Estimation of Limiting Conditional Probabilities for Regularly Varying Time Series

by

Herve Dimy Anguima Ibondzi

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the Master degree in
Mathematics and Statistics

Department of Mathematics and Statistics
Faculty of Sciences
University of Ottawa

Abstract

In this thesis we are concerned with estimation of clustering probabilities for univariate heavy tailed time series. We employ functional convergence of a bivariate tail empirical process to conclude asymptotic normality of an estimator of the clustering probabilities. Theoretical results are illustrated by simulation studies.

Acknowledgements

This thesis is dedicated to the memory of my little sister Ornella Joslette Amboulou Dimy. Thanks to God Almighty Lord Jesus Christ. Thanks to Prof. Rafal Kulik, members of the Department of Mathematics and Statistics at the University of Ottawa, the African Institute for Mathematical Sciences, as well as Dr. J. Carolina Ödman-Govender, Prof. Barry Green and Prof. Jan Govaerts.

Contents

Introduction

A stochastic process is a collection of random variables. Realisations of the stochastic process arranged according to time are called *time series*. We will denote a time series process by $\{X_t\}$, $t \in T$, where T is a set of continuous or discrete indices.

Time series observations occur in a wide range of areas in which studies are done to model a stochastic mechanism of certain series in time and to predict future values based on the history of that series. Commonly we use time series in business with observations of weekly interest rates or daily closing stock prices: in agriculture, with crop and livestock production; in meteorology, with daily high and low temperatures, hourly wind speed; social sciences with birthrates and deathrates; in health care, with blood pressure measurements. In our thesis we focus on models that are used in finance.

Financial data have, among others, a particular feature: large values of such series cluster (they appear together). A goal of this thesis is to estimate *clustering probabilities* using an empirical estimator. In order to do this, we proceed as follows: In Chapter 1 we describe regular variation as a tool to model heavy tails. In Chapter 2, we describe different strictly stationary processes such as linear models, stochastic volatility, ARCH and GARCH models. Also we discuss some tools for time series such as mixing and state those properties for the models considered. Also, we summarize some results on the central limit theorem (CLT) and tightness of stochastic processes. These tools are needed to prove asymptotic normality of our estimator.

In Chapter 3, we discuss the classical empirical process for the iid case and for the dependent case. *Functional convergence* to a Brownian bridge is re-proven there.

We introduce an univariate tail empirical process in Chapter 4. In the iid case we use the regular variation property and Lindeberg's CLT to prove asymptotic normality. For the dependent case, we use an $\alpha-$mixing property with geometric rates to separate the sample into blocks in order to prove the CLT by using Lindeberg's conditions again.

In Chapter 5, we consider an estimator of the limiting conditional probabilities. To do so, we introduce a bivariate tail empirical process. We prove its asymptotic normality.

It leads to the formulation of the appropriate estimator.

In Chapter 6, we perform some simulation studies. We consider an AR(1) model with Pareto innovations in order to estimate the value of the tail index α using a Hill estimator $\widehat{\alpha}_{Hill}$. Then, we perform an analysis of the estimator for the limiting conditional probabilities. We also consider a GARCH model.

We state final conclusions in Chapter 7. `R`-codes are given in the last chapter.

We finish this introduction by noting that our results included in this thesis have been proven in the literature in some form before. For example, a version of the bivariate tail empirical process was considered in [8] without proving tightness. However we are not aware of numerical studies on clustering probabilities.

Chapter 1

Regular Variation

In this section we briefly discuss regular variation and multivariate regular variation. This material is based on [9] and [7].

A random variable is regularly varying with index $-\alpha$ if

$$\mathbb{P}(|X| > x) = x^{-\alpha} L(x) \,, \tag{1.1}$$

where $L(\cdot)$ is slowly varying at infinity. In particular, for all $y \geq 1$,

$$\lim_{x\to\infty} \mathbb{P}(|X| > yx \mid |X| > x) = y^{-\alpha} \,, \qquad y \geq 1 \,.$$

If X is a nonnegative random variable with a marginal distribution F, then the regular variation (1.1) is equivalent to

$$\lim_{x\to\infty} \frac{\overline{F}(yx)}{\bar{F}(x)} = y^{-\alpha} =: \psi(y) \,, \qquad y > 0 \,, \tag{1.2}$$

where $\overline{F}(x) = \mathbb{P}(X > x)$. We note also that (1.2) is equivalent to

$$\lim_{n\to\infty} \frac{\overline{F}(c_n y)}{\bar{F}(c_n)} = y^{-\alpha} = \psi(y) \,,$$

where c_n is an arbitrary sequence such that $c_n \to \infty$ as $n \to \infty$.

