_{y \geq y_0} \end{aligned}$$

to a centered Gaussian process. In particular, as $n \rightarrow \infty$,

$$\begin{aligned} & (nv_n)^{-1} \sum_{i=1}^n \left(\frac{X_{i-1}}{X_i} \right)^\alpha \log \left(\frac{X_{i-1}}{X_i} \right) \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \\ &= \mathbb{E} \left[\left(\frac{X_{-1}}{X_0} \right)^\alpha \log \left(\frac{X_{-1}}{X_0} \right) \mathbf{1} \left(\frac{X_0}{X_{-1}} > y \right) \middle| X_0 > u_n \right] + o_P(1) \\ &\rightsquigarrow \mathbb{E} [\Theta_{-1}^\alpha \log(\Theta_{-1}) \mathbf{1}(1/\Theta_{-1} > y)] =: \ell_y, \end{aligned}$$

uniformly for $y \geq y_0$. By the time-change formula (2.6) with $i = -1$, $s = t = 0$ and $f(x) = -\log(x) \mathbf{1}_{(y, \infty)}(x)$, one has $\ell_y = -\mathbb{E} [\log(\Theta_1) \mathbf{1}(\Theta_1 > y)]$.

It follows that

$$\begin{aligned} (nv_n)^{-1/2} \sum_{i=1}^n \left(\left(\frac{X_{i-1}}{X_i} \right)^{\hat{\alpha}_n} \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) - \mathbb{E} \left[\left(\frac{X_{i-1}}{X_i} \right)^{\alpha} \mathbf{1} \left(\frac{X_i}{X_{i-1}} > y, X_i > u_n \right) \right] \right) \\ = \tilde{Z}_n(\phi_{3,y}) + \ell_y(\alpha Z_n(\phi_1) - \alpha^2 \tilde{Z}_n(\tilde{\psi})) + o_P(1), \quad n \rightarrow \infty. \end{aligned}$$

Proceed as in the proof of Proposition 2.11 to arrive at the assertion. $\square$

2.9 Appendix B - Finite levels approximation assessment

In this part of the appendix we address the bias problem that can arise when approximating the limiting distribution of the tail process by the finite level conditional distribution. The analysis is specialized to the setting in the simulation study (Section 2.6) and in the case study (Section 2.7). For the former, we assess the quality of the approximation by presenting quantile-quantile plots of the limiting distribution, $\mathcal{L}(A_1)$, versus the conditional distributions at finite thresholds. For the latter, we show the stability plots with the estimators of the law of A_1 applied with different thresholds. Additionally, we show that the conclusion made in the case study section is valid for higher thresholds as well.

2.9.1 Approximation assessment for the simulation study

Copula-based Markov process

For the copula-based Markov model (see Section 2.5.1), we choose the symmetric t-distribution with $\nu_1 = 2$ degrees of freedom as the margin G and the t-copula $C_{\nu_2, \rho}^t(u, v)$ with $\nu_2 = 2.5$ and $\rho = 0.2$ to model the temporal dependence structure. Hence, $\mathbb{P}[\Theta_0 = 1] = \mathbb{P}[\Theta_0 = -1] = 1/2$ and the index of regular variation is equal to $\alpha = 2$. Let t_ν denote the cdf of the univariate t-distribution with $\nu > 0$ degrees of freedom and let $t_\nu^\leftarrow$ denote the corresponding quantile function. The distribution function of A_1 is given by

$$\mathbb{P}[A_1 \leq x] = t_{\nu_2+1} \left(\left(\text{sign}(x) |x|^{\nu_1/\nu_2} - \rho \right) \left(\frac{\nu_2 + 1}{1 - \rho^2} \right)^{1/2} \right), \quad x \in \mathbb{R}.$$

Hence, the quantile function is given by

$$F_{\text{lim}}^\leftarrow(q) = \text{sign} \left(t_{\nu_2+1}^\leftarrow(q) \left(\frac{1 - \rho^2}{\nu_2 + 1} \right)^{1/2} + \rho \right) \left| t_{\nu_2+1}^\leftarrow(q) \left(\frac{1 - \rho^2}{\nu_2 + 1} \right)^{1/2} + \rho \right|^{\nu_2/\nu_1},$$

for $q \in (0, 1)$. In the simulation study we use finite-level approximations. We have

$$\mathbb{P}[X_1/X_0 \leq x \mid X_0 = y] = \dot{C}_1(G(y), G(xy)), \quad x \in \mathbb{R}, y > 0,$$

where, in our setting, $G = t_{\nu_1}$, $\dot{C}_1(u, v) = \frac{\partial}{\partial u} C_{\nu_2, \rho}^t(u, v)$, and $y > 0$ is large. Specifically,

$$\frac{\partial}{\partial u} C_{\nu_2, \rho}^t(u, v) = t_{\nu+1} \left(\left\{ \frac{t_{\nu}^{\leftarrow}(v)}{t_{\nu}^{\leftarrow}(u)} - \rho \right\} \left(\frac{\nu_2 + 1}{1 - \rho^2} \right)^{1/2} \left\{ 1 + \frac{\nu_2}{(t_{\nu}^{\leftarrow}(u))^2} \right\}^{-1/2} \right).$$

From this we can calculate the corresponding quantile function of the finite level conditional distribution, i.e.,

$$F_{\text{fin}, y}^{\leftarrow}(q) = \frac{1}{y} t_{\nu_1}^{\leftarrow} \circ t_{\nu_2} \left(\left\{ t_{\nu_2+1}^{\leftarrow}(q) \left(\frac{((t_{\nu_2}^{\leftarrow}(u))^2 + \nu_2)(1 - \rho^2))^{1/2}}{(t_{\nu_2}^{\leftarrow}(u))^2(\nu_2 + 1)} \right) + \rho \right\} t_{\nu_2}^{\leftarrow}(u) \right),$$

where $u = t_{\nu_1}(y)$ and $q \in (0, 1)$.

In the simulation study we set the threshold to be the 97.5% quantile of the absolute values of the sample. This corresponds to $y \approx 6.2$ since $\mathbb{P}[|X_0| > 6.2] \approx 2.5\%$. In the left panel of Figure 2.6 below we plot $(F_{\text{fin}, y}^{\leftarrow}(q), F_{\text{lim}}^{\leftarrow}(q))$ for different values of $q \in (0, 1)$ and $y \in \{2.92, 4.3, 6.2\}$, where y is chosen such that $\mathbb{P}[|X_0| > 2.92] \approx 10\%$ and $\mathbb{P}[|X_0| > 4.3] \approx 5\%$. We added the 45° reference line.

Stochastic recurrence equation

For the stochastic recurrence equation (see Section 2.5.2),

$$X_t = C_t X_{t-1} + D_t, \quad t \in \mathbb{Z},$$

we let C_t and D_t be independent $N(1/3, 8/9)$ and $N(-10, 1)$ random variables, respectively. This choice ensures that $\mathbb{E}[C_t^2] = 1$. As a consequence, a stationary solution $(X_t)_{t \in \mathbb{Z}}$ exists which is regularly varying with index $\alpha = 2$. Moreover, $A_1 \sim N(1/3, 8/9)$ and $(X_1/X_0 \mid X_0 = y) \sim N(1/3 - 10/y, 8/9 + 1/y^2)$ for a finite high level $y > 0$.

In the right-hand panel of Figure 2.6 below we plot the corresponding pair of quantile functions $(F_{\text{fin}, y}^{\leftarrow}(q), F_{\text{lim}}^{\leftarrow}(q))$ for different values of $q \in (0, 1)$ and for $y \in \{47.21, 70.37, 102.18\}$, the latter values being chosen because $\mathbb{P}[|X_0| > 47.21] \approx 10\%$, $\mathbb{P}[|X_0| > 70.37] \approx 5\%$, and $\mathbb{P}[|X_0| > 102.18] \approx 2.5\%$.

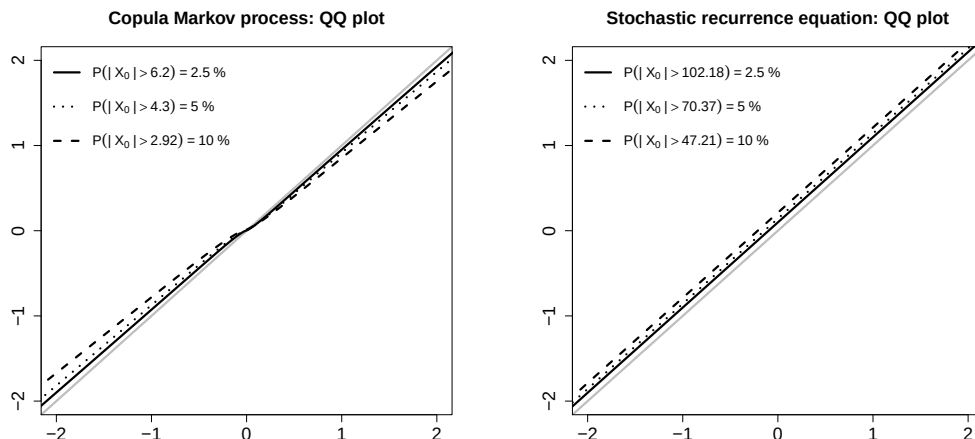


FIGURE 2.6: QQ plots of the finite level conditional distributions and limiting distributions for: copula-based Markov process (left) and stochastic recurrence equation (right).

2.9.2 Stability plots for the case study

In the case study we apply the mixture estimators with a threshold set at the 95% empirical quantiles of the absolute values of transformed log-returns of Google and UBS stock prices. In Figure 2.7 we additionally show estimates of the cdf's of A_1 for different thresholds and different estimators. For small values of $|x|$, the backward estimator is rather unstable as the threshold varies, whereas it is a bit more stable than the forward estimator for large $|x|$. The mixture estimator behaves relatively stable for thresholds larger than 90%.

Figure 2.8 displays the estimated distribution functions of A_1^* , A_{-1}^* , B_1^* , and B_{-1}^* based on the monotonized mixture estimator for Google and UBS. The threshold was set at the 97.5% empirical quantile level. Clearly, the conclusion made in the case study, regarding extremal time reversibility and market overreaction, is even more pronounced for this, higher, threshold.

Acknowledgments

H. Drees was supported by the “Deutsche Forschungsgemeinschaft”, project JA2160/1. J. Segers was supported by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). M. Warchol was funded by a “mandat d’aspirant” of the “Fonds de la Recherche Scientifique” (FNRS). We thank two anonymous referees whose comments helped to improve the presentation of the results.

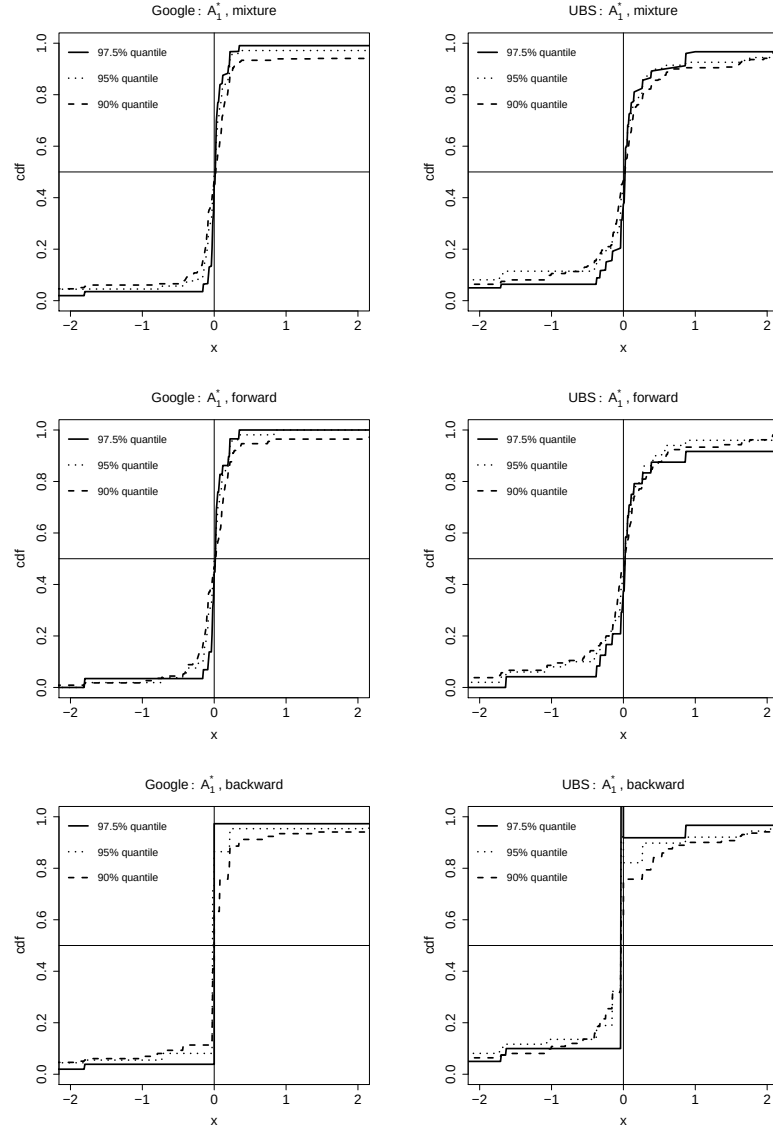


FIGURE 2.7: Estimated distribution functions of A_1^* based on the monotized mixture estimator (top), forward estimator (middle), and backward estimator (bottom) for transformed log-returns on Google (left) and UBS (right) stock prices corresponding to different thresholds: 90% (dashed lines), 95% (dotted lines), and 97.5% (solid lines).

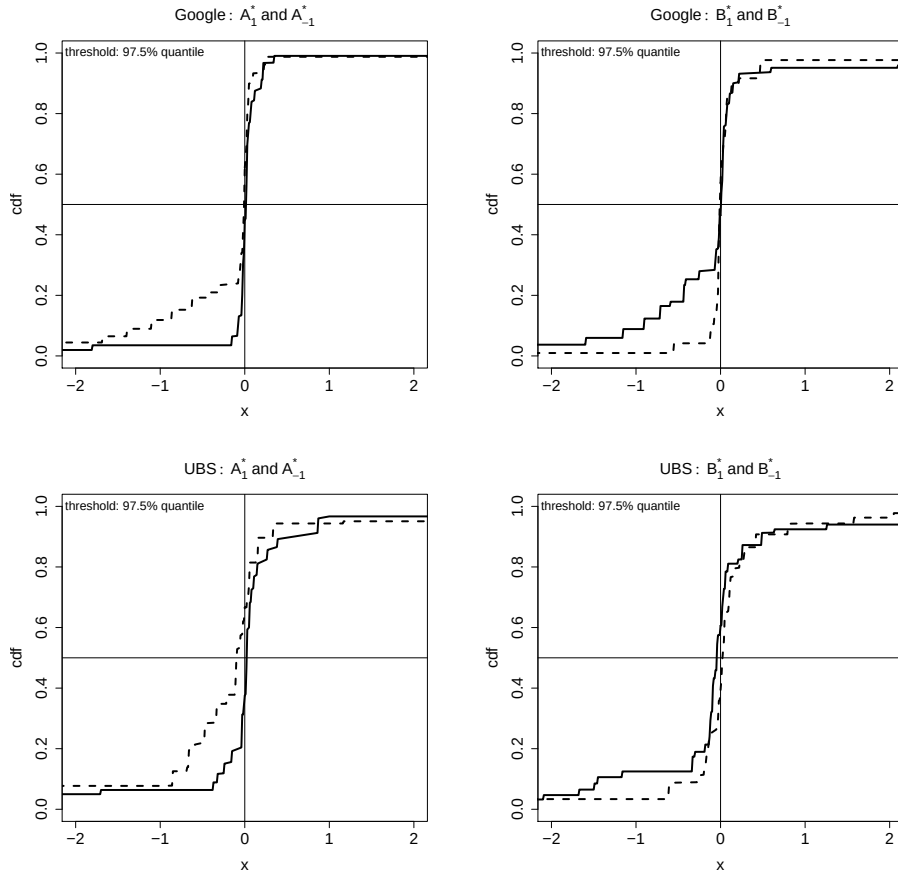


FIGURE 2.8: Estimated distribution functions of A_1^* and A_{-1}^* (left) and of B_1^* and B_{-1}^* (right) based on the monotized mixture estimator for transformed log-returns on Google (top) and UBS (bottom) stock prices with a threshold set at the 97.5% empirical quantile. Solid lines: A_1^* and B_1^* ; dashed lines: A_{-1}^* and B_{-1}^* .

Bibliography

- Basrak, B., R. A. Davis, and T. Mikosch (2002). Regular variation of GARCH processes. *Stochastic Processes and their Applications* 99(1), 95 – 115.
- Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* 119(4), 1055–1080.
- Beare, B. K. and J. Seo (2014, 4). Time irreversible copula-based markov models. *Econometric Theory FirstView*, 1–38.
- Bortot, P. and S. Coles (2000). A sufficiency property arising from the characterization of extremes of markov chains. *Bernoulli* 6(1), 183–190.
- Bowman, R. G. and D. Iverson (1998). Short-run overreaction in the New Zealand stock market. *Pacific-Basin Finance Journal* 6(5), 475–491.
- Chen, X. and Y. Fan (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130(2), 307–335.
- Chen, X., W. B. Wu, and Y. Yi (2009). Efficient estimation of copula-based semiparametric markov models. *The Annals of Statistics* 37(6B), 4214–4253.
- Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions. *Journal of Econometrics* 95(1), 199–218.
- Chen, Y.-T. and C.-M. Kuan (2002). Time irreversibility and EGARCH effects in US stock index returns. *Journal of Applied Econometrics* 17(5), 565–578.
- Davis, R. and T. Mikosch (2009). The extremogram: a correlogram for extreme events. *Bernoulli* 15(4), 977–1009.
- Davis, R. A., T. Mikosch, and Y. Zhao (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications* 123(7), 2575–2602.
- De Bondt, W. F. M. and R. Thaler (1985). Does the stock market overreact? *The Journal of Finance* 40(3), 793–805.

- Dekkers, A., J. Einmahl, and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17(4), 1833–1855.
- Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review* 73(1), 111–129.
- Doukhan, P. (1995). *Mixing. Properties and Examples*. New York: Springer.
- Drees, H. (1998a). A general class of estimators of the extreme value index. *Journal of Statistical Planning and Inference* 66(1), 95–112.
- Drees, H. (1998b). On smooth statistical tail functionals. *Scandinavian Journal of Statistics* 25(1), 187–210.
- Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics* 38(4), 2145–2186.
- Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability* 1(1), 126–166.
- Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In *Copula Theory and Its Applications*, Lecture Notes in Statistics. Springer Berlin Heidelberg.
- Janßen, A. and H. Drees (2013). A stochastic volatility model with flexible extremal dependence structure. Available at <http://arxiv.org/abs/1310.4621>.
- Janßen, A. and J. Segers (2014). Markov tail chains. *Advances in Applied Probability, forthcoming*.
- Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters* 9(1), 75–81.
- Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131(1), 207–248.
- Kulik, R. and P. Soulier (2013). Heavy tailed time series with extremal independence. Available at <http://arxiv.org/abs/1307.1501>.
- Larsson, M. and S. I. Resnick (2012). Extremal dependence measure and extremogram: the regularly varying case. *Extremes* 15(2), 231–256.
- Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 65(2), 291–306.
- Ledford, A. and J. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83(1), 169–187.

- Ledford, A. and J. Tawn (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical society, Series B* 65(2), 521–543.
- Norli, A. (2010). Short run stock overreaction: Evidence from Bursa Malaysia. *Journal of Economics and management* 4(2), 319–333.
- Perfekt, R. (1997). Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$. *Advances in Applied Probability* 29(1), 138–164.
- Pratt, J. W. (1960). On interchanging limits and integrals. *Annals of Mathematical Statistics* 31(1), 74–77.
- Segers, J. (2007). Multivariate regular variation of heavy-tailed markov chains. Technical report, Université catholique de Louvain. Available at <http://arxiv.org/abs/math/0701411>.
- Smith, R. (1987). Estimating tails of probability distributions. *Annals of Statistics* 15(3), 1174–1207.
- Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability* 29(1), 37–45.
- Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* 75(3), 397–415.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence of Empirical Processes*. New York: Springer.
- Yun, S. (2000). The distributions of cluster functionals of extreme events in a d th-order markov chain. *Journal of Applied Probability* 37(1), 29–44.

Chapter 3

Modeling Serial Extremal Dependence

Richard A. Davis¹, Holger Drees², Johan Segers³, and Michał Warchol³

Manuscript submitted for publication⁴.

Abstract

To draw inference on serial extremal dependence within heavy-tailed Markov chains, Drees, Segers and Warchol [Extremes (2015) 18, 369–402] proposed nonparametric estimators of the spectral tail process. The methodology can be extended to the more general setting of a stationary, regularly varying time series. The large-sample distribution of the estimators is derived via empirical process theory for cluster functionals. The finite-sample performance of these estimators is evaluated via Monte Carlo simulations. Moreover, two different bootstrap schemes are employed which yield confidence intervals for the pre-asymptotic spectral tail process: the stationary bootstrap and the multiplier block bootstrap. The estimators are applied to stock price data to study the persistence of positive and negative shocks.