Example 0.1. *A standard Pareto distribution is defined by $\overline{F}(x) = x^{-\alpha}$ for $x > 1$. Hence, for $y > 0$,*

$$\overline{F}(yx) = \mathbb{P}(X_0 > xy) = y^{-\alpha} x^{-\alpha},$$

therefore

$$\lim_{x\to\infty} \frac{\overline{F}(xy)}{\overline{F}(x)} = \frac{y^{-\alpha} x^{-\alpha}}{x^{-\alpha}} = y^{-\alpha} = \psi(y).$$

Define a sequence a_n by

$$\lim_{n\to\infty} n\mathbb{P}(|X| > a_n) = 1$$

and the measure $\nu_{\alpha,p}$ on the Borel sigma-field $\mathcal{B}(\bar{\mathbb{R}} \setminus \{0\})$ by

$$\nu_{\alpha,p}(dy) = \alpha \left[py^{-\alpha-1}1_{\{y>0\}} + (1-p)(-y)^{-\alpha-1}1_{\{y<0\}} \right] ,$$

where

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

and $p = 1$ if X is positive. Then, the regular variation of X is equivalent to the vague convergence on $\mathcal{B}(\bar{\mathbb{R}} \setminus \{0\})$

$$n\mathbb{P}(a_n^{-1}X \in \cdot) \xrightarrow{\text{v}} \nu_{\alpha,p} , \qquad n \to \infty .$$

This means that for any set $B \in \mathcal{B}(\bar{\mathbb{R}} \setminus \{0\})$ which is bounded away from zero and such that $\nu_{\alpha,p}(\partial B) = 0$,

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1}X \in B) = \nu_{\alpha,p}(B) . \tag{1.3}$$

Now, having (1.3), we can extend the univariate regular variation to a multivariate case.

Definition 0.1. *A vector $\boldsymbol{X} = (X_0, \ldots, X_h)$ in $\mathbb{R}^{h+1}$ is (multivariate) regularly varying with index $-\alpha$ if there exists a non zero Radon measure $\boldsymbol{\nu}$ on $\overline{\mathbb{R}}^{h+1} \setminus \{\boldsymbol{0}\}$ that does not place a mass at lines going through infinity, and a sequence of constants $c_n \uparrow \infty$ such that*

$$n\mathbb{P}(c_n^{-1}\boldsymbol{X} \in \cdot) \xrightarrow{\text{v}} \boldsymbol{\nu}$$

on $\mathcal{B}(\overline{\mathbb{R}}^{h+1} \setminus \{\boldsymbol{0}\})$. The measure $\boldsymbol{\nu}$ is called an exponent measure.

We should explain the assumption ”does not place a mass at lines going through infinity”. In dimension one it means that $\boldsymbol{\nu}(\{\infty\}) = 0$. In dimension two it means that $\boldsymbol{\nu}(\mathbb{R} \times \{\infty\}) = 0$ and $\boldsymbol{\nu}(\{\infty\} \times \mathbb{R}) = 0$.

Definition 0.2. *We say that a strictly stationary sequence is regularly varying if all its multivariate distributions are regularly varying in the sense of Definition 0.1.*

In particular, for $u_0 \geq 1$ and $u_1, \ldots, u_h > 0$ we have

$$\lim_{x\to\infty} \mathbb{P}(X_0 > u_0 x, X_1 > u_1 x, \ldots, X_h > u_h x \mid |X_0| > x)$$
$$= \boldsymbol{\nu}((u_0, \infty] \times (u_1, \infty] \times \cdots \times (u_h, \infty]) \; .$$

Also,

$$\lim_{x\to\infty} \mathbb{P}(X_h > u_h x \mid |X_0| > x) = \frac{\boldsymbol{\nu}((1,\infty] \times \mathbb{R}^{h-1} \times (u_h, \infty])}{\boldsymbol{\nu}((1,\infty] \times \mathbb{R}^h)}. \tag{1.4}$$

Definition 0.3 (Extremal independence, extremal dependence)**.** *Let $\boldsymbol{X}$ be a regularly varying random vector in $\mathbb{R}^{h+1}$. It will be said to be extremally independent if its exponent measure $\boldsymbol{\nu}$ is concentrated on the axes (hence (1.4) becomes zero then). It will be said extremally dependent if the exponent measure is not concentrated on the axes.*

Intuitively speaking, in case of extremal independence *extremes do not cluster*, that is a large value is not followed by an observation with the same order of magnitude. In case of extremal dependence *extremes do cluster*, that is a large value is followed by another large value. In this particular case, the limiting conditional probability in (1.4) is positive, whereas in case of extremal independence the limiting conditional probability is zero.

Example 0.2. *Assume that random variables X_0, X_h, $h \neq 0$, are independent and regularly varying. Then*

$$\lim_{x\to\infty} \mathbb{P}(X_h > x \mid |X_0| > x) = \lim_{x\to\infty} \frac{\mathbb{P}(X_h > x, |X_0| > x)}{\mathbb{P}(|X_0| > x)} = 0 \; .$$

Example 0.3. *Examples of strictly stationary sequences that are regularly varying with extremal dependence include GARCH processes, whereas the Stochastic Volatility model is regularly varying with extremal independence. Those examples will be discussed later in the thesis.*

The main goal of this thesis is to estimate the limiting conditional probabilities defined in (1.4).