¹Columbia University, Department of Statistics, New York, USA

²University of Hamburg, Department of Mathematics, Hamburg, Germany

³Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Louvain-la-Neuve, Belgium

⁴Appendix A (Section 3.7) has been added as a supplementary material.

3.1 Introduction

The typical modeling paradigm for a time series often starts by choosing a flexible class of models that captures salient features present in the data. Of course, *features* depends on the type of characteristics one is looking for. For a financial time series consisting of say log-returns of some asset, the key features, often referred to as *stylized facts*, include heavy-tailed marginal distributions and serially uncorrelated but dependent data. These characteristics are readily detected using standard diagnostics such as qq-plots of the marginal distribution and plots of the sample autocorrelation function (ACF) of the data and the squares of the data. The GARCH process (and its variants) as well as the stochastic volatility (SV) process driven by heavy-tailed noise exhibit these attributes and often serve as a starting point for building a model. More recently, considerable attention has been directed towards studying the extremal behavior of both financial and environmental time series, especially as it relates to estimating risk factors. Extremes for such time series can occur in clusters and getting a handle on the nature of clusters both in terms of size and frequency of occurrence is important for evaluating various risk measures. Ultimately, one wants to choose models that adequately describe various extremal dependence features observed in the data. The theme of this paper is to provide additional tools that not only give measures of extremal dependence, but can be used as a basis for assessing the quality of a model's fit to extremal properties present in the data.

The *extremal index* $\theta \in (0, 1]$ ([Leadbetter, 1983](#)) is one such measure of extremal dependence for a stationary time series. It is a measure of extremal clustering ($1/\theta$ is the mean cluster size of extremes) with $\theta < 1$ indicating clustering and $\theta = 1$ signifying no clustering in the limit. Unfortunately, θ is a rather crude measure and does not provide fine detail about extremal dependence. The extremogram, developed in [Davis and Mikosch \(2009\)](#), is an attempt to provide a measure of serial dependence among the extremes in a stationary time series. It was conceived to be used in much the same way as an ACF in traditional time series modeling, but only applied to extreme values.

In this paper, we will use the spectral tail process, as formulated by [Basrak and Segers \(2009\)](#) for heavy-tailed time series, to assess and measure extremal dependence. The spectral tail process provides a more in-depth description of the structure of extremal dependence than the extremogram. The first objective of this paper will be to establish limit theory for nonparametric estimates of the distribution of the spectral tail process for a class of heavy-tailed stationary time series. This builds on earlier work of [Drees et al. \(2015\)](#) (see Chapter 2 of the thesis) for heavy-tailed Markov chains. The nonparametric estimates provide quantitative information about extremal dependence within a time series and as such can be used in both exploratory and confirmatory phases of modeling. As an example, it provides estimates of the probability that an extreme observation will occur at time t , given one has occurred at time 0, and that its absolute value will be even larger. These estimates can also be used for model confirmation, in much the

same way that the ACF is used for assessing quality of fit for second-order models of time series. For example, one can compute a *pre-asymptotic* version (to be defined later) of the distribution of the spectral tail process from a GARCH process, which in most cases can be easily calculated via simulation. Then the estimated distribution of the spectral tail process can be compared with the *pre-asymptotic* version corresponding to a model for compatibility. A good fit would indicate the plausibility of using a GARCH model for capturing serial extremal dependence. The second main objective is then to provide a useful way of measuring compatibility, which we propose using resampling methods.

The key object of study in this paper is the tail process and in particular, its normalized version – the spectral tail process. A *strictly* stationary univariate time series $(X_t)_{t \in \mathbb{Z}}$ is said to have a *tail process* $(Y_t)_{t \in \mathbb{Z}}$ if, for all integers $s \leq t$, we have

$$\mathcal{L}(u^{-1}X_s, \dots, u^{-1}X_t \mid |X_0| > u) \rightsquigarrow \mathcal{L}(Y_s, \dots, Y_t), \quad u \rightarrow \infty, \quad (3.1)$$

with the implicit understanding that the law of $|Y_0|$ is non-degenerate. The law of $|Y_0|$ is then necessarily $\text{Pareto}(\alpha)$ for some $\alpha > 0$ and the function $u \mapsto \mathbb{P}[|X_0| > u]$ is regularly varying at infinity with index $-\alpha$:

$$\lim_{u \rightarrow \infty} \frac{\mathbb{P}[|X_0| > uy]}{\mathbb{P}[|X_0| > u]} = \mathbb{P}[|Y_0| > y] = y^{-\alpha}, \quad y \in [1, \infty). \quad (3.2)$$

The existence of a tail process is equivalent to multivariate regular variation of the finite-dimensional distributions of $(X_t)_{t \in \mathbb{Z}}$ (Basrak and Segers, 2009, Theorem 2.1). In many respects, this condition can be viewed as the heavy-tailed analogue of the condition that a process is Gaussian in the sense that all the finite-dimensional distributions are specified to be of a certain type.

The *spectral tail process* is defined by $\Theta_t = Y_t/|Y_0|$, for $t \in \mathbb{Z}$. By (3.1), it follows that for all integers $s \leq t$, we have

$$\mathcal{L}(X_0/u, X_s/|X_0|, \dots, X_t/|X_0| \mid |X_0| > u) \rightsquigarrow \mathcal{L}(Y_0, \Theta_s, \dots, \Theta_t), \quad u \rightarrow \infty. \quad (3.3)$$

The difference between (3.1) and (3.3) is that in the latter equation, the variables X_t have been normalized by $|X_0|$ rather than by the threshold u . Such auto-normalization allows the tail process to be decomposed into two stochastically independent components, i.e.,

$$Y_t = |Y_0| \Theta_t, \quad t \in \mathbb{Z}.$$

Independence of $|Y_0|$ and $(\Theta_t)_{t \in \mathbb{Z}}$ is stated in Basrak and Segers (2009, Theorem 3.1). The random variable $|Y_0|$ characterizes the magnitudes of extremes, whereas $(\Theta_t)_{t \in \mathbb{Z}}$ captures serial dependence. The spectral tail process at time $t = 0$ yields information on the relative weights of

the upper and lower tails of $|X_0|$: since $\Theta_0 = Y_0/|Y_0| = \text{sign}(Y_0)$, we have

$$p = \mathbb{P}[\Theta_0 = +1] = \lim_{u \rightarrow \infty} \frac{\mathbb{P}[X_0 > u]}{\mathbb{P}[|X_0| > u]}, \quad 1 - p = \mathbb{P}[\Theta_0 = -1]. \quad (3.4)$$

The distributions of the *forward* tail process $(Y_t)_{t \geq 0}$ and the *backward* tail process $(Y_t)_{t \leq 0}$ mutually determine each other (Basrak and Segers, 2009, Theorem 3.1). For all $i, s, t \in \mathbb{Z}$ with $s \leq 0 \leq t$ and for all measurable functions $f: \mathbb{R}^{t-s+1} \rightarrow \mathbb{R}$ satisfying $f(y_s, \dots, y_t) = 0$ whenever $y_0 = 0$, we have, provided the expectations exist,

$$\mathbb{E}[f(\Theta_{s-i}, \dots, \Theta_{t-i})] = \mathbb{E}\left[f\left(\frac{\Theta_s}{|\Theta_i|}, \dots, \frac{\Theta_t}{|\Theta_i|}\right) |\Theta_i|^\alpha \mathbf{1}\{\Theta_i \neq 0\}\right]. \quad (3.5)$$

The indicator variable $\mathbf{1}\{\Theta_i \neq 0\}$ can be omitted because of the presence of $|\Theta_i|^\alpha$, but sometimes, it is useful to mention it explicitly in order to avoid errors arising from division by zero. By exploiting the ‘time-change formula’ (3.5), we will be able to improve upon the efficiency of estimators of the spectral tail process.

Main interest in this paper is in the cumulative distribution function (cdf), $F^{(\Theta_t)}$, of Θ_t . If $F^{(\Theta_t)}$ is continuous at a point x , then

$$\lim_{u \rightarrow \infty} \mathbb{P}[X_t/|X_0| \leq x \mid |X_0| > u] = \mathbb{P}[\Theta_t \leq x] = F^{(\Theta_t)}(x). \quad (3.6)$$

The remainder of the paper is organized as follows: Two empirical estimates of the spectral tail process based on forward and backward representations for the tail process are described in Section 3.2. The companion limit theory for these estimators is formulated in Section 3.3. While the estimates are asymptotically normal, the expressions for the asymptotic variances are too complicated to be useful for constructing confidence regions. To overcome this limitation, inference procedures can be carried out using resampling methods. Two resampling methods for constructing confidence intervals, based on the stationary bootstrap as used in Davis et al. (2012), and the multiplier block bootstrap as described in Drees (2015), are also considered in Section 3.2. The validity of the proposed methodology is confirmed in Section 3.3, whereas the finite-sample performance is investigated through Monte Carlo simulations in Section 3.4. In terms of coverage probabilities, the multiplier block bootstrap generally performed better than the stationary bootstrap procedure. However, with both procedures one has to be careful when applying them for very high thresholds.

In Section 3.5, we apply the methodology to study serial extremal dependence of daily log-returns on the S&P500 index and the P&G stock price. We distinguish between two sources of such dependence - positive and negative shocks - pointing out an asymmetric behavior. Moreover, we compare up to what extent serial extremal dependence is captured by the GARCH(1,1) model and an extension of it allowing for a leverage effect, the APARCH(1,1) model. These examples

illustrate how our methodology can provide useful information on the behavior of extremes that follow both positive and negative shocks, which, in turn, can be used in a model-building context.

All proofs of the main results are collected in Section 3.6.

3.2 Methodology

3.2.1 Estimators

The data consist of a stretch $X_{1-\tilde{t}}, \dots, X_{n+\tilde{t}}$, where $\tilde{t}$ is fixed and corresponds to the maximal lag of interest, drawn from a regularly varying, stationary univariate time series with spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$ and index $\alpha > 0$.

In order to estimate $p = P[\Theta_0 = 1]$, we simply take the empirical version of (3.4), yielding

$$\hat{p}_n = \frac{\sum_{i=1}^n \mathbf{1}(X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}.$$

For $\hat{p}_n$ to be consistent and asymptotically normal, the threshold sequence u_n should tend to infinity at a certain rate described in the next section.

To estimate the cdf, $F^{(\Theta_t)}$, of Θ_t , we propose the *forward estimator*

$$\hat{F}_n^{(f, \Theta_t)}(x) := \frac{\sum_{i=1}^n \mathbf{1}(X_{i+t}/|X_i| \leq x, |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}. \quad (3.7)$$

This is just the empirical version of the left-hand side of (3.6). In equations (3.1) and (3.3), the conditioning event is $\{|X_0| > u\}$, making no distinction between positive extremes, $X_0 > u$, and negative extremes, $X_0 < -u$. However, these two cases can be distinguished by conditioning on the sign of Θ_0 . In particular, we define

$$\hat{F}_n^{(f, \Theta_t | \Theta_0 = \pm 1)}(x) := \frac{\sum_{i=1}^n \mathbf{1}(X_{i+t}/(\pm X_i) \leq x, \pm X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(\pm X_i > u_n)}.$$

The time-change formula (3.5) yields a different representation of the law of Θ_t , motivating a different estimator than the one above, based on different data points.

Lemma 3.1. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary univariate time series, regularly varying with index α and spectral tail process $(\Theta_t)_{t \in \mathbb{Z}}$. Then, for all integer $t \neq 0$,*

$$P[\Theta_t \leq x] = \begin{cases} 1 - E[|\Theta_{-t}|^\alpha \mathbf{1}(\Theta_0/|\Theta_{-t}| > x)] & \text{if } x \geq 0, \\ E[|\Theta_{-t}|^\alpha \mathbf{1}(\Theta_0/|\Theta_{-t}| \leq x)] & \text{if } x < 0. \end{cases} \quad (3.8)$$

Moreover

$$P[\Theta_t \leq x \mid \Theta_0 = 1] = \begin{cases} 1 - \frac{1}{p} E[\Theta_{-t}^\alpha \mathbf{1}(1/\Theta_{-t} > x, \Theta_0 = 1)] & \text{if } x \geq 0, \\ \frac{1}{p} E[\Theta_{-t}^\alpha \mathbf{1}(-1/\Theta_{-t} \leq x, \Theta_0 = -1)] & \text{if } x < 0, \end{cases} \quad (3.9)$$

and

$$P[\Theta_t \leq x \mid \Theta_0 = -1] = \begin{cases} 1 - \frac{1}{1-p} E[(-\Theta_{-t})^\alpha \mathbf{1}(-1/\Theta_{-t} > x, \Theta_0 = 1)] & \text{if } x \geq 0, \\ \frac{1}{1-p} E[(-\Theta_{-t})^\alpha \mathbf{1}(1/\Theta_{-t} \leq x, \Theta_0 = -1)] & \text{if } x < 0. \end{cases} \quad (3.10)$$

Assume for the moment that α is known. Lemma 3.1 suggests the following *backward estimator* of the cdf of Θ_t :

$$\hat{F}_n^{(b, \Theta_t)}(x) := \begin{cases} 1 - \frac{\sum_{i=1}^n \left| \frac{X_{i-t}}{X_i} \right|^\alpha \mathbf{1}(X_i/|X_{i-t}| > x, |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)} & \text{if } x \geq 0, \\ \frac{\sum_{i=1}^n \left| \frac{X_{i-t}}{X_i} \right|^\alpha \mathbf{1}(X_i/|X_{i-t}| \leq x, |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)} & \text{if } x < 0. \end{cases} \quad (3.11)$$

Moreover

$$\hat{F}_n^{(b, \Theta_t | \Theta_0 = \pm 1)}(x) := \begin{cases} 1 - \frac{\sum_{i=1}^n \left(\frac{\pm X_{i-t}}{X_i} \right)^\alpha \mathbf{1}(\pm X_i/X_{i-t} > x, X_i > u_n)}{\sum_{i=1}^n \mathbf{1}(\pm X_i > u_n)} & \text{if } x \geq 0, \\ \frac{\sum_{i=1}^n \left(\frac{\mp X_{i-t}}{X_i} \right)^\alpha \mathbf{1}(\pm X_i/X_{i-t} \leq x, X_i < -u_n)}{\sum_{i=1}^n \mathbf{1}(\pm X_i > u_n)} & \text{if } x < 0. \end{cases}$$

In most applications, however, the value of α is unknown. Therefore, we estimate α separately, for instance, by the Hill-type estimator

$$\hat{\alpha}_n = \frac{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)}{\sum_{i=1}^n \log(|X_i|/u_n) \mathbf{1}(|X_i| > u_n)}. \quad (3.12)$$

We plug the estimated value of α into equation (3.11) to obtain

$$\hat{F}_n^{(\hat{b}, \Theta_t)}(x) := \begin{cases} 1 - \frac{\sum_{i=1}^n \left| \frac{X_{i-t}}{X_i} \right|^{\hat{\alpha}_n} \mathbf{1}(X_i/|X_{i-t}| > x, |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)} & \text{if } x \geq 0, \\ \frac{\sum_{i=1}^n \left| \frac{X_{i-t}}{X_i} \right|^{\hat{\alpha}_n} \mathbf{1}(X_i/|X_{i-t}| \leq x, |X_i| > u_n)}{\sum_{i=1}^n \mathbf{1}(|X_i| > u_n)} & \text{if } x < 0. \end{cases} \quad (3.13)$$

3.2.2 Resampling

We explore two different bootstrap schemes that yield confidence intervals for $F^{(\Theta_t)}(x)$, or rather, for the pre-asymptotic version $P[X_t/|X_0| \leq x \mid |X_0| > u]$: the stationary bootstrap and the multiplier block bootstrap. We apply each of the two resampling schemes to both the forward and backward estimators at various levels x and at different lags t .

The stationary bootstrap goes back to [Politis and Romano \(1994\)](#) and is an adaptation of the block bootstrap by allowing for random block sizes. The resampling scheme was applied to the extremogram in [Davis et al. \(2012\)](#). It consists of generating pseudo-samples $X_1^*, \dots, X_n^*$, drawn from the sample $X_1, \dots, X_n$ by taking the first n values in the sequence

$$X_{K_1}, \dots, X_{K_1+L_1-1}, X_{K_2}, \dots, X_{K_2+L_2-1}, \dots,$$

where $K_1, K_2, \dots$ is an iid sequence of random variables uniformly distributed on $\{1, \dots, n\}$ and $L_1, L_2, \dots$ is an iid sequence of geometrically distributed random variables (independent of $(K_j)_{j \in \mathbb{N}}$) with distribution $P[L_1 = l] = p(1-p)^{l-1}$, $l = 1, 2, \dots$ for some $p = p_n \in (0, 1)$ such that $p_n \rightarrow 0$ and $np_n \rightarrow \infty$. If the index t thus obtained exceeds the sample size n , we replace t by $(t-1 \bmod n) + 1$, i.e., we continue from the beginning of the sample. The estimators are then applied to $X_{1-\tilde{t}}^*, \dots, X_{n+\tilde{t}}^*$.

The multiplier block bootstrap method was applied to cluster functionals in [Drees \(2015\)](#). It consists of splitting the data set into $m_n = \lfloor n/r_n \rfloor$ blocks of length r_n and multiplying the cluster functionals of each block by a random factor. (Here $\lfloor x \rfloor$ denotes the integer part of x .) Specifically, for iid random variables ξ_j , independent of $(X_t)_{t \in \mathbb{Z}}$, with $E[\xi_j] = 0$ and $\text{var } \xi_j = 1$, the bootstrapped forward estimator can be written as

$$\hat{F}_n^{*(f, \Theta_t)}(x) := \frac{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}\left(\frac{X_{i+t}}{|X_i|} \leq x, |X_i| > u_n\right)}{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}(|X_i| > u_n)},$$

where $I_j = \{(j-1)r_n + 1, \dots, jr_n\}$ denotes the set of indices belonging to the j th block. Similarly, the bootstrapped backward estimator for $x > 0$ with estimated index of regular variation is

$$\hat{F}_n^{*(b, \Theta_t)}(x) := 1 - \frac{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \left| \frac{X_{i-t}}{X_i} \right|^{\hat{\alpha}_n^*} \mathbf{1}\left(\left| \frac{X_i}{X_{i-t}} \right| > x, |X_i| > u_n\right)}{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}(|X_i| > u_n)} \quad (3.14)$$

with

$$\hat{\alpha}_n^* := \frac{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}(|X_i| > u_n)}{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \log(|X_i|/u_n) \mathbf{1}(|X_i| > u_n)}. \quad (3.15)$$

If the threshold u_n is high, it may be advisable to construct bootstrap confidence intervals based on lower thresholds and then scale accordingly; see the explanation after Theorem 3.4.

3.2.3 Testing for dependence of extreme observations

For iid random variables, the spectral tail process simplifies to $\Theta_t \equiv 0$ a.s. for all nonzero t . If this occurs for a stationary, regularly varying time series, then we say that the series exhibits *serial extremal independence*. The opposite case is referred to as *serial extremal dependence*, i.e., at least one of the variables Θ_t for $t \neq 0$ is not degenerate at 0. Since the convergence of the pre-asymptotic distribution can be arbitrarily slow, one cannot formally test for extremal dependence within the present framework.

However, if one wants to test whether the exceedances over a given high threshold u are independent then one may check whether the lower bound of a confidence interval for, say, $P[|X_t| \geq |X_0| \mid |X_0| > u]$ constructed by one of the bootstrap methodologies is larger than this probability under the assumption of exact independence of $|X_0|\mathbf{1}(|X_0| > u)$ and $|X_t|\mathbf{1}(|X_t| > u)$, which is easily shown to equal $P[|X_0| > u]/2$.

If one prefers to work with exceedances of the original time series (instead of its absolute values), then the probability under the assumption of independence depends on the relative weights of the upper and lower tails, and can thus not be calculated analytically. In that case, it seems natural to calculate this probability by Monte Carlo simulation. To this end, one generates (conditionally) iid samples according to the empirical distribution of the original time series by drawing with replacement from the observations, which corresponds to the classical bootstrap procedure for iid data. The considered probability under independence can be approximated by the pertaining relative frequency, which is then compared with the lower confidence bound for the probability estimated from the original time series. If the latter is larger this indicates that the exceedances in the time series exhibit a non-negligible serial dependence (see Figure 3.3 in Section 3.5).

Alternatively, one may compare the pre-asymptotic probability estimated from the observed time series using either the forward or the backward estimator with quantiles of the distribution of this estimator under independence. In a similar way as described above, the latter can be approximated by an empirical quantile obtained in Monte Carlo simulations with (conditionally) iid samples (cf. Figure 3.5).

3.3 Large-sample theory

Under certain conditions, the standardized estimation errors of the forward and the backward estimators converge jointly to a centered Gaussian process (Section 3.3.1). Convergence of the multiplier block bootstrap follows under the same conditions (Section 3.3.2).

In order not to overload the presentation, we focus on nonnegative time series. We briefly indicate how the conditions and results must be modified in the real-valued case. In addition, we distinguish between the cases where α is known or unknown.

3.3.1 Asymptotic normality of the estimators

Suppose first that the index of regular variation, α , is known. All estimators under consideration can then be expressed in terms of *generalized tail array sums*. These are statistics of the form $\sum_{i=1}^n \phi(X_{n,i})$, with

$$X_{n,i} := u_n^{-1} (X_{i-\bar{t}}, \dots, X_i, \dots, X_{i+\bar{t}}) \mathbf{1}(X_i > u_n). \quad (3.16)$$

Drees and Rootzén (2010) give conditions under which, after standardization, such statistics converge to a centered Gaussian process, uniformly over appropriate families of functions ϕ . From these results we will deduce a functional central limit theorem for the processes of forward and backward estimators defined in (3.7) and (3.11), respectively.