Chapter 2

Stationary Time Series Models and Stochastic Processes

2.1 Time Series

All material presented in this section is classical and can be found in e.g. [3].

A time series $\{X_t\}$ is strictly stationary if $(X_1, \ldots, X_n)'$ and $(X_{1+h}, \ldots, X_{n+h})'$ have the same joint distribution function for all integers h and $n \geq 0$.

Definition 0.4. *Let $\{X_t\}$ be a time series, its **mean function** and **covariance function** are respectively given by*

$$\mathrm{E}(X_t) = \mu_t$$

and

$$\gamma_x(r, s) = \mathrm{Cov}(X_r, X_s)$$

*for all integers r and s. X_t is **weakly stationary** if*

1. *μ_t is independent of t,*

2. *$\gamma_x(t + h, t)$ is independent of t for each h.*

Definition 0.5. *Let $\{X_t\}$ be a weakly stationary time series with finite variance,* $\mathrm{Var}(X_t) = \sigma_x^2$. *The **autocovariance function** (ACVF) and **autocorrelation function** (ACF) at lag h are given respectively by*

$$\gamma_X(h) = \gamma_x(t, t + h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_0, X_t)$$

and

$$\rho_x(h) = \rho_x(t, t+h) = \frac{\gamma_x(t, t+h)}{\gamma_x(t, t)} = \text{Corr}(X_t, X_{t+h}) = \text{Corr}(X_0, X_h).$$

Let us note a couple of facts:

1. If $\{X_t\}$ is a strictly stationary, then all random variables are identically distributed.
2. $(X_t, X_{t+h})' = (X_1, X_{1+h})'$ in distribution $\forall h$ and t.
3. If $\{X_t\}$ is strictly stationary and $E(X_t^2) < \infty$ then it is also weakly stationary.
4. Weak stationarity does not imply strict stationarity.
5. An iid sequence is strictly stationary.

2.1.1 Linear Models

A time series $\{X_t\}$ is a linear process if it has a representation

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j Z_{t-j}, \qquad \forall t = 1, 2..., \tag{2.1}$$

where Z_t are iid random variables. In order to guarantee existence and strict stationarity one assumes either:

- the innovations Z_t are iid random variables with mean zero and variance σ_z^2, and $\{\psi_j\}$ is a sequence of constants such that $\sum_{j=-\infty}^{+\infty} |\psi_j|^2 < \infty$,
- or the innovations Z_t are iid and regularly varying with index $-\alpha$ and $\sum_{j=-\infty}^{\infty} |\psi_j|^\delta < \infty$, for $\delta \leq \alpha$. See [4, Eq. (2.6), Eq. (2.7)].

This linear process can be represented as

$$X_t = \psi(B) Z_t$$

where

$$\psi(B) = \sum_{j=-\infty}^{+\infty} \psi_j B^j,$$

and B^j is the backshift operator: $B^j Z_t = Z_{t-j}$.

A linear process is causal if $\psi_j = 0 \ \forall j < 0$, i.e, if

$$X_t = \sum_{j=0}^{+\infty} \psi_j Z_{t-j}.$$

Sometimes a linear process is called moving average of order ∞, denoted by $MA(\infty)$.

2.1.2 Stochastic Volatility (SV)

Let $\{\varepsilon_t\}$ be an iid sequence with mean zero and variance $\sigma_\varepsilon^2 < \infty$. Let $\{Y_t\}$ be a strictly stationary sequence independent of $\{\varepsilon_t\}$. Assume that $\sigma(\cdot)$ is a nonnegative function. A sequence given by

$$X_t = \sigma(Y_t)\varepsilon_t = \sigma_t \varepsilon_t \tag{2.2}$$

is called a stochastic volatility process.
The random variables $\{X_t\}$ are uncorrelated, but correlated in squares. Indeed:

$$\begin{aligned}
\mathrm{Cov}(X_t, X_{t+h}) &= \mathrm{E}[X_t X_{t+h}] \\
&= \mathrm{E}[\sigma_t \varepsilon_t \sigma_{t+h} \varepsilon_{t+h}] \\
&= \begin{cases} \mathrm{E}[\varepsilon_t]\mathrm{E}[\varepsilon_{t+h}]\mathrm{E}[\sigma_t \sigma_{t+h}] = 0 & \text{for} \quad h \neq 0 \\ \\ \sigma_\varepsilon^2 \mathrm{E}[\sigma_t^2] = \sigma_\varepsilon^2 \mathrm{E}[\sigma_0^2] & \text{for} \quad h = 0, \end{cases}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(X_t^2, X_{t+h}^2) &= \mathrm{E}[X_t^2 X_{t+h}^2] - \mathrm{E}[X_t^2] E[X_{t+h}^2] \\
&= \mathrm{E}[\sigma_t^2 \varepsilon_t^2 \sigma_{t+h}^2 \varepsilon_{t+h}^2] - \mathrm{E}[\sigma_t^2 \varepsilon_t^2]\mathrm{E}[\sigma_{t+h}^2 \varepsilon_{t+h}^2] \\
&= \mathrm{E}[\varepsilon_t^2 \varepsilon_{t+h}^2]\mathrm{E}[\sigma_{t+h}^2 \sigma_t^2] - \mathrm{E}[\varepsilon_t^2]\mathrm{E}[\varepsilon_{t+h}^2]\mathrm{E}[\sigma_t^2]\mathrm{E}[\sigma_{t+h}^2] \\
&= \begin{cases} \sigma_\varepsilon^4 \mathrm{Cov}(\sigma_t^2, \sigma_{t+h}^2) & \text{for} \quad h \neq 0 \\ \\ \sigma_\varepsilon^4 \mathrm{E}[\sigma_t^4] - (\mathrm{E}[\varepsilon_t^2])^2 (\mathrm{E}[\sigma_t^2])^2 = \sigma_\varepsilon^4 \left(\mathrm{E}[\sigma_0^4] - (\mathrm{E}[\sigma_0^2])^2\right) & \text{for} \quad h = 0, \end{cases}
\end{aligned}$$

given that $\mathrm{E}[\sigma_0^4] < \infty$ and $\mathrm{E}[\varepsilon_0^4] < \infty$.

2.1.3 Regression models and ARCH

Let $(Y_{1,t}, \ldots, Y_{p,t})'$, $t \geq 1$, be a sequence of random vectors in $\mathbf{R}^p$, and consider a regression model given by

$$X_t = f(Y_{1,t}, \ldots, Y_{p,t}) + \varepsilon_t \ ,$$

where $\{\varepsilon_t\}$, $t \geq 1$, are iid random variables, with $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{Var}[\varepsilon_t] = \sigma^2_\varepsilon$, such that for a fixed t, ε_t is independent of $(Y_{1,t}, \ldots, Y_{p,t})$, and $f(\cdot)$ is a linear function of $Y_{1,t}, \ldots, Y_{p,t}$,

$$f(Y_{1,t} \ldots, Y_{p,t}) = \beta_0 + \beta_1 Y_{1,t} + \ldots + \beta_p Y_{p,t}.$$

Hence,

$$\begin{aligned} \mathrm{E}(X_t|Y_{1,t} \ldots, Y_{p,t}) &= \mathrm{E}[f(Y_{1,t}, \ldots, Y_{p,t}) + \varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &= f(Y_{1,t}, \ldots, Y_{p,t}) + \mathrm{E}[\varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &= f(Y_{1,t}, \ldots, Y_{p,t}) + \mathrm{E}[\varepsilon_t] = f(Y_{1,t}, \ldots, Y_{p,t}), \end{aligned}$$

since ε_t is independent of $Y_{1,t}, \ldots, Y_{p,t}$ and $\mathrm{E}[\varepsilon_t] = 0$.
Furthermore

$$\begin{aligned} \mathrm{Var}(X_t|Y_{1,t} \ldots, Y_{p,t}) &= \mathrm{Var}[f(Y_{1,t}, \ldots, Y_{p,t}) + \varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &= \mathrm{Var}[f(Y_{1,t}, \ldots, Y_{p,t})|Y_{1,t} \ldots, Y_{p,t}] + \mathrm{Var}[\varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &\quad +2\mathrm{E}[f(Y_{1,t}, \ldots, Y_{p,t})\varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &= \mathrm{Var}[\varepsilon_t] + 2f(Y_{1,t}, \ldots, Y_{p,t})\mathrm{E}[\varepsilon_t|Y_{1,t} \ldots, Y_{p,t}] \\ &= \mathrm{Var}[\varepsilon_t] + 2f(Y_{1,t}, \ldots, Y_{p,t})E[\varepsilon_t] = \mathrm{Var}[\varepsilon_t], \end{aligned}$$

since conditionally on $Y_{1,t} \ldots, Y_{p,t}$, the variance of $f(Y_{1,t} \ldots, Y_{p,t})$ is zero. Therefore

$$\mathrm{Var}(X_t|Y_{1,t} \ldots, Y_{p,t}) = \sigma^2_\varepsilon.$$

One can compute further to get the unconditional expectation and unconditional variance:

$$\begin{aligned} \mathrm{E}(X_t) &= \mathrm{E}[f(Y_{1,t} \ldots, Y_{p,t})] \\ \mathrm{Var}(X_t) &= \sigma^2_\varepsilon + \mathrm{Var}[f(Y_{1,t} \ldots, Y_{p,t})]. \end{aligned}$$