To ensure consistency, the threshold u_n must tend to infinity in such a way that

$$v_n := \mathbb{P}[X_0 > u_n]$$

tends to 0, but the expected number, nv_n , of exceedances tends to infinity. Moreover, we have to ensure that observations which are sufficiently separated in time are almost independent. The strength of dependence will be assessed by the β -mixing coefficients

$$\beta_{n,k} := \sup_{1 \leq l \leq n-k-1} \mathbb{E} \left[\sup_{B \in \mathcal{B}_{n,l+k+1}^n} |\mathbb{P}[B \mid \mathcal{B}_{n,1}^l] - \mathbb{P}[B]| \right].$$

Here $\mathcal{B}_{n,i}^j$ is the σ -field generated by $(X_{n,l})_{i \leq l \leq j}$.

We assume that there exist sequences $l_n, r_n \rightarrow \infty$ and some $x_0 \geq 0$ such that the following conditions hold:

(A(x_0)) The cdf of Θ_t , $F^{(\Theta_t)}$, is continuous on $[x_0, \infty)$, for $t \in \{1, \dots, \bar{t}\}$.

(B) As $n \rightarrow \infty$, we have $l_n \rightarrow \infty$, $l_n = o(r_n)$, $r_n = o((nv_n)^{1/2})$, $r_n v_n \rightarrow 0$, and $\beta_{n,l_n} n / r_n \rightarrow 0$.

Condition (B) imposes restrictions on the rate at which v_n tends to 0 and thus on the rate at which u_n tends to ∞ . Often, the β -mixing coefficients decay geometrically, i.e., $\beta_{n,k} = O(\eta^k)$ for some $\eta \in (0, 1)$. Then one may choose $l_n = O(\log n)$, and Condition (B) is fulfilled for a suitably chosen r_n if $(\log n)^2 / n = o(v_n)$ and $v_n = o(1/(\log n))$.

(C) For all $k \in \{0, \dots, r_n\}$, there exists

$$s_n(k) \geq \mathbb{P}[X_k > u_n \mid X_0 > u_n]$$

such that $s_\infty(k) = \lim_{n \rightarrow \infty} s_n(k)$ exists and $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} s_\infty(k) < \infty$.

Under these conditions, one can prove the asymptotic normality of relevant generalized tail array sums (see Proposition 3.7 below) and thus the joint asymptotic normality of the forward and the backward estimator of $F^{(\Theta_t)}$ centered by their expectation. However, additional conditions are needed to ensure that their biases are asymptotically negligible:

$$\sup_{x \in [x_0, \infty)} \left| \mathbb{P} \left[\frac{X_t}{X_0} \leq x \mid X_0 > u_n \right] - F^{(\Theta_t)}(x) \right| = o((nv_n)^{-1/2}), \quad (3.17)$$

$$\sup_{y \in [y_0, \infty)} \left| \mathbb{E} \left[\left(\frac{X_{-t}}{X_0} \right)^\alpha \mathbf{1}(X_0/X_{-t} > y) \mid X_0 > u_n \right] - \bar{F}^{(\Theta_t)}(y) \right| = o((nv_n)^{-1/2}), \quad (3.18)$$

for $t \in \{-\tilde{t}, \dots, \tilde{t}\} \setminus \{0\}$. Here $\bar{F}^{(\Theta_t)} := 1 - F^{(\Theta_t)}$ denotes the survival function of Θ_t . These conditions are fulfilled if nv_n tends to ∞ sufficiently slowly, because by definition of the spectral tail process and by (3.8), the left-hand sides in (3.17)–(3.18) tend to 0 if $F^{(\Theta_t)}$ is continuous on $[x_0, \infty)$.

Theorem 3.2. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary, regularly varying process. If (A(x_0)), (B), (C), (3.17) and (3.18) are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0, \infty) \cap (0, \infty)$, then*

$$\begin{aligned} & (nv_n)^{1/2} \left(\begin{array}{c} (\hat{F}_n^{(f, \Theta_t)}(x_t) - F^{(\Theta_t)}(x_t))_{x_t \in [x_0, \infty)} \\ (\hat{F}_n^{(b, \Theta_t)}(y_t) - F^{(\Theta_t)}(y_t))_{y_t \in [y_0, \infty)} \end{array} \right)_{|t| \in \{1, \dots, \tilde{t}\}} \\ & \rightsquigarrow \left(\begin{array}{c} (Z(\phi_{2, x_t}^t) - \bar{F}^{(\Theta_t)}(x_t)Z(\phi_1))_{x_t \in [x_0, \infty)} \\ (Z(\phi_{3, x_t}^t) - \bar{F}^{(\Theta_t)}(y_t)Z(\phi_1))_{y_t \in [y_0, \infty)} \end{array} \right)_{|t| \in \{1, \dots, \tilde{t}\}} \end{aligned}$$

where Z is a centered Gaussian process, indexed by functions defined in (3.22), whose covariance function is given in (3.23).

If the conditions of Theorem 3.2 hold except for the bias conditions (3.17) and (3.18), then the convergence still holds true if one centers with the corresponding pre-asymptotic values $P(X_t/X_0 \leq x_t \mid X_0 > u_n)$ and $1 - E((X_{-t}/X_0)^\alpha \mathbf{1}(X_0/X_{-t} \leq y_t \mid X_0 > u_n))$.

Suppose next that the index of regular variation, α , is unknown. In the definition of the backward estimator, it must then be replaced by an estimator, e.g., by the Hill-type estimator (3.12). To prove asymptotic normality of the resulting backward estimator $\hat{F}_n^{(b, \Theta_t)}(x)$, defined in (3.13), we have to sharpen condition (C). Let $\log^+(x) := \log(x) \mathbf{1}(x > 1)$.

(C') For all $k \in \{0, \dots, r_n\}$, there exists

$$s_n(k) \geq \mathbb{E} \left[\log \left(\frac{X_0}{u_n} \right) \max \left\{ \log \left(\frac{X_k}{u_n} \right), \mathbf{1}(X_k > u_n) \right\} \mid X_0 > u_n \right]$$

such that $s_\infty(k) = \lim_{n \rightarrow \infty} s_n(k)$ exists and $\lim_{n \rightarrow \infty} \sum_{k=1}^{r_n} s_n(k) = \sum_{k=1}^{\infty} s_\infty(k) < \infty$.

Moreover, there exists $\delta > 0$ such that

$$\sum_{k=1}^{r_n} \left(\mathbb{E} \left[\left(\log^+ \left(\frac{X_0}{u_n} \right) \log^+ \left(\frac{X_k}{u_n} \right) \right)^{1+\delta} \mid X_0 > u_n \right] \right)^{1/(1+\delta)} = O(1), \quad n \rightarrow \infty.$$

In addition, we need a condition to ensure that the bias of the Hill estimator is asymptotically negligible:

$$|\mathbb{E}[\log(X_0/u_n) \mid X_0 > u_n] - 1/\alpha| = o((nv_n)^{-1/2}), \quad n \rightarrow \infty. \quad (3.19)$$

Theorem 3.3. *Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary, regularly varying process. If (A(x_0)), (B), (C') and (3.17)–(3.19) are fulfilled for some $x_0 \geq 0$ and $y_0 \in [x_0, \infty) \cap (0, \infty)$, then*

$$(nv_n)^{1/2} \left(\begin{array}{c} (\hat{F}_n^{(f, \Theta_t)}(x_t) - F^{(\Theta_t)}(x_t))_{x_t \in [x_0, \infty)} \\ (\hat{F}_n^{(b, \Theta_t)}(y_t) - F^{(\Theta_t)}(y_t))_{y_t \in [y_0, \infty)} \end{array} \right)_{|t| \in \{1, \dots, \tilde{t}\}} \rightsquigarrow \left(\begin{array}{c} (Z(\phi_{2, x_t}^t) - \bar{F}^{(\Theta_t)}(x_t)Z(\phi_1))_{x_t \in [x_0, \infty)} \\ (Z(\phi_{3, x_t}^t) - \bar{F}^{(\Theta_t)}(y_t)Z(\phi_1) + (\alpha^2 Z(\phi_0) - \alpha Z(\phi_1)) \mathbb{E}[\log(\Theta_t) \mathbf{1}(\Theta_t > y_t)])_{y_t \in [y_0, \infty)} \end{array} \right)_{|t| \in \{1, \dots, \tilde{t}\}}$$

where Z is the same centered Gaussian process as in Theorem 3.2.

3.3.2 Consistency of the multiplier block bootstrap

Here we discuss the asymptotic behavior of the multiplier block bootstrap version of the forward and backward estimators. For the sake of brevity, we focus on estimators of $F^{(\Theta_t)}(x)$ for a fixed x in the case when α is unknown.

Drees (2015) has shown convergence of bootstrap versions of empirical processes of tail array sums under the same conditions needed for convergence of the original empirical processes. Let P_ξ denote the probability w.r.t. $\xi = (\xi_j)_{j \in \mathbb{N}}$, i.e., the conditional probability given $(X_{n,i})_{1 \leq i \leq n}$.

Theorem 3.4. Let ξ_j , $j \in \mathbb{N}$, be iid random variables independent of $(X_t)_{t \in \mathbb{Z}}$ with $E[\xi_j] = 0$ and $\text{var } \xi_j = 1$. Then, under the conditions of Theorem 3.3, for all $x \geq x_0, y \geq y_0$,

$$\begin{aligned} \sup_{r, s \in \mathbb{R}^{2\tilde{t}}} & \left| P_\xi \left[(nv_n)^{1/2} \left(\hat{F}_n^{*(f, \Theta_t)}(x) - \hat{F}_n^{(f, \Theta_t)}(x) \right) \leq r_t, \right. \right. \\ & \left. (nv_n)^{1/2} \left(\hat{F}_n^{*(\hat{b}, \Theta_t)}(y) - \hat{F}_n^{(\hat{b}, \Theta_t)}(y) \right) \leq s_t, \forall |t| \in \{1, \dots, \tilde{t}\} \right] \\ & - P \left[(nv_n)^{1/2} \left(\hat{F}_n^{(f, \Theta_t)}(x) - F^{(\Theta_t)}(x) \right) \leq r_t, \right. \\ & \left. (nv_n)^{1/2} \left(\hat{F}_n^{(\hat{b}, \Theta_t)}(y) - F^{(\Theta_t)}(y) \right) \leq s_t, \forall |t| \in \{1, \dots, \tilde{t}\} \right] \Big| \rightarrow 0 \end{aligned}$$

in probability.

In particular, if a and b are such that $P_\xi \left[\hat{F}_n^{*(\hat{b}, \Theta_t)}(y) \in [a, b] \right] = \beta$, then

$$\left[2\hat{F}_n^{(\hat{b}, \Theta_t)}(y) - b, 2\hat{F}_n^{(\hat{b}, \Theta_t)}(y) - a \right]$$

is a confidence interval for $F^{(\Theta_t)}(y)$ with approximative coverage level β . However, if the number of exceedances over a given threshold is too small, one may prefer to construct confidence intervals based on bootstrap estimators corresponding to lower thresholds. Let $\tilde{u}_n$ denote another threshold sequence, let $\tilde{v}_n = P[X_0 > \tilde{u}_n]$ denote the corresponding exceedance probabilities, and let $\hat{\tilde{F}}_n^{(\hat{b}, \Theta_t)}(y)$ and $\hat{\tilde{F}}_n^{*(\hat{b}, \Theta_t)}(y)$ denote the backward estimator and the bootstrap version thereof, respectively, based on the exceedances over $\tilde{u}_n$. The conditional distribution of

$$(n\tilde{v}_n)^{1/2} \left(\hat{\tilde{F}}_n^{*(\hat{b}, \Theta_t)}(y) - \hat{\tilde{F}}_n^{(\hat{b}, \Theta_t)}(y) \right)$$

given the data is approximately the same as the unconditional distribution of

$$(nv_n)^{1/2} \left(\hat{F}_n^{(\hat{b}, \Theta_t)}(y) - F^{(\Theta_t)}(y) \right).$$

Hence, if a and b are such that $P_\xi \left[\hat{\tilde{F}}_n^{*(\hat{b}, \Theta_t)}(y) \in [a, b] \right] = \beta$ and if $\hat{\tilde{v}}_n/\hat{v}_n$ is a suitable estimator of $\tilde{v}_n/v_n$, then

$$\left[\left(\frac{\hat{\tilde{v}}_n}{\hat{v}_n} \right)^{1/2} \left(\hat{\tilde{F}}_n^{(\hat{b}, \Theta_t)}(y) - b \right) + \hat{F}_n^{(\hat{b}, \Theta_t)}(y), \left(\frac{\hat{\tilde{v}}_n}{\hat{v}_n} \right)^{1/2} \left(\hat{\tilde{F}}_n^{(\hat{b}, \Theta_t)}(y) - a \right) + \hat{F}_n^{(\hat{b}, \Theta_t)}(y) \right] \quad (3.20)$$

is a confidence interval for $F^{(\Theta_t)}(y)$ with approximative coverage probability β . In practice, one will often use large order statistics as thresholds, say the k_n -th and $\tilde{k}_n$ -th largest observations, respectively. In that case, $\tilde{v}_n/v_n$ can be replaced by $\tilde{k}_n/k_n$. A similar approach, namely to use a variance estimator which is based on a lower threshold, has successfully been employed in Drees (2003, Section 5).

Of course, confidence intervals based on the bootstrap version of the forward estimator can be constructed analogously.

Remark 3.5. It is possible to generalize Theorem 3.4 to cover the joint limit distribution of the bootstrap estimators for all $x \geq x_0$ and $y \geq y_0$. Technically, this requires to endow the space of probability measures on spaces of bounded functions from $[x_0, \infty)$ (resp. $[y_0, \infty)$) to $\mathbb{R}^{2\bar{t}}$ with a metric that induces weak convergence, e.g., the bounded Lipschitz metric. This is the approach in Drees (2015) to establish the consistency of a bootstrap method for estimating the extremogram, a close cousin of the tail process. For brevity, we omit the details.

Remark 3.6. For time series which are not necessarily positive, the forward and backward estimators of $F^{(\Theta_t)}$ can be represented in terms of generalized tail array sums constructed from

$$X_{n,i}^{\bar{t}} = u_n^{-1} (X_{i-\bar{t}}, \dots, X_i, \dots, X_{i+\bar{t}}) \mathbf{1}(|X_i| > u_n).$$

When $x < 0$, for example, the backward estimator $\hat{F}_n^{(\mathbf{b}, \Theta_t)}(x)$ is equal to the ratio of the generalized tail array sums pertaining to the functions

$$\begin{aligned} (y_{-\bar{t}}, \dots, y_0, \dots, y_{-\bar{t}}) &\mapsto |y_{-t}/y_0|^\alpha \mathbf{1}(y_0/|y_{-t}| \leq x, |y_0| > 1), \\ (y_{-\bar{t}}, \dots, y_0, \dots, y_{-\bar{t}}) &\mapsto \mathbf{1}(|y_0| > 1). \end{aligned}$$

Limit theorems can be obtained by the same methods as in the positive case under obvious analogues to the conditions (A(x_0)), (B) and (C) or (C'), respectively, with $v_n := \mathbb{P}[|X_0| > u_n]$.

3.4 Finite-sample performance

In Section 3.4.1, we show results from a numerical simulation study designed to test the performance of the forward (3.7) and the backward (3.13) estimators. We continue in Section 3.4.2 by evaluating the performance of two bootstrap schemes, the multiplier block bootstrap and the stationary bootstrap, described in Section 3.2.2.

The simulations are based on pseudo-random samples from two widely used models for financial time series. Both models are of the form $X_t = \sigma_t Z_t$ where σ_t and Z_t are independent. First, we consider the GARCH(1,1) model with $\sigma_t^2 = 0.1 + 0.14X_{t-1}^2 + 0.84\sigma_{t-1}^2$, the innovations Z_t being independent t_4 random variables, standardized to have unit variance. The second model is the stochastic volatility (SV) process with $\log \sigma_t = 0.9 \log \sigma_{t-1} + \epsilon_t$, with independent standard normal innovations ϵ_t and independent innovations Z_t with common distribution $t_{2.6}$. The parameters have been chosen to ensure that both time series are regularly varying with index $\alpha = 2.6$ (Davis and Mikosch, 2001; Mikosch and Stărică, 2000).

3.4.1 Forward and backward estimators

We estimate $P[\Theta_t \leq x]$ for both the GARCH(1,1) and the SV model, for various arguments x and lags t , via the forward and the backward estimator, with estimated tail index α . The threshold is set at the empirical 95% quantile of the absolute values of a time series of length $n = 2000$. We do 1000 Monte Carlo repetitions and calculate bias, standard deviation, and root mean squared error (RMSE) with respect to the pre-asymptotic values $P[X_t/|X_0| \leq x \mid |X_0| > F_{|X|}^{\leftarrow}(0.95)]$ in the forward representation. The true quantile $F_{|X|}^{\leftarrow}(0.95)$ of $|X_0|$ and the true pre-asymptotic values were calculated numerically via 10000 Monte Carlo simulations based on time series of length 10000.

It was already reported in the context of Markovian time series that for $t = 1$ and $|x|$ large, the backward estimators usually have a smaller variance than the forward estimators (Drees et al., 2015). Here, numerical simulations suggest that this is true for non-Markovian time series and for higher lags as well. The results are presented in Figure 3.1.

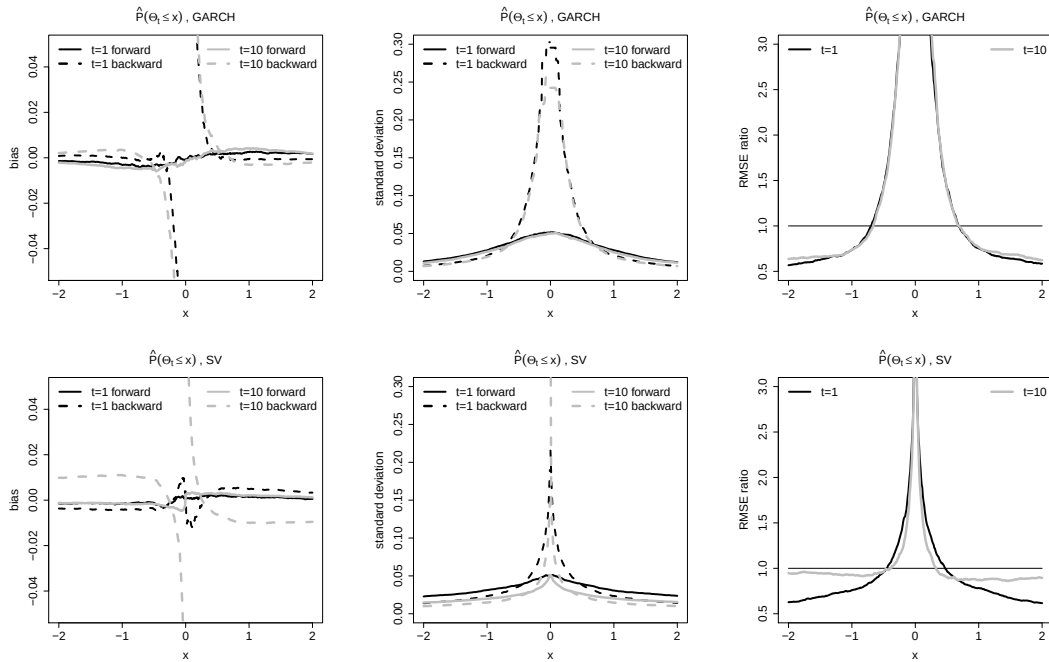


FIGURE 3.1: Performance of the forward and the backward estimators: bias (left), standard deviation (middle), ratio of root mean square errors (right) with respect to the pre-asymptotic values in the forward estimator, for GARCH(1,1) model (top) and SV model (bottom).

The right column, which shows the RMSE of the backward estimator divided by the RMSE of the forward estimators (both with respect to the pre-asymptotic values of the spectral tail process in the forward representation), shows that the backward estimator outperforms the forward estimator if x is sufficiently large in absolute value. This phenomenon was also observed at

other lags (not shown). For some other models, however, such as certain stochastic recurrence equation or copula Markov models, the advantage of the backward estimator was observed only for smaller lags ($t = 1, \dots, 4$).

3.4.2 Bootstrapped spectral tail process

We assess the performance of the two bootstrap schemes, the stationary bootstrap and the multiplier block bootstrap. To do so, we estimate the coverage probability of the bootstrapped confidence intervals with respect to the true pre-asymptotic spectral tail process in the forward representation. We focus on probabilities of the form $P[|\Theta_t| > 1]$. This particular value can be of interest due to its interpretation as the probability of a shock being followed by an even larger aftershock, i.e., $|X_t|$ being larger than $|X_0|$ conditionally on $|X_0|$ exceeding some threshold already. The true pre-asymptotic values in the forward representation were calculated numerically via 10 000 Monte Carlo simulations with time series of length 10 000.

In Figure 3.2, we plot the results for the GARCH(1,1) model and for the SV model, for the forward and the backward estimators. The expected block size for the stationary bootstrap (represented by gray lines) was chosen as 100. For the multiplier block bootstrap (black lines), the block size was fixed at 100 and the multiplier variables ξ_j were drawn independently from the standard normal distribution. Estimates of the coverage probabilities are based on 1 000 simulations. In each such sample, we use 1 000 bootstrap samples for calculating the confidence intervals with nominal coverage probability 95%. We use two different thresholds, i.e., the 95% and 98% empirical quantiles of the absolute values of a time series of length $n = 2 000$. For the higher threshold, the confidence intervals were calculated either directly (indicated by the solid lines) or using a rescaled bootstrap estimator that was based on the exceedances over the 95% empirical quantile as in (3.20) (dashed lines).

In all cases, the multiplier block bootstrap produces a better coverage probability than the stationary bootstrap. Moreover, the backward estimator is more stable than the forward one, at least for $x = 1$, and this translates into higher stability of the bootstrapped confidence intervals. The effect is especially visible for higher thresholds, e.g., at the 98% quantile, leaving insufficiently many pairs of exceedances for accurate inference. Finally, rescaled confidence intervals (3.20) based on lower thresholds can have a much better coverage than confidence intervals based on higher thresholds.

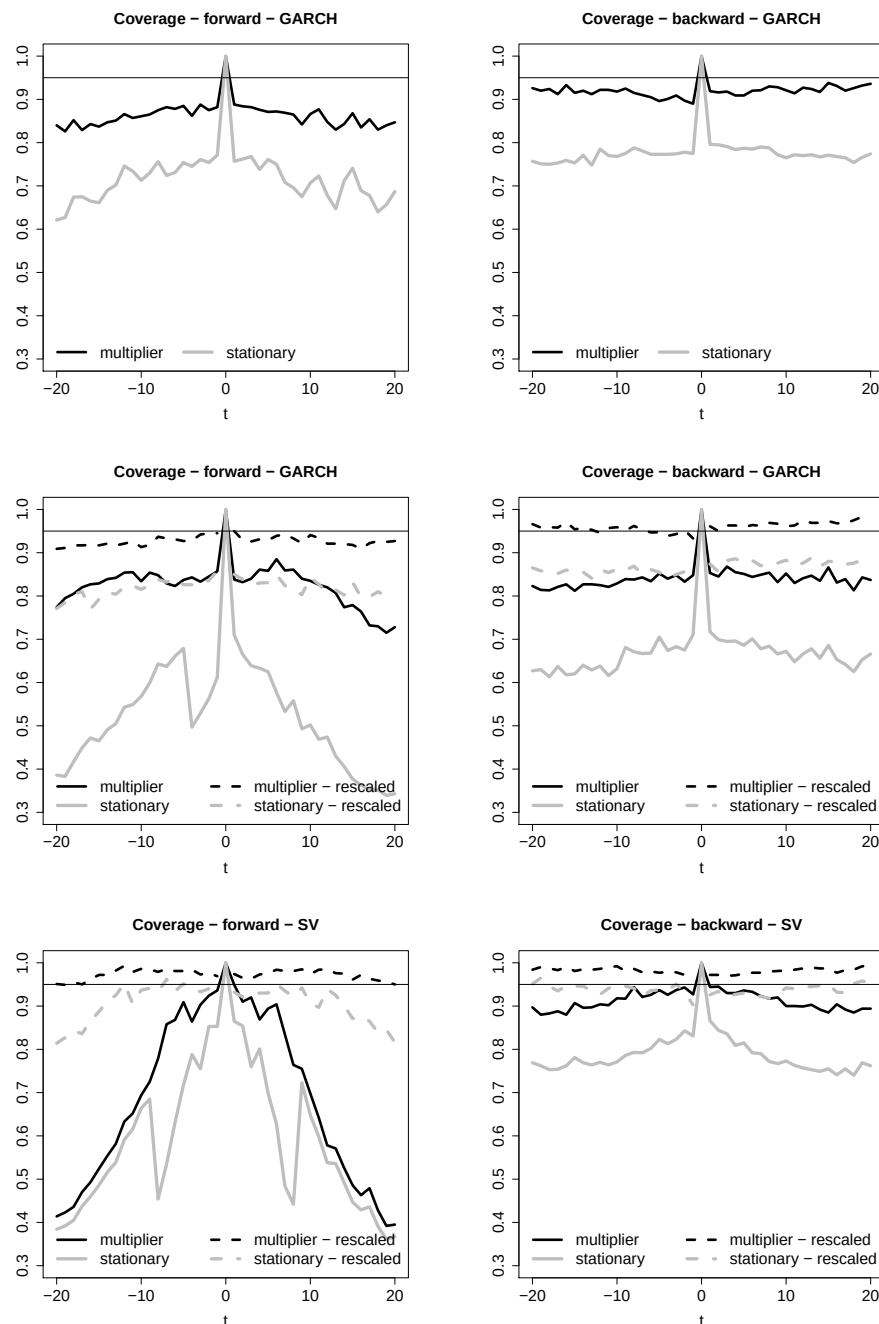


FIGURE 3.2: Coverage probabilities of confidence intervals for $P(|X_t/X_0| > 1 \mid X_0 > u_n)$ (left: forward estimator; right: backward estimator) based on the stationary bootstrap (gray) and the multiplier block bootstrap (black). The top and the middle plots correspond to the GARCH(1,1) model with thresholds set at the 95% and the 98% empirical quantiles, respectively. The bottom plots correspond to the SV simulation study with threshold set at the 98% empirical quantile. In the latter two cases, the dashed lines correspond to the coverage probabilities of the rescaled confidence intervals (3.20). The horizontal black line is the 0.95 reference line.

3.5 Application

We first consider daily log-returns on the S&P500 stock market index between 1995-01-01 and 2005-01-01 taken from Yahoo Finance⁵. In Figure 3.3, we plot the sample spectral tail process based on the backward estimator with 98% empirical quantile taken as a threshold and the 80% confidence intervals from the multiplier bootstrap scheme rescaled via the 95% quantile as threshold as in (3.20). The estimated index of regular variation is $\hat{\alpha} = 4.31$. The left-hand plot corresponds to conditioning on a positive extreme at the current time instant, whereas the right-hand plot corresponds to conditioning on a negative shock. The former plot indicates much weaker serial extremal dependence than the latter one: negative shocks are more persistent than positive ones. This characteristic is shared by other stock's daily returns which were tested but not reported here.

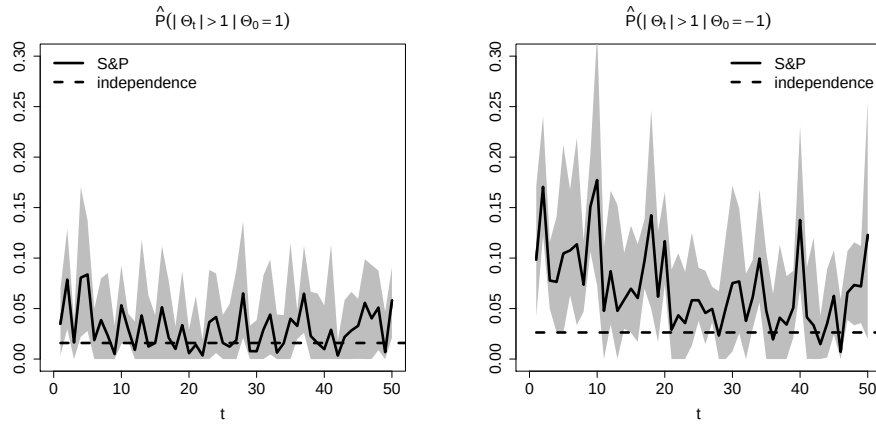


FIGURE 3.3: Sample spectral tail process (solid black line) for S&P500 daily log-returns based on the backward estimator. The gray area corresponds to the 80% confidence intervals for the pre-asymptotic spectral tail probabilities based on the multiplier bootstrap with 1000 replications. The horizontal dashed line corresponds to the pre-asymptotic spectral tail probabilities under independence.

Consider two widely used financial models of the type $X_t = \sigma_t Z_t$: first, the GARCH(1, 1) process, where

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

and second, the APARCH(1,1) process (Ding et al., 1993) with

$$\sigma_t^\delta = \omega + \alpha_1 (|X_{t-1}| - \gamma_1 X_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta.$$

Both models allow for volatility clustering in the limit. Additionally, the APARCH model captures asymmetry in the volatility of returns. That is, volatility tends to increase more when returns are negative, as compared to positive returns of the same magnitude if $\gamma_1 > 0$. The

⁵<http://finance.yahoo.com/>

		ω	α_1	β_1	δ	γ_1
S&P500	GARCH	9×10^{-7} (3×10^{-7})	0.079 (0.011)	0.918 (0.011)	-	-
	APARCH	28×10^{-5} (5×10^{-5})	0.068 (0.007)	0.927 (0.008)	0.976 (0.144)	0.999 (0.015)
P&G	GARCH	7×10^{-7} (2×10^{-7})	0.041 (0.004)	0.958 (0.004)	-	-
	APARCH	19×10^{-5} (3×10^{-5})	0.054 (0.004)	0.956 (0.003)	0.847 (0.094)	0.732 (0.07)

TABLE 3.1: Parameters of the models fitted to daily log-returns of the S&P500 index (top) and the P&G stock price (bottom). Standard errors in parentheses.

asymmetric response of volatility to positive and negative shocks is well known in the finance literature as the *leverage effect* of the stock market returns (Black, 1976).

We fit those two models to daily log-returns of the S&P500 index. We use the `garchFit` function from the `fGarch` library available in R, the function being based on maximum likelihood estimation (Wuertz et al., 2013). The innovations, Z_t , are assumed to be standard normally distributed. The fitted parameters are given in the top part of Table 3.1.

In Figure 3.4 we plot the pre-asymptotic spectral tail process probabilities based on the forward estimator for the fitted GARCH and APARCH models, together with the spectral tail process probabilities for S&P500 daily log-returns estimated by the backward estimator. The pre-asymptotic values corresponding to the fitted models were calculated numerically via 10 000 Monte Carlo simulations with time series of length 10 000. Clearly, the APARCH model captures the asymmetry which the GARCH model cannot.

As a second example, we study daily log-returns on the P&G stock price between 1993-01-04 and 2012-12-31. The tail index is estimated at $\hat{\alpha} = 3.16$. We fit the GARCH(1,1) and APARCH(1,1) models to the time series and show the estimated parameters in the bottom part of Table 3.1.

In Figure 3.5 we plot the spectral tail process probabilities based on the daily log-returns themselves and on the residuals of the fitted GARCH(1,1) and APARCH(1,1) models. The top-right plot indicates that there is significant serial extremal dependence in the P&G daily log-returns triggered by the negative shocks. Due to high asymmetry in volatility, this feature is still present in the residuals of the fitted GARCH model whereas it is better removed by the APARCH filter.

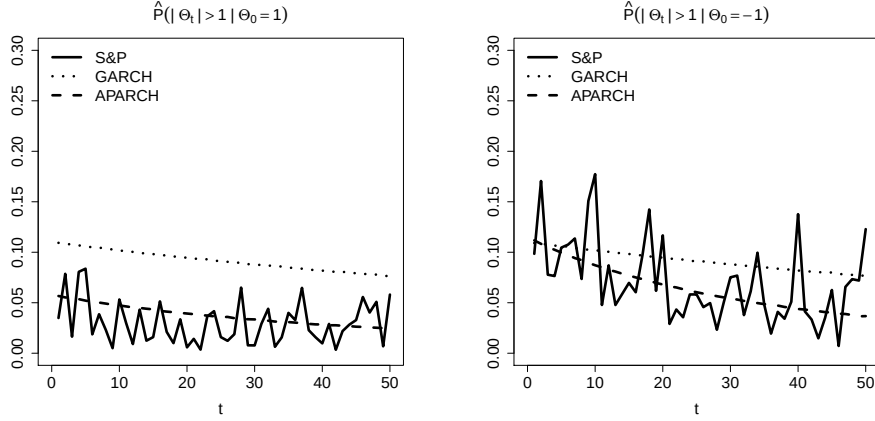


FIGURE 3.4: Sample spectral tail process probabilities (solid black line) for the S&P500 daily log-returns based on the backward estimator and pre-asymptotic spectral tail probabilities of the fitted GARCH(1,1) (dotted line) and APARCH(1,1) (dashed line) models.

3.6 Proofs

Proof of Lemma 3.1. To prove (3.8), apply the time-change formula (3.5) with $s = t = 0$, $i = -h$, and $f(y_0) = \mathbf{1}(y_0 \leq x) - \mathbf{1}(0 \leq x)$ to see that

$$\mathbb{P}[\Theta_h \leq x] - \mathbf{1}(0 \leq x) = \mathbb{E}[|\Theta_{-h}|^\alpha \mathbf{1}(\Theta_0/|\Theta_{-h}| \leq x)] - \mathbf{1}(0 \leq x) \mathbb{E}[|\Theta_{-h}|^\alpha].$$

For $x \geq 0$ in (3.9), apply the time-change formula (3.5) with $s = -h$, $t = 0$, $i = -h$ and $f(y_{-h}, \dots, y_0) = \mathbf{1}(y_0 > x, y_{-h} = 1)$ to get

$$\begin{aligned} \mathbb{P}[\Theta_h > x, \Theta_0 = 1] &= \mathbb{E}[|\Theta_{-h}|^\alpha \mathbf{1}(\Theta_0/|\Theta_{-h}| > x, \Theta_{-h} > 0)] \\ &= \mathbb{E}[\Theta_{-h}^\alpha \mathbf{1}(1/\Theta_{-h} > x, \Theta_0 = 1)], \end{aligned}$$

whereas for $x < 0$, take $f(y_{-h}, \dots, y_0) = \mathbf{1}(y_0 \leq x, y_{-h} = 1)$ to obtain

$$\mathbb{P}[\Theta_h \leq x, \Theta_0 = 1] = \mathbb{E}[\Theta_{-h}^\alpha \mathbf{1}(-1/\Theta_{-h} \leq x, \Theta_{-h} > 0, \Theta_0 = -1)].$$

Similarly, in (3.10) choose $f(y_{-h}, \dots, y_0) = \mathbf{1}(y_0 > x, y_{-h} = -1)$ and $f(y_{-h}, \dots, y_0) = \mathbf{1}(y_0 \leq x, y_{-h} = -1)$ for $x \geq 0$ and $x < 0$, respectively. $\square$

Next we turn to the asymptotic normality of the forward and backward estimators. Recall the definition of $X_{n,i}$ in (3.16). Consider the empirical process

$$\tilde{Z}_n(\psi) := (nv_n)^{-1/2} \sum_{i=1}^n (\psi(X_{n,i}) - \mathbb{E}[\psi(X_{n,i})]), \quad (3.21)$$

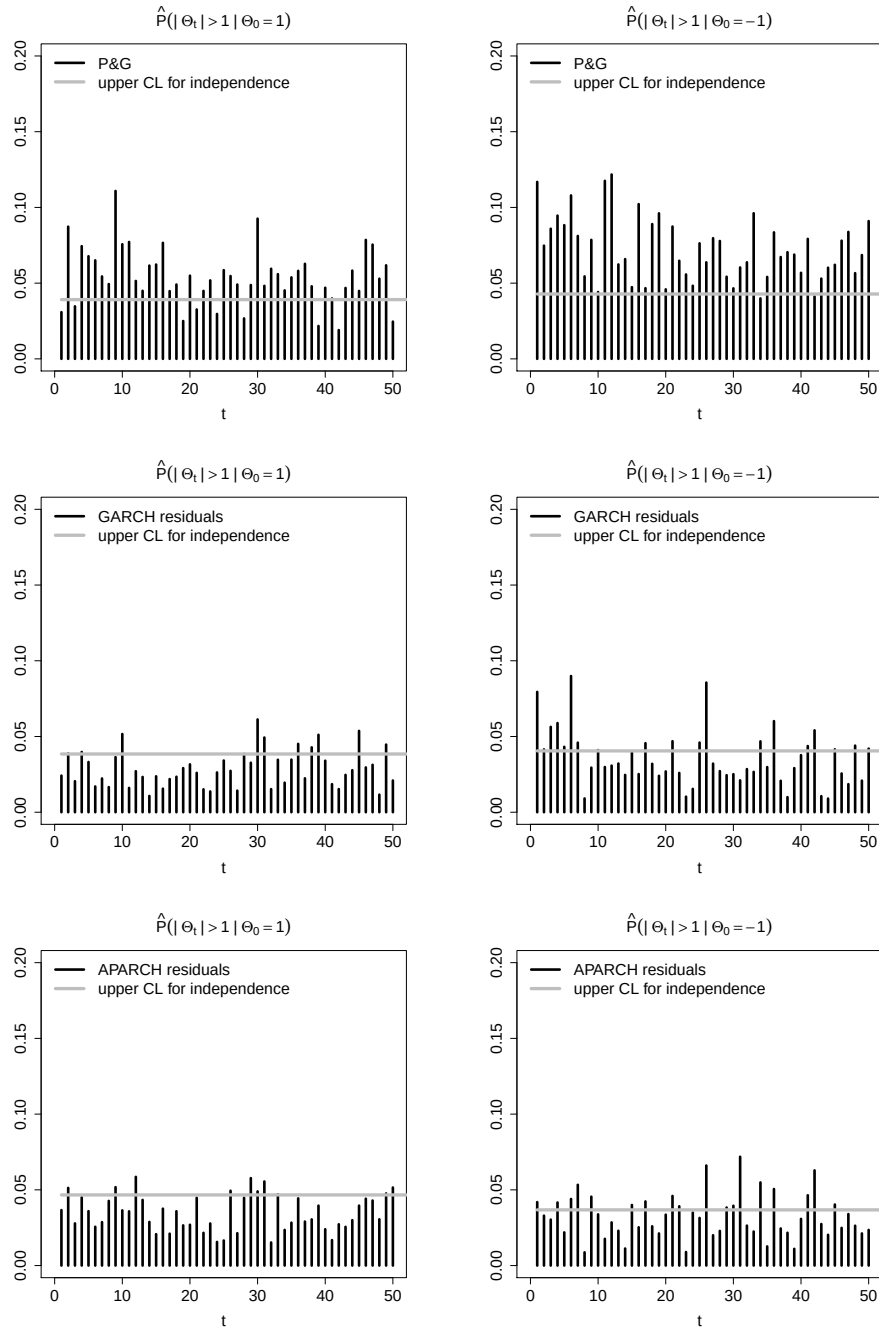


FIGURE 3.5: Sample spectral tail process (black bars) for P&G daily log-returns (top), GARCH(1,1) residuals (middle), and APARCH(1,1) residuals (bottom) based on the backward estimator. The horizontal gray lines correspond to the empirical 80% quantile of the backward estimator under independence obtained from 10 000 simulations.

where ψ is one of the following functions:

$$\begin{aligned}\phi_0(y_{-\tilde{t}}, \dots, y_0, \dots, y_{\tilde{t}}) &= \log^+(y_0), \\ \phi_1(y_{-\tilde{t}}, \dots, y_0, \dots, y_{\tilde{t}}) &= \mathbf{1}(y_0 > 1), \\ \phi_{2,x}^t(y_{-\tilde{t}}, \dots, y_0, \dots, y_{\tilde{t}}) &= \mathbf{1}(y_t/y_0 > x, y_0 > 1), \\ \phi_{3,x}^t(y_{-\tilde{t}}, \dots, y_0, \dots, y_{\tilde{t}}) &= (y_{-t}/y_0)^\alpha \mathbf{1}(y_0/y_{-t} > x, y_0 > 1)\end{aligned}\quad (3.22)$$

for $|t| \in \{1, \dots, \tilde{t}\}$ and $x \geq 0$. The asymptotic behavior of $\tilde{Z}_n$ can be derived from more general results by [Drees and Rootzén \(2010\)](#).

Proposition 3.7. *Let $(X_t)_{t \in \mathbb{Z}}$ be a positive, stationary, regularly varying time series with tail process $(Y_t)_{t \in \mathbb{Z}}$. Assume that conditions [\(A\(\$x_0\$ \)\)](#), [\(B\)](#) and [\(C'\)](#) are fulfilled for some $x_0 \geq 0$. Then, for all $y_0 \in [x_0, \infty) \cap (0, \infty)$, the sequence of processes*

$$\left(\tilde{Z}_n(\phi_0), \tilde{Z}_n(\phi_1), \left[(\tilde{Z}_n(\phi_{2,x}^t))_{x \in [x_0, \infty)}, (\tilde{Z}_n(\phi_{3,y}^t))_{y \in [y_0, \infty)} \right]_{|t| \in \{1, \dots, \tilde{t}\}} \right)$$

converges weakly to a centered Gaussian process Z with covariance function given by

$$\text{cov}(Z(\psi_1), Z(\psi_2)) = \sum_{j=-\infty}^{\infty} \mathbb{E} [\psi_1(Y_{-j}, \dots, Y_{\tilde{t}}) \psi_2(Y_{j-\tilde{t}}, \dots, Y_{j+\tilde{t}})] =: c(\psi_1, \psi_2) \quad (3.23)$$

for all $\psi_1, \psi_2 \in \{\phi_0, \phi_1, \phi_{2,x}^t, \phi_{3,y}^t \mid x \geq x_0, y \geq y_0, |t| \in \{1, \dots, \tilde{t}\}\}$.