The idea of Autoregressive Conditional Heteroscedasticity (ARCH) model is to introduce a time-dependent conditional variance. Define

$$X_t = f(Y_{1,t}, \dots, Y_{p,t}) + \sigma(Y_{1,t} \dots, Y_{p,t})\varepsilon_t,$$

where σ is a nonnegative function:

$$\sigma : \mathbb{R}^p \to \mathbb{R}_+.$$

If we take $f(Y_{1,t}, \dots, Y_{p,t}) = 0$ and denote $\sigma(Y_{1,t} \dots, Y_{p,t})$ by σ_t instead, we have

$$X_t = \sigma_t \varepsilon_t.$$

In particular, σ_t can be defined by

$$\sigma_t^2 = \alpha_0 + \Sigma_{i=1}^p \alpha_i X_{t-i}^2,$$

with $\alpha_0 > 0$ and $\alpha_j > 0$, $j = 1, \dots, p$. Consider for simplicity $p = 1$. By taking the squares of X_t we obtain $X_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2}\varepsilon_t$, thus

$$\begin{aligned} X_t^2 &= \alpha_0 \varepsilon_t^2 + \alpha_1 X_{t-1}^2 \varepsilon_t^2 \\ &\vdots \\ &= \alpha_0 \Sigma_{j=0}^{\infty} \alpha_1^j \varepsilon_t^2 \varepsilon_{t-1}^2 \cdots \varepsilon_{t-j}^2 + \alpha_1^{n+1} X_{t-n-1}^2 \varepsilon_t^2 \varepsilon_{t-1}^2 \cdots \varepsilon_{t-n}^2. \end{aligned}$$

Under the condition $|\alpha_1| < 1$, the last term converges to zero almost surely as $n \to \infty$. Hence if $\mathrm{E}[\varepsilon_t^2] = 1$, then

$$X_t^2 = \alpha_0 \Sigma_{j=0}^{\infty} \alpha_1^j \varepsilon_t^2 \varepsilon_{t-1}^2 \cdots \varepsilon_{t-j}^2$$

and

$$\mathrm{E}(X_t^2) = \frac{\alpha_0}{1 - \alpha_1}. \tag{2.3}$$

Hence, if $|\alpha_1| < 1$, then there exists a strictly stationary solution with finite variance. Below we will discuss the case of infinite variance as well.

2.1.4 GARCH

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is given by

$$X_t = \sigma_t \varepsilon_t, \tag{2.4}$$

where σ_t is defined by

$$\sigma_t^2 = \alpha_0 + \Sigma_{i=1}^p \alpha_i X_{t-i}^2 + \Sigma_{i=1}^q \beta_i \varepsilon_{t-i}^2. \tag{2.5}$$

2.2 Weak Dependence Properties of Time Series

In order to measure weak dependence in time series, we recall the α-mixing condition (see [5]).

Definition 0.6. *Let $\{X_t, t \in \boldsymbol{Z}\}$ be a strictly stationary time series and for $-\infty \leq a \leq b \leq \infty$, let $\mathcal{F}_a^b$ be the sigma-field generated by $X_a, \ldots, X_b$. Define*

$$\alpha_n = \sup_{C \in \mathcal{F}_{-\infty}^0, D \in \mathcal{F}_n^\infty} |\mathbb{P}(C \cap D) - \mathbb{P}(C)\mathbb{P}(D)|.$$

The sequence $\{X_t, t \in \boldsymbol{Z}\}$ is called strongly mixing if $\alpha_n \to 0$ as $n \to \infty$. We say that the sequence has geometric rates if $\alpha_n = \rho^n$, $\rho \in (0,1)$.

Example 0.4. *Let $X_t = X$ for all $t \in \mathbb{Z}$. Let $C = \{X_{-7} > 1, X_{-1} > 2\}$ and $D = \{X_n > 5\}$. Then $C \cap D = \{X_{-7} > 1, X_{-1} > 2, X_n > 5\}$, $C = \{X > 1, X > 2\} = \{X > 2\}$ and $D = \{X > 5\}$, so $C \cap D = \{X > 5\}$. Therefore the sequence is not α-mixing.*

The α-mixing condition implies the following bound:

$$\mathrm{Cov}(f(X_t), g(X_s)) \leq \alpha_{t-s}^{1/q_1} \|f(X_t)\|_{q_2} \|g(X_s)\|_{q_3} \ ,$$

where $\|U\|_q = \mathrm{E}^{1/q}[|U|^q]$ and $1/q_1 + 1/q_2 + 1/q_3 = 1$. In particular,

$$\mathrm{Cov}(\mathbf{1}_{x_1 < X_0 \leq x_2}, \mathbf{1}_{x_1 < X_t \leq x_2}) \leq \mathbb{P}^{1/q_2 + 1/q_3}(x_1 < X_0 \leq x_2) \alpha_t^{1/q_1} \ .$$