If [\(C'\)](#) is replaced with the weaker condition [\(C\)](#), then we still have weak convergence of

$$\left(\tilde{Z}_n(\phi_1), \left[(\tilde{Z}_n(\phi_{2,x}^t))_{x \in [x_0, \infty)}, (\tilde{Z}_n(\phi_{3,y}^t))_{y \in [y_0, \infty)} \right]_{|t| \in \{1, \dots, \tilde{t}\}} \right).$$

The weak convergence statements in Proposition 3.7 hold in the space of bounded functions on $\{\phi_0, \phi_1, \phi_{2,x}^t, \phi_{3,y}^t \mid x \geq x_0, y \geq y_0, |t| \in \{1, \dots, \tilde{t}\}\}$ equipped with the supremum norm; see [van der Vaart and Wellner \(1996, Section 1.5\)](#) for details.

Proof of Proposition 3.7. One can argue similarly as in the proof of Proposition 8.4 of [Drees et al. \(2015\)](#), because the asymptotic equicontinuity of the process can be established for each t separately. Note that the discussion in [Drees and Rootzén \(2016\)](#) shows that part (ii) of condition (B) of [Drees et al. \(2015\)](#) is not needed.

By stationarity, the covariance of $Z(\psi_1)$ and $Z(\psi_2)$ is obtained as the limit of

$$\frac{1}{r_n v_n} \mathbb{E} \left[\sum_{i=1}^{r_n} \psi_1(X_{n,i}) \sum_{j=1}^{r_n} \psi_2(X_{n,j}) \right] = \frac{1}{v_n} \sum_{k=-r_n+1}^{r_n-1} \left(1 - \frac{|k|}{r_n} \right) \mathbb{E} [\psi_1(X_{n,0}) \psi_2(X_{n,k})].$$

This sum can be shown to converge to $c(\psi_1, \psi_2)$ using Pratt's lemma and Condition (C') or (C) respectively, as in Drees et al. (2015). $\square$

Remark 3.8. The covariances can be expressed in terms of the spectral tail process. For example,

$$\begin{aligned} c(\phi_{3,x}^t, \phi_0) &= \sum_{j=-\infty}^{\infty} \mathbb{E} [\Theta_{-t}^\alpha \mathbf{1}(1/\Theta_{-t} > x) \log^+(Y_0 \Theta_j)] \\ &= \sum_{j=-\infty}^{\infty} \mathbb{E} [\Theta_{-t}^\alpha \mathbf{1}(1/\Theta_{-t} > x) (\Theta_j^\alpha \wedge 1) (\log^+ \Theta_j + \alpha^{-1})]. \end{aligned}$$

Here we have used that Y_0 is independent of $(\Theta_s)_{s \in \mathbb{Z}}$ with distribution $\mathbb{P}[Y_0 > y] = y^{-\alpha}$ for $y \geq 1$.

Theorems 3.2 and 3.3 can now be proved in the same way as Theorems 4.1 and 4.6 in Drees et al. (2015). We omit the details, which can also be inferred from the more involved discussion of the bootstrap estimator below.

Drees (2015) has shown that under roughly the same conditions as used by Drees and Rootzén (2010), conditionally on the data, the following bootstrap version of the empirical process $\tilde{Z}_n$ has the same asymptotic behavior as $\tilde{Z}_n$:

$$Z_{n,\xi}(\psi) := (nv_n)^{-1/2} \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} (\psi(X_{n,i}) - \mathbb{E}[\psi(X_{n,i})]), \quad (3.24)$$

with $I_j := \{(j-1)r_n + 1, \dots, jr_n\}$ and $m_n := \lfloor n/r_n \rfloor$. In what follows, the symbol $\mathbb{E}_\xi$ denotes the expectation w.r.t. $\xi = (\xi_j)_{j \in \mathbb{N}}$, i.e., the expectation conditionally on $(X_{n,i})_{1 \leq i \leq n}$. Moreover, let BL_1 denote the set of all functions $g : \mathbb{R}^{4\tilde{t}+2} \rightarrow \mathbb{R}$ such that $\sup_{z \in \mathbb{R}^{4\tilde{t}+2}} |g(z)| \leq 1$ and $|g(z_1) - g(z_2)| \leq \|z_1 - z_2\|$ for all $z_1, z_2 \in \mathbb{R}^{4\tilde{t}+2}$.

Proposition 3.9. *Suppose that $(X_t)_{t \in \mathbb{Z}}$ is a positive, stationary, regularly varying time series and that the conditions (A(x_0)), (B) and (C') are fulfilled for some $x_0 \geq 0$. Then, for all $x \geq x_0$ and all $y_0 \in [x_0, \infty) \cap (0, \infty)$, one has*

$$\begin{aligned} &\left(Z_{n,\xi}(\phi_0), Z_{n,\xi}(\phi_1), [Z_{n,\xi}(\phi_{2,x}^t), Z_{n,\xi}(\phi_{3,y}^t)]_{|t| \in \{1, \dots, \tilde{t}\}} \right) \\ &\rightsquigarrow \left(Z(\phi_0), Z(\phi_1), [Z(\phi_{2,x}^t), Z(\phi_{3,y}^t)]_{|t| \in \{1, \dots, \tilde{t}\}} \right) \end{aligned}$$

with Z as defined in Theorem 3.2. Moreover,

$$\begin{aligned} \sup_{g \in BL_1} \left| \mathbb{E}_\xi g \left(Z_{n,\xi}(\phi_0), Z_{n,\xi}(\phi_1), [Z_{n,\xi}(\phi_{2,x}^t), Z_{n,\xi}(\phi_{3,y}^t)]_{|t| \in \{1, \dots, \tilde{t}\}} \right) \right. \\ \left. - \mathbb{E} g \left(Z(\phi_0), Z(\phi_1), [Z(\phi_{2,x}^t), Z(\phi_{3,y}^t)]_{|t| \in \{1, \dots, \tilde{t}\}} \right) \right| \rightarrow 0 \end{aligned} \quad (3.25)$$

in probability.

Proposition 3.9 follows immediately from Drees (2015, Theorem 2.1), because in the proof of Theorem 3.2 (cf. the proof of Proposition 8.4 of Drees et al. (2015)) it is shown that the assumptions of Drees (2015, Theorem 2.1) follow from the conditions of Proposition 3.9.

Now we are ready to prove the consistency of the multiplier block bootstrap procedure.

Proof of Theorem 3.4. We only prove consistency of the bootstrap version of the backward estimator, as the proof for the forward estimator is considerably simpler. For simplicity, we assume that $n = m_n r_n$. Let

$$\alpha_n := \frac{1}{\mathbb{E} [\log^+(X_0/u_n) \mid X_0 > u_n]} = \frac{v_n}{\mathbb{E} [\phi_0(X_{n,1})]}.$$

Recall $\tilde{Z}_n$ and $Z_{n,\xi}$ in (3.21) and (3.24) respectively, recall $I_j = \{(j-1)r_n + 1, \dots, jr_n\}$, and recall $\hat{\alpha}_n$ and $\hat{\alpha}_n^*$ in (3.12) and (3.15), respectively. Then

$$\begin{aligned} (nv_n)^{1/2}(\hat{\alpha}_n^* - \hat{\alpha}_n) &= (nv_n)^{1/2} \frac{\sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \mathbf{1}(X_i > u_n) - \hat{\alpha}_n \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \log^+(X_i/u_n)}{\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \log^+(X_i/u_n)} \\ &= \frac{Z_{n,\xi}(\phi_1) - \hat{\alpha}_n Z_{n,\xi}(\phi_0) + (r_n v_n)^{1/2} m_n^{-1/2} \sum_{j=1}^{m_n} \xi_j (1 - \hat{\alpha}_n/\alpha_n)}{\alpha_n^{-1} (1 + m_n^{-1} \sum_{j=1}^{m_n} \xi_j) + (nv_n)^{-1/2} \{\tilde{Z}_n(\phi_0) + Z_{n,\xi}(\phi_0)\}}. \end{aligned}$$

Since $m_n^{-1/2} \sum_{j=1}^{m_n} \xi_j$ and $\tilde{Z}_n$ are stochastically bounded and $\hat{\alpha}_n \rightarrow \alpha$ in probability, the assumptions $nv_n \rightarrow \infty$, $r_n v_n \rightarrow 0$, and $\alpha_n \rightarrow \alpha$, $m_n^{-1/2}$ and Proposition 3.9 ensure that

$$(nv_n)^{1/2}(\hat{\alpha}_n^* - \hat{\alpha}_n) = \alpha Z_{n,\xi}(\phi_1) - \alpha^2 Z_{n,\xi}(\phi_0) + o_P(1), \quad (3.26)$$

which converges weakly to $\alpha Z(\phi_1) - \alpha^2 Z(\phi_0)$. Moreover, conditionally on the data, it converges to the same limit weakly in probability in the sense of (3.25).

Next, recall $\hat{F}_n^{(\hat{b}, \Theta_t)}(y)$ and $\hat{F}_n^{*(\hat{b}, \Theta_t)}(y)$ in (3.13) and (3.14), respectively. For $y > 0$, we have

$$\left(1 - \hat{F}_n^{(\hat{b}, \Theta_t)}(y)\right) \sum_{i=1}^n \mathbf{1}(X_i > u_n) = \sum_{i=1}^n (X_{i-t}/X_i)^{\hat{\alpha}_n} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n).$$

It follows that

$$\begin{aligned}
& \hat{F}_n^{(\hat{b}, \Theta_t)}(y) - \hat{F}_n^{*(\hat{b}, \Theta_t)}(y) \\
&= \left[\sum_{i=1}^n \left(\left(\frac{X_{i-t}}{X_i} \right)^{\hat{\alpha}_n^*} - \left(\frac{X_{i-t}}{X_i} \right)^{\hat{\alpha}_n} \right) \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \right. \\
&\quad + \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \left(\frac{X_{i-t}}{X_i} \right)^{\hat{\alpha}_n^*} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&\quad \left. - \{1 - \hat{F}_n^{(\hat{b}, \Theta_t)}(y)\} \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \mathbf{1}(X_i > u_n) \right] / \left[\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}(X_i > u_n) \right].
\end{aligned}$$

For any pair $(\underline{\alpha}, \bar{\alpha})$ such that $0 < \underline{\alpha} < \alpha < \bar{\alpha}$, there exists a constant $0 < C < \infty$ such that for all $\tilde{\alpha} \in [\underline{\alpha}, \bar{\alpha}]$ and, for suitable constants $\lambda = \lambda(\tilde{\alpha}) \in (0, 1)$, we have, on the event $\{X_i/X_{i-t} > y_0\}$,

$$\begin{aligned}
& \left| \left(\frac{X_{i-t}}{X_i} \right)^{\tilde{\alpha}} - \left(\frac{X_{i-t}}{X_i} \right)^{\alpha} - \left(\frac{X_{i-t}}{X_i} \right)^{\alpha} \log \left(\frac{X_{i-t}}{X_i} \right) (\tilde{\alpha} - \alpha) \right| \\
&= \frac{1}{2} \left(\frac{X_{i-t}}{X_i} \right)^{\alpha + \lambda(\tilde{\alpha} - \alpha)} \log^2 \left(\frac{X_{i-t}}{X_i} \right) (\tilde{\alpha} - \alpha)^2 \leq C(\tilde{\alpha} - \alpha)^2.
\end{aligned}$$

Hence

$$\begin{aligned}
& \hat{F}_n^{(\hat{b}, \Theta_t)}(y) - \hat{F}_n^{*(\hat{b}, \Theta_t)}(y) \\
&= \left[\sum_{i=1}^n \left(\frac{X_{i-t}}{X_i} \right)^{\alpha} \log \left(\frac{X_{i-t}}{X_i} \right) (\hat{\alpha}_n^* - \hat{\alpha}_n) \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \right. \\
&\quad + \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \left\{ \left(\frac{X_{i-t}}{X_i} \right)^{\alpha} + \left(\frac{X_{i-t}}{X_i} \right)^{\alpha} \log \left(\frac{X_{i-t}}{X_i} \right) (\hat{\alpha}_n^* - \alpha) \right\} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&\quad \left. - (1 - \hat{F}_n^{(\hat{b}, \Theta_t)}(y)) \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \mathbf{1}(X_i > u_n) + R_n(y) \right] / \left[\sum_{j=1}^{m_n} (1 + \xi_j) \sum_{i \in I_j} \mathbf{1}(X_i > u_n) \right] \quad (3.27)
\end{aligned}$$

with

$$\begin{aligned}
|R_n(y)| &\leq C(\hat{\alpha}_n^* - \hat{\alpha}_n)^2 \sum_{i=1}^n \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&\quad + C(\hat{\alpha}_n^* - \alpha)^2 \sum_{j=1}^{m_n} |\xi_j| \sum_{i \in I_j} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\
&= O_P((nv_n)^{-1}nv_n + (nv_n)^{-1}m_nr_nv_n) = O_P(1), \quad n \rightarrow \infty.
\end{aligned}$$

Consider the function

$$\phi_{4,x}^t(y_{-\tilde{t}}, \dots, y_0, \dots, y_{\tilde{t}}) = (y_{-t}/y_0)^\alpha \log(y_{-t}/y_0) \mathbf{1}(y_0/y_{-t} > x, y_{-t} > 0, y_0 > 1).$$

One may show as in the proof of Theorem 3.2 that $\tilde{Z}_n(\phi_{4,y}^t)$ and $Z_{n,\xi}(\phi_{4,y}^t)$ both converge weakly to $Z(\phi_{4,y}^t)$. In particular, as $n \rightarrow \infty$,

$$\begin{aligned} & (nv_n)^{-1} \sum_{i=1}^n \left(\frac{X_{i-t}}{X_i} \right)^\alpha \log \left(\frac{X_{i-t}}{X_i} \right) \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\ &= \mathbb{E} \left[\left(\frac{X_{-t}}{X_0} \right)^\alpha \log \left(\frac{X_{-t}}{X_0} \right) \mathbf{1}(X_0/X_{-t} > y) \mid X_0 > u_n \right] + O_P((nv_n)^{-1/2}) \\ &\rightarrow \mathbb{E} [\Theta_{-t}^\alpha \log(\Theta_{-t}) \mathbf{1}(1/\Theta_{-t} > y)] \\ &= -\mathbb{E} [\log(\Theta_t) \mathbf{1}(\Theta_t > y)], \end{aligned}$$

where the last step follows from the time-change formula (3.5) applied with $f(y_0) = -\log(y_0) \mathbf{1}(y_0 > y)$ and $(-t, 0, -t)$ instead of (s, t, i) . Therefore

$$\begin{aligned} & \sum_{i=1}^n \left(\frac{X_{i-t}}{X_i} \right)^\alpha \log \left(\frac{X_{i-t}}{X_i} \right) (\hat{\alpha}_n^* - \hat{\alpha}_n) \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\ &= -(nv_n)^{1/2} (\mathbb{E} [\log(\Theta_t) \mathbf{1}(\Theta_t > y)] + o_P(1)) (nv_n)^{1/2} (\hat{\alpha}_n^* - \hat{\alpha}_n). \end{aligned} \quad (3.28)$$

Likewise, one can conclude that

$$\begin{aligned} & (nv_n)^{-1/2} \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \left(\frac{X_{i-t}}{X_i} \right)^\alpha \log \left(\frac{X_{i-t}}{X_i} \right) \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\ &= Z_{n,\xi}(\phi_{4,y}^t) + (rv_n)^{1/2} m_n^{-1/2} \sum_{j=1}^{m_n} \xi_j \mathbb{E} \left[\left(\frac{X_{-t}}{X_0} \right)^\alpha \log \left(\frac{X_{-t}}{X_0} \right) \mathbf{1}(X_0/X_{-t} > y) \mid X_0 > u_n \right] \\ &= O_P(1). \end{aligned}$$

As a consequence,

$$\begin{aligned} & \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \left\{ \left(\frac{X_{i-t}}{X_i} \right)^\alpha + \left(\frac{X_{i-t}}{X_i} \right)^\alpha \log \left(\frac{X_{i-t}}{X_i} \right) (\hat{\alpha}_n^* - \alpha) \right\} \mathbf{1}(X_i/X_{i-t} > y, X_i > u_n) \\ &= (nv_n)^{1/2} Z_{n,\xi}(\phi_{3,y}^t) + \sum_{j=1}^{m_n} \xi_j r_n v_n (\mathbb{E} [\Theta_{-t}^\alpha \mathbf{1}(1/\Theta_{-t} > y)] + o(1)) + O_P(1) \\ &= (nv_n)^{1/2} \left(Z_{n,\xi}(\phi_{3,y}^t) + O_P((r_n v_n)^{1/2}) + O_P((nv_n)^{-1/2}) \right). \end{aligned} \quad (3.29)$$

Moreover, we find, as $r_n v_n \rightarrow 0$ and $\sum_{j=1}^{m_n} \xi_j = O_P(m_n^{1/2})$, that

$$\begin{aligned} \sum_{j=1}^{m_n} \xi_j \sum_{i \in I_j} \mathbf{1}(X_i > u_n) &= (nv_n)^{1/2} Z_{n,\xi}(\phi_1) + r_n v_n \sum_{j=1}^{m_n} \xi_j \\ &= (nv_n)^{1/2} (Z_{n,\xi}(\phi_1) + o_P(1)), \quad n \rightarrow \infty. \end{aligned} \quad (3.30)$$

The denominator of (3.27) equals $nv_n + O_P((nv_n)^{1/2})$. Combining (3.27)–(3.30) and (3.26) yields

$$\begin{aligned}
 & (nv_n)^{1/2} (\hat{F}_n^{(\hat{b}, \Theta_t)}(y) - \hat{F}_n^{*(\hat{b}, \Theta_t)}(y)) \\
 &= -E[\log(\Theta_t) \mathbf{1}(\Theta_t > y)] (nv_n)^{1/2} (\hat{\alpha}_n^* - \hat{\alpha}_n) + Z_{n,\xi}(\phi_{3,y}^t) \\
 &\quad - (1 - \hat{F}_n^{(\hat{b}, \Theta_t)}(y)) Z_{n,\xi}(\phi_1) + o_P(1) \\
 &= Z_{n,\xi}(\phi_{3,y}^t) - \bar{F}^{(\Theta_t)}(y) Z_{n,\xi}(\phi_1) \\
 &\quad - E[\log(\Theta_t) \mathbf{1}(\Theta_t > y)] (\alpha Z_{n,\xi}(\phi_1) - \alpha^2 Z_{n,\xi}(\phi_0)) + o_P(1).
 \end{aligned}$$

Now the assertion is a direct consequence of Proposition 3.9 and Theorem 3.3. $\square$

3.7 Appendix A - Additional simulation study

In this section we show results from an additional numerical simulation study where we test the performance of the forward (3.7) and the backward (3.13) estimators. The simulation is based on pseudo-random samples from the GARCH(1,1) model fitted to daily log-returns of the S&P500 index used in Section 3.5, see Table 3.1. Such a model could be considered as more realistic than the ones used in the simulation study in Section 3.4.

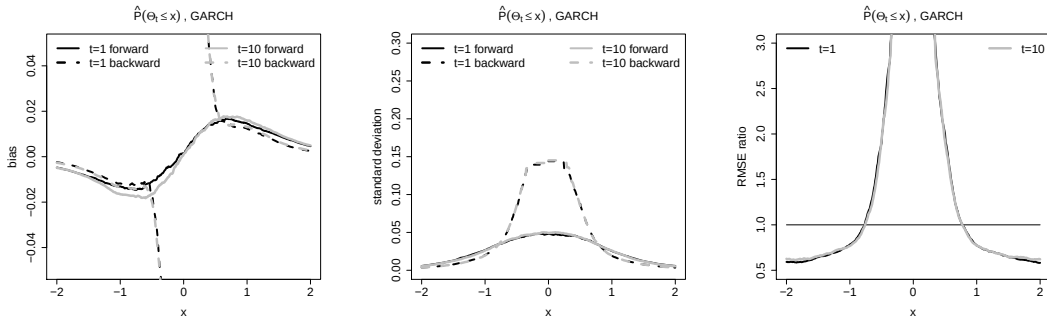


FIGURE 3.6: Performance of the forward and the backward estimators: bias (left), standard deviation (middle), ratio of root mean square errors (right) with respect to the pre-asymptotic values in the forward estimator, for the GARCH(1,1) model fitted to daily log-returns of the S&P500 index.

Except for considering a different model, the setting of the simulation study is the same as in Section 3.4.1. The results are presented in Figure 3.6. The right column shows the RMSE of the backward estimator divided by the RMSE of the forward estimators (both with respect to the pre-asymptotic values of the spectral tail process in the forward representation). We see that the backward estimator outperforms the forward estimator if x is sufficiently large in absolute value. This is in line with the results of the simulation study reported in Section 3.4.1 for different models.

Acknowledgements

J. Segers gratefully acknowledges funding by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). The research of M. Warchol was funded by a PhD grant of the “Fonds de la Recherche Scientifique” (F.R.S.-FNRS). H. Drees was partially supported by DFG research grant JA 2160/1. R. Davis was supported in part by ARO MURI grant W911NF-12-1-0385.

Bibliography

- Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* 119(4), 1055–1080.
- Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Section*, pp. 177–181.
- Davis, R. A. and T. Mikosch (2001). Point process convergence of stochastic volatility processes with application to sample autocorrelation. *Journal of Applied Probability* 38A, 93–104.
- Davis, R. A. and T. Mikosch (2009). The extremogram: a correlogram for extreme events. *Bernoulli* 15(4), 977–1009.
- Davis, R. A., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* 170(1), 142–152.
- Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1), 83–106.
- Drees, H. (2003). Extreme quantile estimation for dependent data with applications to finance. *Bernoulli* 9(4), 617–657.
- Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. *arXiv preprint arXiv:1511.00420*.
- Drees, H. and H. Rootzén (2010). Limit theorems for empirical processes of cluster functionals. *The Annals of Statistics* 38(4), 2145–2186.
- Drees, H. and H. Rootzén (2016). Correction note to “limit theorems for empirical processes of cluster functionals”. *The Annals of Statistics* (to appear).
- Drees, H., J. Segers, and M. Warchol (2015). Statistics for tail processes of markov chains. *Extremes* 18(3), 369–402.
- Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 65(2), 291–306.

- Mikosch, T. and C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *The Annals of Statistics* 28(5), 1427–1451.
- Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical association* 89(428), 1303–1313.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence of Empirical Processes*. New York: Springer.
- Wuertz, D., Y. C. with contribution from Michal Miklovic, C. Boudt, P. Chausse, and others (2013). *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 3010.82.

Chapter 4

A Euclidean Likelihood Estimator for Bivariate Tail Dependence

Miguel de Carvalho^{1 2}, Boris Oumow³, Johan Segers⁴, and Michał Warchol⁴

Manuscript published in *Communication in Statistics - Theory and Methods* (2013) 42(7), 1176–1192.

Abstract

The spectral measure plays a key role in the statistical modeling of multivariate extremes. Estimation of the spectral measure is a complex issue, given the need to obey a certain moment condition. We propose a Euclidean likelihood-based estimator for the spectral measure which is simple and explicitly defined, with its expression being free of Lagrange multipliers. Our estimator is shown to have the same limit distribution as the maximum empirical likelihood estimator of J. H. J. Einmahl and J. Segers, *Annals of Statistics* 37(5B), 2953–2989 (2009). Numerical experiments suggest an overall good performance and identical behavior to the maximum empirical likelihood estimator. We illustrate the method in an extreme temperature data analysis.

¹Universidade Nova de Lisboa, Centro de Matemática e Aplicações, Portugal

²Pontificia Universidade Católica de Chile, Santiago, Chile

³Ecole Polytechnique Fédérale de Lausanne, Department of Statistics, Switzerland

⁴Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Louvain-la-Neuve, Belgium

4.1 Introduction

When modeling dependence for bivariate extremes, only an infinite-dimensional object is flexible enough to capture the ‘spectrum’ of all possible types of dependence. One of such infinite-dimensional objects is the spectral measure, describing the limit distribution of the relative size of the two components in a vector, normalized in a certain way, given that at least one of them is large; see, for instance, [Kotz and Nadarajah \(2000, §3\)](#) and [Beirlant et al. \(2004, §8–9\)](#). The normalization of the components induces a moment constraint on the spectral measure, making its estimation a nontrivial task.

In the literature, a wide range of approaches has been proposed. [Kotz and Nadarajah \(2000, §2–3\)](#) survey many parametric models for the spectral measure, and new models continue to be invented ([Cooley et al., 2010](#); [Boldi and Davison, 2007](#); [Ballani and Schlather, 2011](#)). Here we are mostly concerned with semiparametric and nonparametric approaches. [Einmahl and Segers \(2009\)](#) propose an enhancement of the empirical spectral measure in [Einmahl et al. \(2001\)](#) by enforcing the moment constraints with empirical likelihood methods. A nonparametric Bayesian method based on the censored-likelihood approach in [Ledford and Tawn \(1996\)](#) is proposed in [Guillotte et al. \(2011\)](#).

In this paper we introduce a Euclidean likelihood-based estimator related with the maximum empirical likelihood estimator of [Einmahl and Segers \(2009\)](#). Our estimator replaces the empirical likelihood objective function by the Euclidean distance between the barycenter of the unit simplex and the vector of probability masses of the spectral measure at the observed pseudo-angles ([Owen, 1991, 2001](#); [Crépet et al., 2009](#)). This construction allows us to obtain an empirical likelihood-based estimator which is simple and explicitly defined. Its expression is free of Lagrange multipliers, which not only simplifies computations but also leads to a more manageable asymptotic theory. We show that the limit distribution of the empirical process associated with the maximum Euclidean likelihood estimator measure is the same as the one of the maximum empirical likelihood estimator in [Einmahl and Segers \(2009\)](#). Note that standard large-sample results for Euclidean likelihood methods ([Xu, 1995](#); [Lin and Zhang, 2001](#)) cannot be applied in the context of bivariate extremes.

The paper is organized as follows. In the next section we discuss the probabilistic and geometric frameworks supporting models for bivariate extremes. In [Section 4.3](#) we introduce the maximum Euclidean likelihood estimator for the spectral measure. Large-sample theory is provided in [Section 4.4](#). Numerical experiments are reported in [Section 4.5](#) and an illustration with extreme temperature data is given in [Section 4.6](#). Proofs and some details on a smoothing procedure using Beta kernels are given in the [Appendix 4.7.1](#) and [4.7.2](#), respectively.

4.2 Background

Let $(X_1, Y_1), (X_2, Y_2), \dots$ be independent and identically distributed bivariate random vectors with continuous marginal distributions F_X and F_Y . For the purposes of studying extremal dependence, it is convenient to standardize the margins to the unit Pareto distribution via $X_i^* = 1/\{1 - F_X(X_i)\}$ and $Y_i^* = 1/\{1 - F_Y(Y_i)\}$. Observe that X_i^* exceeds a threshold $t > 1$ if and only if X_i exceeds its tail quantile $F_X^{-1}(1 - 1/t)$; similarly for Y_i^* . The transformation to the unit Pareto distribution serves to measure the magnitudes of the two components according to a common scale which is free from the actual marginal distributions.

Pickands' representation theorem (Pickands, 1981) asserts that if the vector of rescaled, componentwise maxima

$$\mathbf{M}_n^* = \left(\frac{1}{n} \max_{i=1, \dots, n} X_i^*, \frac{1}{n} \max_{i=1, \dots, n} Y_i^* \right),$$

converges in distribution to a non-degenerate limit, then the limiting distribution is a bivariate extreme value distribution G with unit-Fréchet margins given by

$$G(x, y) = \exp \left\{ -2 \int_{[0,1]} \max \left(\frac{w}{x}, \frac{1-w}{y} \right) dH(w) \right\}, \quad x, y > 0. \quad (4.1)$$

The spectral (probability) measure H is a probability distribution on $[0, 1]$ that is arbitrary apart from the moment constraint

$$\int_{[0,1]} w dH(w) = 1/2, \quad (4.2)$$

induced by the marginal distributions $G(z, \infty) = G(\infty, z) = \exp(-1/z)$ for $z > 0$.

The spectral measure H can be interpreted as the limit distribution of the pseudo-angle $W_i = X_i^*/(X_i^* + Y_i^*)$ given that the pseudo-radius $R_i = X_i^* + Y_i^*$ is large. Specifically, weak convergence of $\mathbf{M}_n^*$ to G is equivalent to

$$P[W_i \in \cdot \mid R_i > t] \rightsquigarrow H(\cdot), \quad t \rightarrow \infty. \quad (4.3)$$

The pseudo-angle W_i is close to 0 or to 1 if one of the components X_i^* or Y_i^* dominates the other one, given that at least one of them is large. Conversely, the pseudo-angle W_i will be close to 1/2 if both components X_i^* and Y_i^* are of the same order of magnitude. In case of asymptotic independence, $G(x, y) = \exp(-1/x - 1/y)$, the spectral measure puts mass 1/2 at the atoms 0 and 1, whereas in case of complete asymptotic dependence, $G(x, y) = \exp\{-1/\min(x, y)\}$, the spectral measure H reduces to a unit point mass at 1/2.

Given a sample $(X_1, Y_1), \dots, (X_n, Y_n)$, we may construct proxies for the unobservable pseudo-angles W_i by setting

$$\hat{W}_i = \hat{X}_i^*/(\hat{X}_i^* + \hat{Y}_i^*), \quad \hat{R}_i = \hat{X}_i^* + \hat{Y}_i^*,$$

where $\hat{X}_i^* = 1/\{1 - \hat{F}_X(X_i)\}$ and $\hat{Y}_i^* = 1/\{1 - \hat{F}_Y(Y_i)\}$ and where $\hat{F}_X = \hat{F}_{X,n}$ and $\hat{F}_Y = \hat{F}_{Y,n}$ are estimators of the marginal distribution functions F_X and F_Y . A robust choice for $\hat{F}_X$ and $\hat{F}_Y$ is the pair of univariate empirical distribution functions, normalized by $n+1$ rather than by n to avoid division by zero. In this case, $\hat{X}_i^*$ and $\hat{Y}_i^*$ are functions of the ranks.

For a high enough threshold $t = t_n$, the collection of angles $\{\hat{W}_i : i \in K\}$ with $K = K_n = \{i = 1, \dots, n : \hat{R}_i > t\}$ can be regarded as if it were a sample from the spectral measure H . Parametric or nonparametric inference on H may then be based upon the sample $\{\hat{W}_i : i \in K\}$. Nevertheless, two complications occur:

1. The choice of the threshold t comes into play both through the rate of convergence in (4.3) and through the effective size $|K| = k$ of the sample of pseudo-angles.
2. The standardization via the estimated margins induces dependence between the pseudo-angles, even when the original random vectors (X_i, Y_i) are independent.

For the construction of estimators of the spectral measure, we may thus pretend that $\{\hat{W}_i : i \in I\}$ constitutes a sample from H . However, for the theoretical analysis of the resulting estimators, the two issues above must be taken into consideration. Failure to do so would lead to a wrong assessment of both the bias and the standard errors of estimators of extremal dependence.

4.3 Maximum Euclidean Likelihood Estimator

We propose to use Euclidean likelihood methods (Owen, 2001, pp. 63–66) to estimate the spectral measure. Let $w_1, \dots, w_k \in [0, 1]$ be a sample of pseudo-angles, for example the observed values of the random variables $\hat{W}_i$, $i \in K$, in the previous section, with $k = |K|$. The Euclidean loglikelihood ℓ_E ratio for a candidate spectral measure H supported on $\{w_1, \dots, w_k\}$ and assigning probability mass $p_i = H(\{w_i\})$ to w_i is formally defined as

$$\ell_E(\mathbf{p}) = -\frac{1}{2} \sum_{i=1}^k (kp_i - 1)^2.$$

The Euclidean loglikelihood ratio can be viewed as a Euclidean measure of the distance of $\mathbf{p} = (p_1, \dots, p_k)$ to the barycenter $(k^{-1}, \dots, k^{-1})$ of the $(k-1)$ -dimensional unit simplex. In this sense, the Euclidean likelihood ratio is similar to the empirical loglikelihood ratio

$$\ell(\mathbf{p}) = \sum_{i=1}^k \log(kp_i),$$

which can be understood as another measure of the distance from $\mathbf{p}$ to $(k^{-1}, \dots, k^{-1})$. Note that $\ell_E(\mathbf{p})$ results from $\ell(\mathbf{p})$ by truncation of the Taylor expansion $\log(1+x) = x - x^2/2 + \dots$ and the fact that $p_1 + \dots + p_k = 1$, making the linear term in the expansion disappear.

We seek to maximize $\ell_E(\mathbf{p})$ subject to the empirical version of the moment constraint (4.2). Our estimator $\hat{H}$ for the distribution function of the spectral measure is defined as

$$\hat{H}(w) = \sum_{i=1}^k \hat{p}_i \mathbf{1}(w_i \leq w), \quad w \in [0, 1], \quad (4.4)$$

the vector of probability masses $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_k)$ solving the optimization problem

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}^k} \quad & -\frac{1}{2} \sum_{i=1}^k (kp_i - 1)^2 \\ \text{s.t.} \quad & \sum_{i=1}^k p_i = 1 \\ & \sum_{i=1}^k w_i p_i = 1/2. \end{aligned} \quad (4.5)$$

This quadratic optimization problem with linear constraints can be solved explicitly with the method of Lagrange multipliers, yielding

$$\hat{p}_i = \frac{1}{k} \{1 - (\bar{w} - 1/2)S^{-2}(w_i - \bar{w})\}, \quad i = 1, \dots, k, \quad (4.6)$$

where $\bar{w}$ and S^2 denote the sample mean and sample variance of $w_1, \dots, w_k$, that is,

$$\bar{w} = \frac{1}{k} \sum_{i=1}^k w_i, \quad S^2 = \frac{1}{k} \sum_{i=1}^k (w_i - \bar{w})^2.$$

The weights $\hat{p}_i$ could be negative, but our numerical experiments suggest that this is not as problematic as it may seem at first sight, in agreement with [Antoine et al. \(2007\)](#) and [Crépet et al. \(2009\)](#), who claim that the weights p_i are nonnegative with probability tending to one. The second equality constraint in (4.5) implies that $\hat{H}$ satisfies the moment constraint (4.2), as $\int_{[0,1]} w d\hat{H}(w) = \sum_i w_i \hat{p}_i = 1/2$, which can be easily verified directly.

The empirical spectral measure estimator of [Einmahl et al. \(2001\)](#) and the maximum empirical likelihood estimator of [Einmahl and Segers \(2009\)](#) can be constructed as in (4.5) through suitable changes in the objective function and the constraints. The empirical spectral measure $\hat{H}(w) = \sum_i \hat{p}_i \mathbf{1}(w_i \leq w)$ solves the optimization problem

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}_+^k} \quad & \sum_{i=1}^k \log p_i \\ \text{s.t.} \quad & \sum_{i=1}^k p_i = 1. \end{aligned}$$

yielding $\dot{p}_i = 1/k$ for $i = 1, \dots, k$. In contrast, the maximum empirical likelihood estimator $\dot{H}(w) = \sum_i \ddot{p}_i \mathbf{1}(w_i \leq w)$ has probability masses given by the solution of

$$\begin{aligned} \max_{\mathbf{p} \in \mathbb{R}_+^k} \quad & \sum_{i=1}^k \log p_i \\ \text{s.t.} \quad & \sum_{i=1}^k p_i = 1 \\ & \sum_{i=1}^k w_i p_i = 1/2. \end{aligned} \tag{4.7}$$

By the method of Lagrange multipliers, the solution is given by

$$\ddot{p}_i = \frac{1}{k} \frac{1}{1 + \lambda(w_i - 1/2)}, \quad i = 1, \dots, k,$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier associated to the second equality constraint in (4.7), defined implicitly as the solution to the equation

$$\frac{1}{k} \sum_{i=1}^k \frac{w_i - 1/2}{1 + \lambda(w_i - 1/2)} = 0,$$

see also [Qin and Lawless \(1994\)](#).

4.4 Large-Sample Theory

The maximum Euclidean likelihood estimator $\hat{H}$ in (4.4) can be expressed in terms of the empirical spectral measure $\dot{H}$ given by

$$\dot{H}(w) = \frac{1}{k} \sum_{i=1}^k \mathbf{1}(w_i \leq w), \quad w \in [0, 1].$$

Indeed, $\bar{w}$ and S^2 are just the mean and the variance of $\dot{H}$, and the expression (4.6) for the weights $\hat{p}_i$ can be written as

$$\frac{d\hat{H}}{d\dot{H}}(v) = 1 - (\bar{w} - 1/2) S^{-2} (v - \bar{w}), \quad v \in [0, 1].$$

Integrating out this ‘likelihood ratio’ over $v \in [0, w]$ yields the identity

$$\hat{H}(w) = (\Phi(\dot{H}))(w), \quad w \in [0, 1],$$

where the transformation Φ is defined as follows. Let $\mathbb{D}_\Phi$ be the set of cumulative distribution functions of non-degenerate probability measures on $[0, 1]$. For $F \in \mathbb{D}_\Phi$, the function $\Phi(F)$ on

$[0, 1]$ is defined by

$$(\Phi(F))(w) = F(w) - (\mu_F - 1/2) \sigma_F^{-2} \int_{[0, w]} (v - \mu_F) dF(v), \quad w \in [0, 1].$$

Here $\mu_F = \int_{[0, 1]} v dF(v)$ and $\sigma_F^2 = \int_{[0, 1]} (v - \mu_F)^2 dF(v)$ denote the mean and the (non-zero) variance of F .

We view $\mathbb{D}_\Phi$ as a subset of the Banach space $\ell^\infty([0, 1])$ of bounded, real-valued functions on $[0, 1]$ equipped with the supremum norm $\|\cdot\|_\infty$. The map Φ takes values in $\ell^\infty([0, 1])$ as well. Weak convergence in $\ell^\infty([0, 1])$ is denoted by the arrow ‘ $\rightsquigarrow$ ’ and is to be understood as in [van der Vaart and Wellner \(1996\)](#).

Asymptotic properties of the empirical spectral measure together with smoothness properties of Φ lead to asymptotic properties of the maximum Euclidean likelihood estimator:

- Continuity of the map Φ together with consistency of the empirical spectral measure yields consistency of the maximum Euclidean likelihood estimator (continuous mapping theorem).
- Hadamard differentiability of the map Φ together with asymptotic normality of the empirical spectral measure yields asymptotic normality of the maximum Euclidean likelihood estimator (functional delta method).

The following theorems are formulated in terms of maps $\dot{H}_n$ taking values in $\mathbb{D}_\Phi$. The case to have in mind is the empirical spectral measure $\dot{H}_n(w) = k_n^{-1} \sum_{i \in K_n} \mathbf{1}(\hat{W}_i \leq w)$ with $\{\hat{W}_i : i \in K_n\}$ and $k_n = |K_n|$ as in Section 4.2. In Theorem 3.1 and equation (7.1) of [Einmahl and Segers \(2009\)](#), asymptotic normality of $\dot{H}_n$ is established under certain smoothness conditions on H and growth conditions on the threshold sequence t_n .

Theorem 4.1 (Consistency). *If $\dot{H}_n$ are maps taking values in $\mathbb{D}_\Phi$ and if $\|\dot{H}_n - H\|_\infty \rightarrow 0$ in outer probability for some nondegenerate spectral measure H , then, writing $\hat{H}_n = \Phi(\dot{H}_n)$, we also have $\|\hat{H}_n - H\|_\infty \rightarrow 0$ in outer probability.*

The proof of this and the next theorem is given in Appendix 4.7.1. In the next theorem, the rate sequence r_n is to be thought of as $\sqrt{k_n}$. Let $\mathcal{C}([0, 1])$ be the space of continuous, real-valued functions on $[0, 1]$.

Theorem 4.2 (Asymptotic normality). *Let $\dot{H}_n$ and H be as in Theorem 4.1. If H is continuous and if*

$$r_n(\dot{H}_n - H) \rightsquigarrow \beta, \quad n \rightarrow \infty,$$

in $\ell^\infty([0, 1])$, with $0 < r_n \rightarrow \infty$ and with β a random element of $\mathcal{C}([0, 1])$, then also

$$r_n(\hat{H}_n - H) \rightsquigarrow \gamma, \quad n \rightarrow \infty$$

with

$$\gamma(w) = \beta(w) - \sigma_H^{-2} \int_0^1 \beta(v) dv \int_0^w (1/2 - v) dH(v), \quad w \in [0, 1]. \quad (4.8)$$

Comparing the expression for γ in (4.8) with the one for γ in (4.7) in Einmahl and Segers (2009), we see that the link between the processes β and γ here is the same as the one between the processes β and γ in Einmahl and Segers (2009). It follows that tuning the empirical spectral measure via either maximum empirical likelihood or maximum Euclidean likelihood makes no difference asymptotically. The numerical experiments below confirm this asymptotic equivalence. To facilitate comparisons with Einmahl and Segers (2009), note that our pseudo-angle $w \in [0, 1]$ is related to their radial angle $\theta \in [0, \pi/2]$ via $\theta = \arctan\{w/(1-w)\}$, and that the function f in (4.2) in Einmahl and Segers (2009) reduces to $f(\theta) = (\sin \theta - \cos \theta)/(\sin \theta + \cos \theta) = 2w - 1$.

How does the additional term $\sigma_H^{-2} \int_0^1 \beta(v) dv \int_0^w (1/2 - v) dH(v)$ influence the asymptotic distribution of the maximum Euclidean/empirical estimator? Given the complicated nature of the covariance function of the process β , see (3.7) and (4.7) in Einmahl and Segers (2009), it is virtually impossible to draw any conclusions theoretically. However, Monte Carlo simulations in Einmahl and Segers (2009, §5.1) confirm that the maximum empirical likelihood estimator is typically more efficient than the ordinary empirical spectral measure. These findings are confirmed in the next section.