Hence, if $\{X_t\}$ is strictly stationary with geometric rates, then

$$\sum_{t=1}^{\infty} \mathrm{Cov}(\mathbf{1}_{x_1 < X_0 \leq x_2}, \mathbf{1}_{x_1 < X_t \leq x_2}) \leq \mathbb{P}^{1/q_2 + 1/q_3}(x_1 < X_0 \leq x_2) \underbrace{\sum_{t=1}^{\infty} \alpha_t^{1/q_1}}_{<\infty} \ . \tag{2.6}$$

2.3 Regular variation and mixing properties of time series models

In what follows, we will quote results on

1. Tail behaviour;

2. Multivariate regular variation;

3. Mixing properties.

2.3.1 Linear processes

Recall (2.1). Assume that the process is causal, that is

$$X_t = \sum_{j=0}^{+\infty} \psi_j Z_{t-j}.$$

1. If the innovations Z_t are regularly varying with index $-\alpha$ and $\sum_{j=0}^{\infty} |\psi_j|^\delta < \infty$, for $\delta \leq \alpha$, $\delta \leq 1$, then X_t is regularly varying. See [4, Eq. (2.6), Eq. (2.7)].

2. For each $t \geq 1$ the vector $(X_1, \ldots, X_t)$ is regularly varying, however, the exponent measure is given in a complicated form.

3. Assume that $\mathrm{E}[|Z_t|^\delta] < \infty$ and if $\delta \geq 1$, $\mathrm{E}[Z_t] = 0$. Assume that $\sum_{j=0}^{\infty} |\psi_j|^\delta < \infty$. Assume also that the innovations have a density f_Z such that

   $$\int_{-\infty}^{\infty} |f_Z(u+x) - f_Z(u)| du \leq c|x| \; .$$

   Then the process is α-mixing. In particular, if the coefficients ψ_j decay exponentially fast, then the process is α-mixing with geometric rates. See [5, Section 2.3.1].

2.3.2 Stochastic volatility models

Recall (2.2).

1. If the innovations ε_t are regularly varying with index $-\alpha$ and $\mathrm{E}[\sigma^{\alpha+\delta}(Y_t)] < \infty$, for $\delta > 0$, then X_t is regularly varying. This is the Breiman Lemma.

2. If the process $\{Y_t\}$ is regularly varying then the process $\{X_t\}$ is also regularly varying.

3. If the process $\{Y_t\}$ is α-mixing with geometric rates, then the process $\{X_t\}$ is also α-mixing with geometric rates.

2.3.3 GARCH

Recall (2.4) and (2.5). We consider for simplicity the case $p = q = 1$ only. Setting $A_t = \alpha_1 \varepsilon_{t-1}^2 + \beta_1$, $B_t = \alpha_0$, we have

$$\sigma_t^2 = A_t \sigma_{t-1}^2 + B_t.$$

Hence, σ_t^2 fulfills the Stochastic Recurrence Equation. It is clear that $\{\sigma_t^2\}$ and hence $\{X_t^2\}$ is a Markov chain.

1. Application of Kesten's Theorem (see [1]) yields that under appropriate conditions, σ_t^2 are regularly varying with some index $-\alpha$. In particular, this holds when the innovations ε_t are normal. Consequently, X_t^2 are regularly varying with index $-\alpha$.

2. For each $t \geq 1$ the vector $(X_1, \ldots, X_t)$ is regularly varying, however, the exponent measure is given in a complicated form.

3. Since $\{\sigma_t^2\}$ is a Markov chain, then under appropriate regularity conditions (like existence of a density of the innovations ε_t) it is α-mixing with geometric rates. See [1].

Since the exponent measure is usually very complicated, in most examples there is no explicit form of the limiting conditional probabilities. One of the very few exceptions is the AR(1) model, $X_t = \rho X_{t-1} + \varepsilon_t$, where the innovations ε_t are regularly varying with index $-\alpha$. Then, the limiting conditional probability has the form $\rho^{h\alpha}$.

2.4 CLT via Lindeberg's conditions

Theorem 1 ([6]). *Let $\{X_{n,t}, t \geq 1, n \geq 1\}$ be an array such that for each n, $X_{n,t}$, $t \geq 1$ is a sequence of independent random variables. Assume that*

- $\sum_{t=1}^{n} \mathrm{E}[X_{n,t}^2] \to \sigma^2 > 0$ *as* $n \to \infty$*;*
- *For each* $\epsilon > 0$*,* $\sum_{t=1}^{n} \mathrm{E}[X_{n,t}^2 \mathbf{1}_{\{|X_{n,t}|>\epsilon\}}] \to 0$ *as* $n \to \infty$*.*

Then

$$\sum_{t=1}^{n} X_{n,t} \to \mathcal{N}(0, \sigma^2).$$

2.5 Stochastic Processes

2.5.1 Brownian motion

Definition 1.1. *A process $B(t)$ is called a Brownian motion with starting point $x \in \boldsymbol{R}$ if the following holds:*

- $B(0) = x$,
- *the process has independent increments, i.e. for all times* $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ *the increments* $B(t_n) - B(t_{n-1})$, $B(t_{n-1}) - B(t_{n-2})$,..., $B(t_2) - B(t_1)$ *are independent random variables,*
- *for all* $t \geq 0$ *and* $h > 0$, *the increments* $B(t+h) - B(t)$ *are normally distributed with expectation zero and variance* h,
- *almost surely, the function* $t \mapsto B(t)$ *is continuous.*

We say that $\{B(t) : t \geq 0\}$ *is a* ***standard Brownian motion*** *if* $x = 0$.