4.5 Monte Carlo Simulations

In this section, the maximum Euclidean likelihood estimator is compared with the empirical spectral measure and the maximum empirical likelihood estimator by means of Monte Carlo simulations. The comparisons are made on the basis of the mean integrated squared error,

$$\text{MISE}(\cdot) = E \left[\int_0^1 \{\cdot - H(w)\}^2 dw \right].$$

The bivariate extreme value distribution G with logistic dependence structure is defined by

$$G(x, y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^\alpha\}, \quad x, y > 0, \quad (4.9)$$

in terms of a parameter $\alpha \in (0, 1]$. Smaller values of α yield stronger dependence and the limiting cases $\alpha \rightarrow 0$ and $\alpha = 1$ correspond to complete dependence and exact independence, respectively. For $0 < \alpha < 1$, the spectral measure H_α is absolutely continuous with density

$$\frac{dH_\alpha}{dw}(w) = \frac{1}{2}(\alpha^{-1} - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2}, \quad 0 < w < 1. \quad (4.10)$$

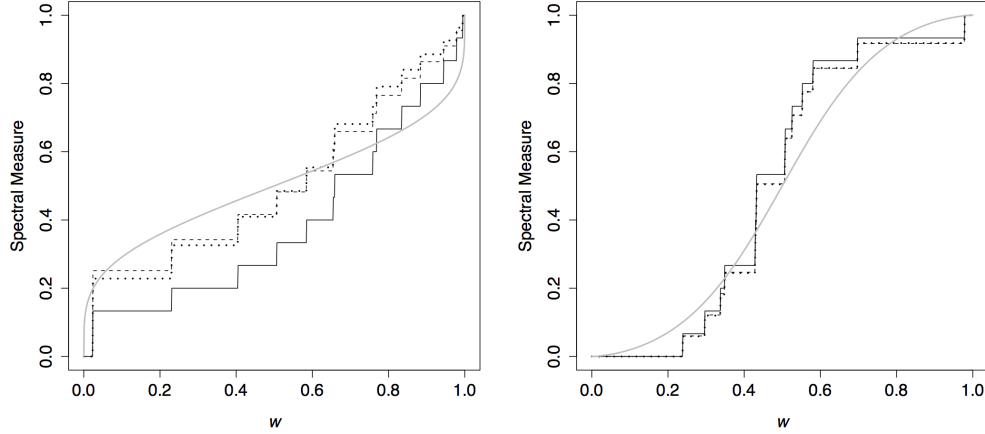


FIGURE 4.1: Examples of trajectories of the empirical spectral measure (solid), the maximum empirical likelihood estimator (dashed), and the maximum Euclidean likelihood estimator (dotted). The solid grey line corresponds to the true spectral measure H_α coming from the bivariate logistic extreme value distribution with parameters $\alpha_I = 0.8$ (left) and $\alpha_{II} = 0.4$ (right).

Here we consider $\alpha_I = 0.8$ and $\alpha_{II} = 0.4$, with stronger extremal dependence corresponding to case II. For each of these two models, 1000 Monte Carlo samples of size 1000 were generated. The thresholds were set at the empirical quantiles of the radius $\hat{R}$ given by $t = 75\%, 75.5\%, \dots, 99.5\%$ for case I and $t = 50\%, 50.5\%, \dots, 99.5\%$ for case II. The margins were estimated parametrically, by fitting univariate extreme value distributions using maximum likelihood.

In Figure 4.1, a typical trajectory of the estimators is shown, illustrating the closeness of the maximum Euclidean and empirical likelihood estimators. The good performance of the maximum Euclidean/empirical spectral measure is confirmed by Figure 4.2. For larger k (lower threshold t), the bias coming from the approximation error in (4.3) is clearly visible.

Numerical experiments in Einmahl and Segers (2009, §5.2) show that the presence of atoms at the endpoints 0 and 1 has an adverse effect on maximum Euclidean/empirical likelihood estimates, and this finding is further confirmed by Guilleotte et al. (2011, §7.1). Indeed, by construction, the pseudo-angles $\hat{W}_i$ will never be exactly 0 or 1. The empirical spectral measure therefore does not assign any mass at 0 and 1, and this situation cannot be remedied by the maximum empirical or Euclidean likelihood estimators, having the same support as the empirical spectral measure.

The weights $\hat{p}_i$ of the maximum Euclidean likelihood estimator can be negative. However, as can be seen from Figure 4.3, the weights tend to be positive overall, except for extremely high thresholds, with the proportion of negative weights being smaller in case I. This suggests that the closer we get to exact independence, the lower the proportion of negative weights.

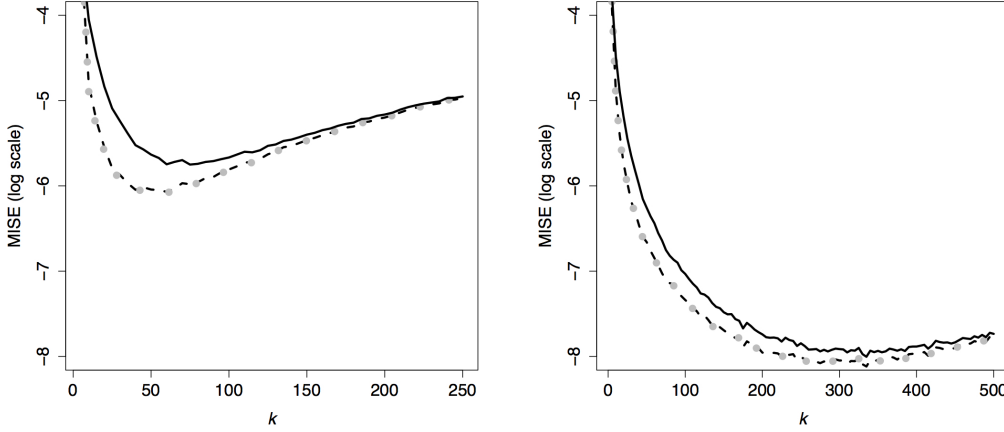


FIGURE 4.2: Logarithms of the mean integrated squared errors of the spectral measure estimates based on 1000 samples of size 1000 from the bivariate logistic extreme value distribution with parameters $\alpha_I = 0.8$ (left) and $\alpha_{II} = 0.4$ (right). The solid, dashed, and dotted lines correspond to the empirical spectral measure, the maximum empirical likelihood estimator, and the maximum Euclidean likelihood estimator, respectively.

4.6 Extreme Temperature Data Analysis

4.6.1 Data Description and Preliminary Considerations

The data were gathered from the Long-term Forest Ecosystem Research database, which is maintained by LWF (Langfristige Waldökosystem-Forschung), and consist of daily average meteorological measurements made in Beatenberg's forest in the canton of Bern, Switzerland. More information on these data can be found at <http://www.wsl.ch>, and for an extensive study see Ferrez et al. (2011). Two time series of air temperature data are available: One in the open field and the other in a nearby site under the forest cover. Our aim is to understand how the extremes in the open relate with those under the canopy; comparison of open-site and below-canopy climatic conditions is a subject of considerable interest in Forestry and Meteorology (Renaud and Rebetez, 2009; Ferrez et al., 2011; Renaud et al., 2011). The raw data are plotted in Figure 4.4, but before we are able to measure extremal dependence of open air and forest cover temperatures we first need to preprocess the data. The preprocessing step is the same as in Ferrez et al. (2011, §3.1) and further details can be found in there.

We consider daily maxima of the residual series that result from removal of the annual cycle in both location and scale, and we then take the residuals at their 98% quantile; hence the threshold boundary is defined as $U = \{(x, y) \in [0, \infty)^2 : x + y = \hat{F}_R^{-1}(0.98) = 105.83\}$, so that there are $|U| = k = 57$ exceedances.

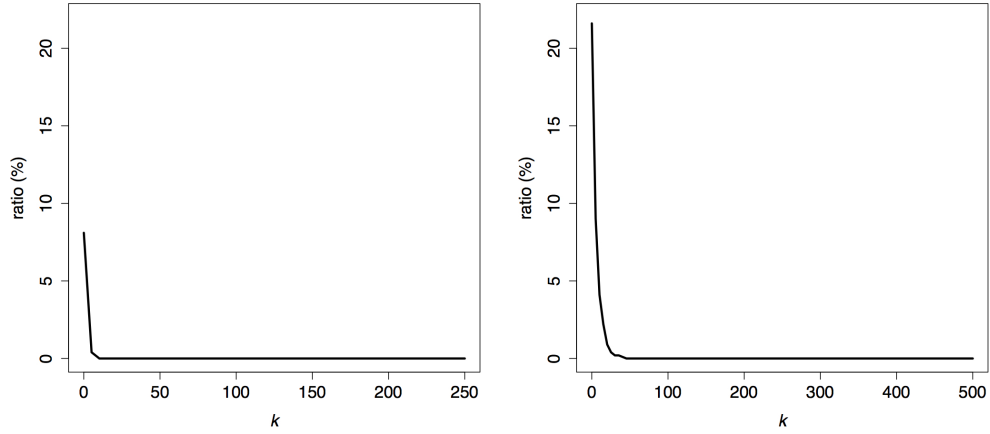


FIGURE 4.3: Proportion of negative weights based on 1000 samples of size 1000 from the bivariate logistic extreme value distribution with parameters $\alpha_I = 0.8$ (left) and $\alpha_{II} = 0.4$ (right).

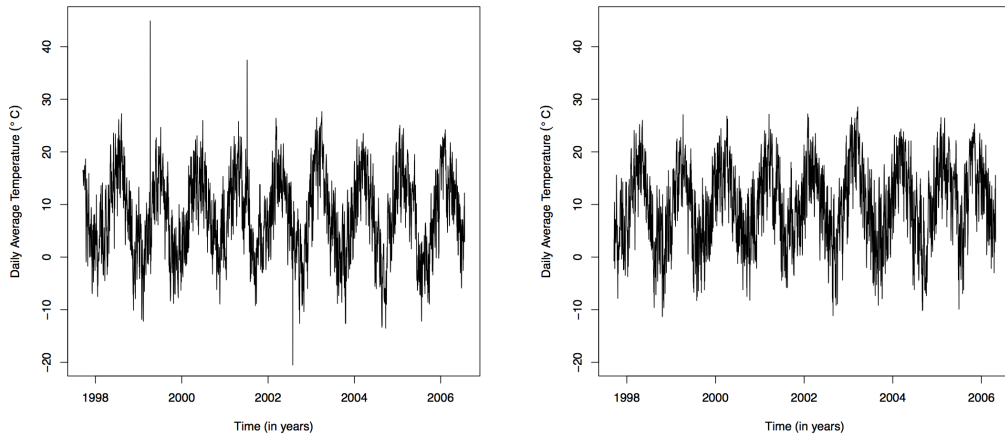


FIGURE 4.4: Daily average air temperatures from the meteorological station in Beatenberg's forest: under the forest cover (left) and in the open field (right).

The dependence between open field and forest cover temperatures can be observed in Figure 4.5, where we plot a log-log scale scatterplot of the unit Fréchet data, and where we note that after log transformation the linearity of the threshold boundary U is perturbed.

Spectral measures are only appropriate for modeling asymptotically dependent data. Although this issue has been already addressed in Ferrez et al. (2011), for exploratory purposes we present in Figure 4.6 the empirical estimates of the dependence coefficients $\chi(u)$ and $\bar{\chi}(u)$, for $0 < u < 1$,

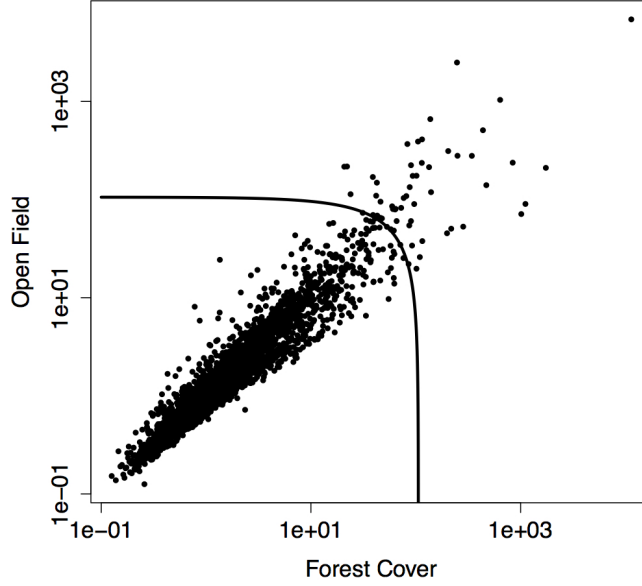


FIGURE 4.5: Scatterplot of air temperature data after transformation to unit Fréchet scale; the solid line corresponds to the boundary threshold in the log-log scale, with both axes being logarithmic.

defined in [Coles et al. \(1999\)](#) as

$$\chi(u) = 2 - \log \frac{\Pr[F_X(X) < u, F_Y(Y) < u]}{\log \Pr[F_X(X) < u]}, \quad \bar{\chi}(u) = \frac{2 \log(1 - u)}{\log \Pr[F_X(X) > u, F_Y(Y) > u]} - 1.$$

Let $\chi = \lim_{u \uparrow 1} \chi(u)$ and $\bar{\chi} = \lim_{u \uparrow 1} \bar{\chi}(u)$. Under asymptotic independence $\chi = 0$ and $\bar{\chi} \in (-1, 1)$, whereas under asymptotic dependence $\chi \in (0, 1]$ and $\bar{\chi} = 1$. Although plots such as Figure 4.6 fail to have a clear-cut interpretation given the large uncertainty entailed in the estimation, the point estimates seem to be consistent with asymptotic dependence as already noticed by [Ferrez et al. \(2011\)](#).

4.6.2 Extremal Dependence of Open Air and Forest Cover Temperatures

We now apply the maximum Euclidean likelihood estimator to measure extremal dependence of open air and forest cover temperatures. The estimated spectral measure is shown in Figure 4.7. All weights are positive, i.e. $\hat{p}_i > 0$, for $i = 1, \dots, 57$.

By construction, the estimate of the spectral measure is discrete. A smooth version which still obeys the moment constraint (4.2) can easily be obtained by smoothing the maximum Euclidean

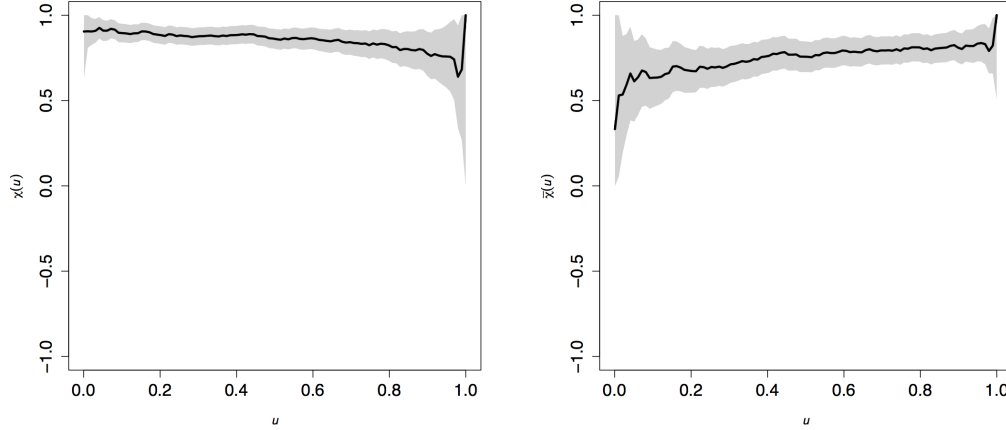


FIGURE 4.6: Empirical estimates and 95% pointwise confidence intervals for $\chi(u)$ (left) and $\bar{\chi}(u)$ (right) as a function of $u \in (0, 1)$.

or empirical likelihood estimator with a Beta kernel. Related ideas are already explored in [Hall and Presnell \(1999\)](#) and [Chen \(1997\)](#). Details are given in [Appendix 4.7.2](#).

A cross-validatory procedure was used to select the bandwidth, yielding a concentration parameter of $\nu \approx 163$. Numerical experiments in [Warchol \(2012\)](#) suggest that convoluting empirical likelihood-based estimators with a Beta kernel yields a further reduction in mean integrated squared error. The Beta kernel even outperforms Chen’s kernel ([Chen, 1999](#)), which is asymptotically optimal under some conditions ([Bouezmarni and Rolin, 2003](#)), but which is unable to conserve the moment constraint.

From the smoothed spectral measure, we obtain an estimate of the spectral density and plug-in estimators for the Pickands dependence function $A(w) = 1 - w + 2 \int_0^w H(v) dv$, $w \in [0, 1]$, and the bivariate extreme value distribution in (4.1). The estimated spectral density is compared with the fit obtained from the asymmetric logistic model

$$H_{\alpha, \psi_1, \psi_2}(w) = \frac{1}{2} \left[1 + \psi_1 + \psi_2 - \{ \psi_1^{1/\alpha} (1-w)^{1/\alpha-1} - \psi_2^{1/\alpha} w^{1/\alpha-1} \} \right. \\ \left. \{ \psi_1^{1/\alpha} (1-w)^{1/\alpha} + \psi_2^{1/\alpha} w^{1/\alpha} \}^{\alpha-1} \right], \quad w \in [0, 1],$$

with parameter estimates $\hat{\psi}_1 = 0.78$ (standard error 0.03), $\hat{\psi}_2 = 0.90$ (0.03) and $\hat{\alpha} = 0.30$ (0.02). The asymmetric logistic model was considered by [Ferrez et al. \(2011\)](#) as the parametric model that achieved the “best overall fit.”

In [Figure 4.7](#) we also plot the smooth spectral measure and corresponding spectral density which are obtained by suitably convoluting the empirical Euclidean spectral measure with a Beta kernel

as described in (4.11) and (4.12). Since more mass concentrated over $1/2$ corresponds to more extremal dependence, and more mass concentrated on 0 and 1 corresponds more independence in the extremes, a rough interpretation for our context is as follows: The lower the shelter ability of the forest, the more mass should be concentrated around $1/2$, whereas higher shelter ability corresponds to the case where the spectral measure gets more mass concentrated at 1; more mass concentrated at 0 suggests relatively more extreme events under the forest cover, suggesting that the forest has the ability to retain heat during extreme events.

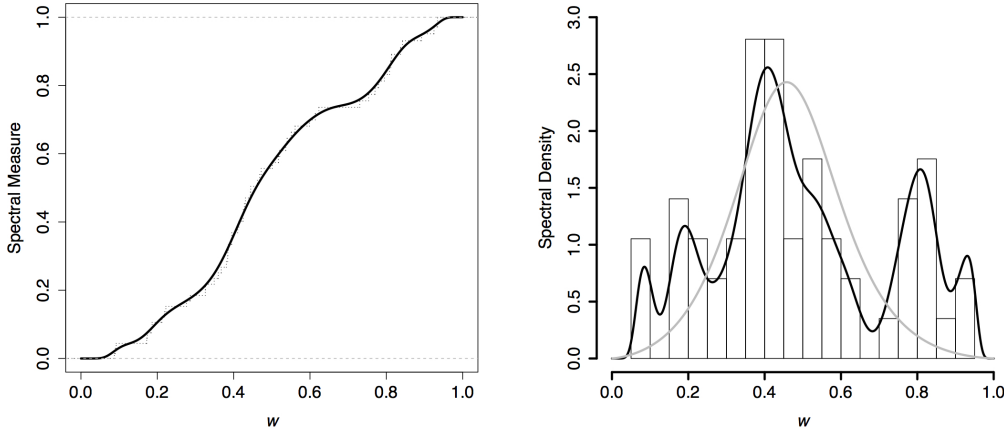


FIGURE 4.7: Estimates of the spectral measure (left) and the spectral density (right). Left: the dotted line corresponds to the empirical Euclidean spectral measure, and the solid line corresponds to its smooth version constructed using (4.12). Right: The solid line corresponds to the smooth spectral density obtained from the Euclidean likelihood weights using (4.11), and the gray line represents the fit from the asymmetric logistic model with $(\hat{\psi}_1, \hat{\psi}_2) = (0.78, 0.90)$ and $\hat{\rho} = 0.30$.

In Figure 4.8 we plot the corresponding Pickands dependence function. More extremal dependence corresponds to lower Pickands dependence functions, and the deeper these are on the right the less frequent are the extreme events under the forest cover relatively to the open field. Our analysis suggests that extreme high temperatures under the forest cover are more frequent than expected from a corresponding parametric analysis. This somewhat surprising finding is already predicted in Ferrez et al. (2011, Fig. 4). The phenomenon may be due to the ability of some forests to retain heat, acting like a greenhouse, or it may be connected with the way that other features of the forest's structure can alter its microclimate (Renaud et al., 2011). Along with the Pickands dependence function, we also plot in Figure 4.8.1 the pseudo-angles which provide further evidence of a marked right skewness.

The joint behavior of temperatures in the open and under forest cover can also be examined from the estimated bivariate extreme value distribution function plotted in Figure 4.8.2, which

was constructed by convoluting the empirical Euclidean spectral measure with a Beta kernel as described in (4.15).

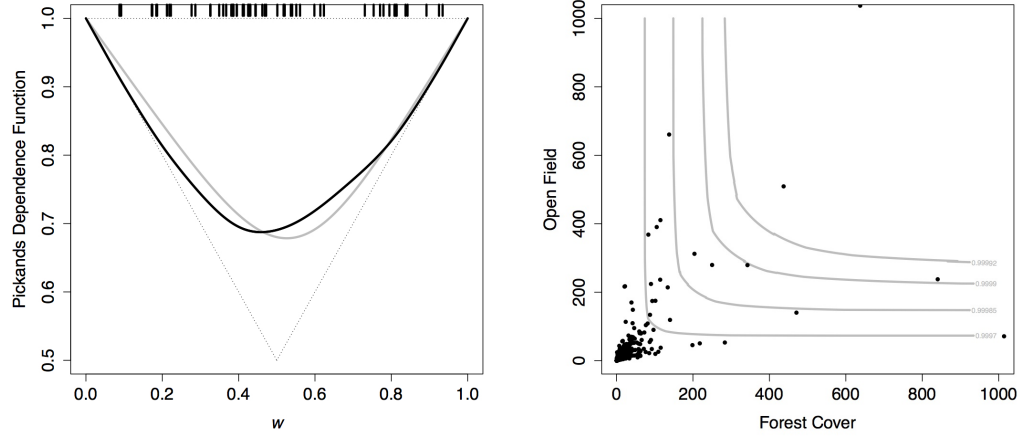


FIGURE 4.8: Estimates of the Pickands dependence function and contours of the bivariate extreme value distribution. Left: The solid line represents the smooth Pickands dependence function obtained from the Euclidean likelihood weights using (4.14), and the gray line represents the fit from the asymmetric logistic model with $(\hat{\psi}_1, \hat{\psi}_2) = (0.78, 0.90)$ and $\hat{\rho} = 0.30$. The pseudo-angles are presented at the top. Right: The solid gray line represents the smooth bivariate extreme value distribution obtained from the Euclidean likelihood weights using (4.15). Air temperature data are plotted on the unit Fréchet scale.

Acknowledgements

We thank Anthony Davison, Vanda Inácio, Feridun Turkman, and Jacques Ferrez for discussions and we thank the editors and anonymous referees for helpful suggestions and recommendations, that led to a significant improvement of an earlier version of this article. Miguel de Carvalho's research was partially supported by the Swiss National Science Foundation, CCES project EXTREMES, and by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA). Johan Segers's research was supported by IAP research network grant No. P6/03 of the Belgian government (Belgian Science Policy) and by contract No. 07/12/002 of the Projet d'Actions de Recherche Concertées of the Communauté française de Belgique, granted by the Académie universitaire Louvain.

4.7 Appendix A

4.7.1 Proofs

Proof of Theorem 4.1. Let $F_n, F \in \mathbb{D}_\Phi$. By Fubini's theorem,

$$\begin{aligned} \int_{[0,w]} v \, dF(v) &= w F(w) - \int_0^w F(v) \, dv, \quad w \in [0, 1], \\ \int_{[0,1]} v^2 \, dF(v) &= 1 - \int_0^1 2v F(v) \, dv, \end{aligned}$$

and similarly for F_n . It follows that $\|F_n - F\|_\infty \rightarrow 0$ implies that $\mu_{F_n} \rightarrow \mu_F$, $\sigma_{F_n}^2 \rightarrow \sigma_F^2$, and $\int_{[0,w]} v \, dF_n(v) \rightarrow \int_{[0,w]} v \, dF(v)$ uniformly in $w \in [0, 1]$. Hence $\|\Phi(F_n) - \Phi(F)\|_\infty \rightarrow 0$. Therefore, the map $\Phi : \mathbb{D}_\Phi \rightarrow \ell^\infty([0, 1])$ is continuous. The lemma now follows from the fact that $\Phi(H) = H$ together with the continuous mapping theorem, see Theorem 1.9.5 in [van der Vaart and Wellner \(1996\)](#). $\square$

Proof of Theorem 4.2. Write

$$\beta_n = r_n(\dot{H}_n - H), \quad \gamma_n = r_n(\hat{H}_n - H).$$

Let $\mathbb{D}_n$ denote the set of functions $f \in \ell^\infty([0, 1])$ such that $H + r_n^{-1}f$ belongs to $\mathbb{D}_\Phi$. Since $H + r_n^{-1}\beta_n = \dot{H}_n$ takes values in $\mathbb{D}_\Phi$, it follows that β_n takes values in $\mathbb{D}_n$. Define $g_n : \mathbb{D}_n \rightarrow \ell^\infty([0, 1])$ by

$$g_n(f) = r_n\{\Phi(H + r_n^{-1}f) - H\}.$$

Observe that

$$g_n(\beta_n) = r_n\{\Phi(\dot{H}_n) - H\} = \gamma_n.$$

Further, define the map $g : \mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ by

$$(g(f))(w) = f(w) - \sigma_H^{-2} \int_0^1 f(v) \, dv \int_0^w (1/2 - v) \, dH(v), \quad w \in [0, 1].$$

A straightforward computation shows that if $f_n \in \mathbb{D}_n$ is such that $\|f_n - f\|_\infty \rightarrow 0$ for some $f \in \mathcal{C}([0, 1])$, then $\|g_n(f_n) - g(f)\|_\infty \rightarrow 0$; note in particular that $f_n(1) = 0$ and thus also $f(1) = 0$. The extended continuous mapping theorem ([van der Vaart and Wellner, 1996](#), Theorem 1.11.1) implies that

$$\gamma_n = g_n(\beta_n) \rightarrow g(\beta) = \gamma, \quad n \rightarrow \infty,$$

as required. Note that we have actually shown that Φ is Hadamard differentiable at H tangentially to $\mathcal{C}([0, 1])$ with derivative given by $\Phi'_F = g$. The result then also follows from the functional delta method. $\square$

4.7.2 Beta-Kernel Smoothing of Discrete Spectral Measures

We only consider the case of the empirical Euclidean spectral measure using a Beta kernel, but the same applies to the empirical likelihood spectral measure by replacing the $\hat{p}_i$ with $\tilde{p}_i$ in (4.3). The smooth Euclidean spectral density is thus defined as

$$\tilde{h}(w) = \sum_{i=1}^k \hat{p}_i \beta\{w; w_i \nu, (1 - w_i) \nu\}, \quad w \in (0, 1), \quad (4.11)$$

where $\nu > 0$ is the concentration parameter (inverse of the squared bandwidth, to be chosen via cross-validation) and $\beta(w; p, q)$ denotes the Beta density with parameters $p, q > 0$. The corresponding smoothed spectral measure is defined as

$$\tilde{H}(w) = \int_0^w \tilde{h}(v) dv = \sum_{i=1}^k \hat{p}_i \mathcal{B}\{w; w_i \nu, (1 - w_i) \nu\}, \quad w \in [0, 1], \quad (4.12)$$

where $\mathcal{B}(w; p, q)$ is the regularized incomplete beta function, with $p, q > 0$. Since

$$\int_0^1 w \tilde{h}(w) dw = \sum_{i=1}^k \hat{p}_i \left\{ \frac{\nu w_i}{\nu w_i + \nu(1 - w_i)} \right\} = \sum_{i=1}^k \hat{p}_i w_i = 1/2, \quad (4.13)$$

the moment constraint is satisfied. Plug-in estimators for the Pickands dependence function and the bivariate extreme value distribution follow directly from

$$\tilde{A}(w) = 1 - w + 2 \sum_{i=1}^k \hat{p}_i \int_0^w \mathcal{B}\{u; w_i \nu, (1 - w_i) \nu\} du, \quad w \in [0, 1], \quad (4.14)$$

$$\tilde{G}(x, y) = \exp \left\{ -\frac{2}{k} \sum_{i=1}^k \hat{p}_i \int_0^1 \max \left(\frac{u}{x}, \frac{1-u}{y} \right) \beta\{u; w_i \nu, (1 - w_i) \nu\} du \right\}, \quad x, y > 0. \quad (4.15)$$

Bibliography

- Antoine, B., Bonnal, H., Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *J. Econometrics* 138(2): 461–487.
- Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* 85(3): 535–547.
- Ballani, F., Schlather, M. (2011). A construction principle for multivariate extreme value distributions. *Biometrika* 98(3): 633–645.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. New York: Wiley.
- Boldi, M.-O., Davison, A. C. (2007). A mixture model for multivariate extremes. *J. R. Statist. Soc. B* 69(2): 217–229.
- Bouezmarni, T., Rolin, J.-M. (2003). Consistency of the beta kernel density function estimator. *Canadian J. Statist.* 31(1): 89–98.
- Chen, S. X. (1997). Empirical likelihood-based kernel density estimation. *Austral. J. Statist* 39(1): 47–56.
- Chen, S. X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* 31(2): 131–145.
- Coles, S. G., Heffernan, J., Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes* 2(4): 339–365.
- Cooley, D., Davis, R. A., Naveau, P. (2010). The pairwise beta distribution: A flexible parametric multivariate model for extremes. *J. Mult. Anal.* 101(9): 2103–2117.
- Crépet, A., Harari-Kermadec, H., Tressou, J. (2009). Using empirical likelihood to combine data: Application to food risk assessment. *Biometrics* 65(1): 257–266.
- Einmahl, J. H. J., de Haan, L., Piterbarg, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* 29(5): 1401–1423.

- Einmahl, J. H. J., Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* 37(5B): 2953–2989.
- Ferre, J. F., Davison, A. C., Rebetez, M. (2011). Extreme temperature analysis under forest cover compared to an open field. *Agric. Forest Meteo.* 151(7): 992–1001.
- Guillotte, S., Perron, F., Segers, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Statist. Soc. B* 73(3): 377–406.
- Hall, P., Presnell, B. (1999). Density estimation under constraints. *J. Comput. Graph. Statist.* 8(2): 259–277.
- Kotz, S., Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.
- Ledford, A. W., Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* 83(1): 169–187.
- Lin, L., Zhang, R. (2001). Blockwise empirical Euclidean likelihood for weakly dependent processes. *Statist. Prob. Lett.* 53(2): 143–152.
- Owen, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* 19(4): 1725–1747.
- Owen, A. B. (2001). *Empirical Likelihood*. Boca Raton: Chapman and Hall.
- Pickands, J., I. (1981). Multivariate extreme value distributions. *Bulletin Int. Statist. Inst., Proc. 43rd Sess.* (Buenos Aires), pp. 859–878. Voorburg, Netherlands: ISI.
- Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22(1): 300–325.
- Renaud, V., Innes, J. L., Dobbertin, M., Rebetez, M. (2011). Comparison between open-site and below-canopy climatic conditions in Switzerland for different types of forests over 10 years (1998–2007). *Theor. Appl. Climatol.* 105(1–2): 119–127.
- Renaud, V., Rebetez, M. (2009). Comparison between open-site and below-canopy climatic conditions in Switzerland during the exceptionally hot summer of 2003. *Agric. Forest Meteo.* 149: 873–880.
- Van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.
- Warchol, M. (2012). *Smoothing Methods for Bivariate Extremes*. Joint MSc thesis, Ecole Polytechnique Fédérale de Lausanne and Uniwersytet Jagielloński.
- Xu, L. (1995). Large sample properties of the empirical Euclidean likelihood estimation for semiparametric model. *Chinese J. Appl. Prob. Statist.* 10: 344–352.

Chapter 5

Discussion

Optimal weights for the mixture estimator

In Chapter 2 we propose estimators for the spectral tail process exploiting additional information that comes from the time change formula. For well-chosen weights, convex combinations of the forward and the backward estimators can achieve a lower asymptotic variance than each of the estimators individually. Since the expression for the asymptotic covariance of the forward and backward estimators is intractable, it remains an open issue how to choose the mixture weights in order to minimize the asymptotic variance. Possibly more efficient solutions exist than the one presented in Chapter 2.

The signed rank transformation

Exploiting the time change property and using the backward estimator requires the value of the index of regular variation α . However, in most applications this index is unknown. In Chapter 2 we propose two approaches to deal with this issue. First, estimate the index of regular variation separately, for instance, by the Hill-type estimator, and plug in the estimated values of α in the backward and mixture estimators. Second, apply the signed rank transformation to ensure that, after transformation, $\alpha = 1$. The numerical simulations in Section 2.6 indicate that the mixture estimator based on the rank-transformed data performs better than the plug-in version. However, the asymptotic properties for the backward estimator have been worked out only in the plug-in setting. The large sample theory for the estimators based on the rank-transformed data as well as other applications of the signed rank transformation could be explored more deeply.

The signed rank transformation can be potentially useful when modeling cross-sectional serial extremal dependence within multivariate time series. In this context, the margins need to be transformed to a common scale for a meaningful analysis. For this purpose, if the focus is on

either positive or negative extremes, a standard rank transformation can be applied. Conversely, the signed rank transformation standardizes the margins (up to the tail balance index p) covering both lower and upper tails simultaneously.

Markov spectral tail chain conditions for the copula Markov models

The focus in Chapter 2 is on one particular family of spectral tail processes, i.e., the Markov spectral tail chain. Such spectral tail processes typically arise when $(X_t)_{t \in \mathbb{Z}}$ is a stationary, first-order, and regularly varying Markov chain. In Section 2.5 we present two examples of regularly varying Markov chains: copula-based Markov processes and solutions to stochastic recurrence equations. For the latter, the Markov spectral tail chain exists and the distribution of the spectral tail process at lag $t = 1$ can be characterized in a simple way. For the former, although we show the link between the copula and the distribution of the spectral tail process at lag $t = 1$, we do not provide a proof for the existence of the Markov tail chain. Conditions for a Markov chain to have a Markov tail chain are given in Janßen and Segers (2014). A particular family of copulas satisfying those conditions could be specified in more detail.

More precisely, for the copula Markov models, as considered in Proposition 2.9, we can write

$$X_t = Q(\epsilon_t \mid X_{t-1}), \quad t = 1, 2, \dots,$$

where $\epsilon_1, \epsilon_2, \dots$ are independent and uniformly distributed on $[0, 1]$ and independent of X_0 , and for all $x \in \mathbb{R}$, $u \mapsto Q(u \mid x)$ is the quantile function of the conditional law of X_1 given $X_0 = x$, i.e., the quantile function of

$$y \mapsto \mathbb{P}[X_1 \leq y \mid X_0 = x] = \dot{C}_1(F(x), F(y)).$$

For a copula Markov model specified by the marginal distribution F and copula C , with partial derivative $\dot{C}_1$, one can verify Condition 2.2 in Janßen and Segers (2014) with $\Phi(X_{t-1}, \epsilon_t) = Q(\epsilon_t \mid X_{t-1})$ to show whether the Markov tail chain exists.

Time reversibility at extreme levels

In Section 2.7 we visually test the compatibility of the forward and the backward estimates. In particular, we compare the estimated distributions of (Θ_0, Θ_1) and (Θ_0, Θ_{-1}) , which provides insight into short-range dependence between extreme observations. The comparison may hint on the presence or absence of time reversibility patterns leading to improvement in model selection. Formal tests for time reversibility regarding the bulk of the distribution are developed for instance in Beare and Seo (2014); Chen et al. (2000). A similar rigorous testing procedure, but with respect to extreme values, could be of interest as well.

Markov spectral tail chain at higher lags

In Chapter 2 we propose estimators for the distribution of the Markov tail chain at lags $t \in \{-1, 0, 1\}$. The estimation procedure is generalized in Chapter 3 to higher lags and to time series that do not necessarily have Markov tail chains. The generalized estimators from Chapter 3 can be readily applied in the Markov tail chain context at higher lags. However, potentially more efficient estimators exist that take advantage of the spectral tail process having the simplified structure of the multiplicative random walk. In particular, for a positive time series with a Markov tail chain, we have $\mathcal{L}(\Theta_2) = \mathcal{L}(\Theta_1 \Theta_1^*)$, where Θ_1^* is an independent copy of Θ_1 . This alternative representation of the law of Θ_2 suggests the following estimator:

$$\hat{F}_n^{(f, \Theta_2)}(x) = \frac{\sum_{i=1}^n \sum_{i < j} \mathbf{1}\left(\frac{X_{i+1} X_{j+1}}{X_i X_j} > x, X_i > u_n, X_j > u_n\right)}{\sum_{i=1}^n \sum_{i < j} \mathbf{1}(X_i > u_n, X_j > u_n)}.$$

Preliminary numerical simulations suggest that such estimator can have a smaller asymptotic variance and provide smoother estimates along x than its classical counterpart. Large sample properties could be derived via U-statistics theory or the functional delta method. Calculations of the asymptotic variance should lead to a closed expression (at least for $t = 2$).

Bootstrapped spectral tail process

In Section 3.2.2 we employ two different bootstrap schemes which yield confidence intervals for the pre-asymptotic spectral tail process: the stationary bootstrap and the multiplier bootstrap. Both were first applied in the extremogram context (Davis et al., 2012; Drees, 2015). In the numerical simulations presented in Section 3.4.2 and as reported by Drees (2015), the multiplier bootstrap outperforms the stationary bootstrap. It is not clear why the stationary bootstrap produces relatively low coverage probabilities, especially for higher lags (see Figure 3.2). This could be studied in more detail and potentially lead to further improvement of the two methods.

Tail processes of multivariate time series

Probabilistic theory for the tail processes of multivariate time series is developed in Basrak and Segers (2009). Although statistical methodology presented in Chapter 3 is restricted to univariate time series, it could be extended to a multivariate setting. Large sample theory could be developed, if not via cluster functionals, then in a classical way like in Davis and Mikosch (2009) in the extremogram context. This can be done by writing the spectral tail process in terms of the extremogram. Statistical methodology dedicated to multivariate time series significantly broadens the scope of possible applications. It is of interest, for instance in quantitative risk management, to draw inference on serial extremal dependence across coordinates. The theory

for the tail processes could be extended to max-stable spatial processes framework as well. Such processes attracted some attention in extreme value literature. See for instance [Cho et al. \(2016\)](#), where the extremogram is defined in the spatial domain with the data being observed on a lattice or at location distributed as a Poisson point process in d -dimensional space.

Generalized empirical likelihood-based estimator

In Chapter 4 we propose a simple empirical likelihood-based estimator for the spectral measure, whose asymptotic efficiency is comparable to the empirical likelihood spectral measure of [Einmahl and Segers \(2009\)](#). The fact that our estimator has the same limit distribution as the empirical likelihood spectral measure, suggests that a more general result may hold for other members of the Cressie–Read class, of which these estimators are particular cases, similarly to what was established by [Baggerly \(1998\)](#) in a context different from ours. We focus on the spectral measure defined over the L_1 -norm, but only for a matter of simplicity, and there is no problem in defining our estimator for the spectral measure defined over the L_p -norm, with $p \in [1, \infty]$. For real data applications smooth versions of the empirical estimator may be preferred, but these can be readily constructed by suitably convoluting the weights of our empirical likelihood-based method with a kernel on the simplex.

The maximum Euclidean likelihood estimator under serial dependence

In Section 4.6 we apply the developed methodology to study the extremal dependence of open air and forest cover temperatures. The data consists of two time series of daily average air temperature at both locations. The data is preprocessed as in [Ferrez et al. \(2011, §3.1\)](#), where the seasonal variation is ruled out with a periodic cubic spline. The preprocessed data are stationary series that do not display long-term dependence. However, the influence of short range dependence is not considered in the case study.

The maximum Euclidean likelihood estimator as well as the maximum empirical likelihood estimator are developed under the iid setting. This assumption is too restrictive but potentially could be relaxed to various weak dependence concepts. For instance, it is shown in [Bücher and Volgushev \(2013\)](#) that similar empirical processes converge for sequentially dependent data. Presence of serial dependence can influence the limiting distribution of the estimators but the consistency result should still hold. Further generalizations of the large sample theory could be investigated in more detail.

Applications

In Sections 2.7 and 3.5 we apply the developed statistical methodology to financial time series. The tools prove to be useful in detecting short-term stock market overreaction and serial extremal dependence. An R software package could be developed to make the methodology described in the thesis more convenient to apply. Currently, selected estimators are embedded in the web applications available at https://extreme.shinyapps.io/serial_extremal_dependence and https://extreme.shinyapps.io/bivariate_tail_dependence, which correspond to serial extremal dependence (Chapter 3) and bivariate tail dependence modeling (Chapter 4) respectively. The online applications are currently in the demo versions and serve more as a promotion tool. More features could be added leading to an extended dashboard for extreme risk monitoring, see for example the V-lab of Robert Engle (<http://vlab.stern.nyu.edu>). Not only can such online applications promote research but also help to exchange knowledge between academia and industry. The demand for quantitative expertise focused on extreme scenarios is growing in the financial industry. Extreme value techniques are already playing an important role in this field, see for instance the RiskLab lead by Paul Embrechts (<http://www.risklab.ch>) or the Real World Risk Institute (<http://realworldrisk.com>) of Nassim Nicholas Taleb.

Bibliography

- Baggerly, K. A. (1998). Empirical likelihood as a goodness-of-fit measure. *Biometrika* 85(3), 535–547.
- Basrak, B. and J. Segers (2009). Regularly varying multivariate time series. *Stochastic Processes and Their Applications* 119(4), 1055–1080.
- Beare, B. K. and J. Seo (2014, 4). Time irreversible copula-based markov models. *Econometric Theory FirstView*, 1–38.
- Bücher, A. and S. Volgushev (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis* 119, 61–70.
- Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions. *Journal of Econometrics* 95(1), 199–218.
- Cho, Y. B., R. A. Davis, and S. Ghosh (2016). Asymptotic properties of the empirical spatial extremogram. *Scandinavian Journal of Statistics*.
- Davis, R. and T. Mikosch (2009). The extremogram: a correlogram for extreme events. *Bernoulli* 15(4), 977–1009.
- Davis, R. A., T. Mikosch, and I. Cribben (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* 170(1), 142 – 152.
- Drees, H. (2015). Bootstrapping empirical processes of cluster functionals with application to extremograms. *arXiv preprint arXiv:1511.00420*.
- Einmahl, J. H. J. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics* 37, 2953–2989.
- Janßen, A. and J. Segers (2014). Markov tail chains. *Advances in Applied Probability* 51(4), 1133–1153.