2.5.2 Brownian Bridge

Definition 1.2. *A Brownian Bridge is defined as a stochastic process* $\{B_0(t), t \in [0,1]\}$ *with the following properties:*

- $B_0(0) = B_0(1) = 0$,
- *for* $t_1 < t_2 < \cdots < t_k$, *a random vector* $(B_0(t_1), B_0(t_2), \ldots, B_0(t_k))$ *has a mean zero multivariate normal distribution,*
- $\mathrm{Cov}(B_0(t), B_0(h)) = \min\{t, h\} - th$,
- *Almost surely, the function* $t \to B_0(t)$ *is continuous.*

2.5.3 Tightness

We start with two examples of stochastic processes considered in this thesis.

Example 1.1. Assume that $X_1, \ldots, X_n$ are observations from a stationary sequence $\{X_t\}$ with a marginal distribution F. We define the **empirical process** by

$$R_n(x) = \sqrt{n}\left(\widehat{F}_n(x) - F(x)\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbf{1}_{\{X_t \leq x\}} - \mathrm{E}[\mathbf{1}_{\{X_t \leq x\}}]). \tag{2.7}$$

Example 1.2. Assume that $\{X_t\}$ is a strictly stationary regularly varying time series such that its marginal distribution fulfills (1.2). Let u_n be a sequence such that $u_n \to \infty$ and $n\bar{F}(u_n) \to \infty$. We define the **Tail Empirical Process** by

$$\tau_n(x) = \frac{1}{\sqrt{n\overline{F}(u_n)}} \sum_{t=1}^{n} \left(\mathbf{1}_{\{X_t > u_n x\}} - \overline{F}(u_n x)\right). \tag{2.8}$$

We quote the following result (see [2]) that is valid for the *empirical process* and the *the tail empirical process.*

Lemma 1.1. *Let* $(Y_n(\cdot))$*,* $n \in \mathbb{N}$*, be a sequence of stochastic process with values in* $D(0,\infty)$*. We assume that either* $Y_n(\cdot)$ *is the empirical process or the tail empirical process. The sequence* $(Y_n(\cdot))$ *is tight if it satisfies the following condition:*

1. *There exist constants* $\gamma \geq 0$ *and* $\beta > 1$ *and a nondecreasing, continuous function* g *on* $(0,\infty)$ *such that*

$$P\left(|Y_n(x_2) - Y_n(x_2)| \geq \epsilon\right) \leq \left(\frac{1}{\epsilon}\right)^{\gamma} |g(x_2) - g(x_1)|^{\beta}$$

holds for all x_1*,* x_2*, all positive* λ *and all* $t \in \mathbb{N}$*.*

We note that the second condition is implied by

$$\mathrm{E}\left[|Y_n(x_2) - Y_n(x_1)|^{\gamma}\right] \leq C|g(x_2) - g(x_1)|^{\beta}.$$

Usually, it is hard to obtain $\beta > 1$ in the above criterion. Hence, [2] argues that tightness is implied by the bound

$$\mathrm{E}\left[|Y_n(x_2) - Y_n(x_1)|^4\right] \leq \frac{C}{n}(g_n(x_2) - g_n(x_1)) + (g_n(x_2) - g_n(x_1))^2 \,, \tag{2.9}$$

where g_n is a function that converges (as $n \to \infty$) to a continuous non-decreasing function (and hence converges uniformly).

Chapter 3

Some background on empirical processes

In this section we review some theory for standard empirical processes based on iid and on dependent random variables. The material below is based on [2].

3.1 Introduction

Assume that $X_1, \ldots, X_n$ are observations from a stationary sequence $\{X_t\}$ with a marginal distribution F. We define the **empirical distribution function** as

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{X_t \le x\}}, \tag{3.1}$$

with $x \in \mathbf{R}$. We have

$$\mathrm{E}[\widehat{F}_n(x)] = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}[\mathbf{1}_{\{X_t \le x\}}] = P(X \le x) = F(x),$$

regardless whether the observations are independent or dependent. We define the **empirical process** by

$$R_n(x) = \sqrt{n}\left(\widehat{F}_n(x) - F(x)\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbf{1}_{\{X_t \le x\}} - \mathrm{E}[\mathbf{1}_{\{X_t \le x\}}]). \tag{3.2}$$

We note that $\mathrm{E}[R_n(x)] = 0$.

3.2 Empirical process for iid random variables

Let now $X_1, \ldots, X_n$ be an iid sample from a distribution function $F(x)$. Then (3.1) is still valid and (3.2) defines the empirical process. Now, we compute the variance of the empirical distribution function. Because of independence, we have

$$\begin{aligned} \mathrm{Var}[\widehat{F}_n(x)] &= \frac{1}{n^2}\left\{\sum_{t=1}^{n}\mathrm{Var}[\mathbf{1}_{\{X_t\le x\}}] + 2\sum_{t=1}^{n-1}\mathrm{Cov}(\mathbf{1}_{\{X_1\le x\}}, \mathbf{1}_{\{X_{t+1}\le x\}})\right\} \\ &= \frac{1}{n}\mathrm{Var}[\mathbf{1}_{\{X_0\le x\}}] = \frac{1}{n}F(x)\overline{F}(x). \end{aligned}$$

Analogously, we can also write the formula for $\mathrm{Cov}(R_n(x), R_n(y))$:

$$\mathrm{Cov}(R_n(x), R_n(y)) = F(x\wedge y) - F(x)F(y), \tag{3.3}$$

where $x \wedge y = \min\{x, y\}$. Therefore by the central limit theorem $\forall x \in \mathbf{R}$,

$$R_n(x) \to \mathcal{N}(0, F(x)\overline{F}(x)) \tag{3.4}$$

in distribution as $n \to \infty$. However, we know an even more detailed result, so called *functional convergence of the empirical process*.

Theorem 2. *Assume that $\{X_t\}$ is an iid sequence from the distribution F. Then,*

$$R_n(x) \Rightarrow B_0(F(x)), \tag{3.5}$$

where $B_0(y)$ is a Brownian bridge on $[0,1]$ and $\Rightarrow$ denotes weak convergence in $D(-\infty,\infty)$.

Note that $B_0(F(x))$ has covariance $F(x\wedge y) - F(x)F(y)$ which is the same as the covariance given in (3.3).

We provide a proof of this theorem, since a similar approach will be used for the tail empirical process and for dependent random variables. In order to prove the theorem one needs to prove finite dimensional convergence (note that the one-dimensional convergence has been already stated in (3.4)) and *tightness*.

In what follows, we will verify that $R_n(x)$ is tight. Let $x_1 < x_2$. Define

$$U_t = \mathbf{1}_{\{x_1 < X_t \le x_2\}} - \mathrm{E}[\mathbf{1}_{\{x_1 < X_t \le x_2\}}]\ .$$

Then $R_n(x_2) - R_n(x_1) = (1/\sqrt{n})\sum_{t=1}^n U_t$. We compute

$$\begin{aligned} \mathrm{E}|R_n(x_2) - R_n(x_1)|^4 &= \frac{1}{n^2}\sum_{t=1}^{n}\mathrm{E}\left[|U_t|^4\right] + \frac{1}{n^2}\sum_{1\le t\ne s\le n}\mathrm{E}[U_t^2U_s^2] \\ &\le \frac{C}{n}P(x_1\le X_t\le x_2) + (P(x_1\le X_t\le x_2)P(x_1\le X_s\le x_2))^2 \\ &= \frac{C}{n}(F(x_2)-F(x_1)) + (F(x_2)-F(x_1))^2\ . \end{aligned}$$

Hence, (2.9) is fulfilled and the process is tight.

3.3 Empirical process for dependent random variables

3.3.1 Partial sums of strictly stationary sequences

Before we talk about the empirical process for dependent random variables, we discuss briefly sample means based on dependent observations. Let $\{X_t\}$ be a strictly stationary process with mean μ and finite variance. Then

$$\mathrm{E}[\overline{X}_n] = \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}[X_t] = \mu$$

and

$$\begin{aligned}
\mathrm{Var}[\overline{X}_n] &= \mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{n} X_t\right] \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\mathrm{Var}[X_t] + \frac{2}{n^2}\sum_{t=1}^{n-1}\sum_{k=t+1}^{n}\mathrm{Cov}[X_t, X_k] \\
&= \frac{1}{n^2}\sum_{t=1}^{n}\mathrm{Var}[X_t] + \frac{2}{n^2}\sum_{t=1}^{n-1}\sum_{k=t+1}^{n}\mathrm{Cov}[X_0, X_{k-t}].
\end{aligned}$$

Hence

$$\mathrm{Var}[\overline{X}_n] = \frac{1}{n}\gamma(0) + 2\frac{1}{n}\sum_{k=1}^{n-1}(1 - k/n)\gamma(k),$$

where $\gamma(h) = \mathrm{Cov}(X_0, X_h)$. Consequently,

$$\mathrm{Var}[\sqrt{n}\,\overline{X}_n] \sim \gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k), \qquad n \to \infty$$

given that (in particular) $\sum_{k=1}^{\infty} |\gamma(k)| < \infty$. Hence, the dependence structure influences the variance of the sum. The same will happen in case of empirical processes.

3.3.2 Empirical distribution for strictly stationary processes

Let $\{X_t\}$ be a strictly stationary process with a marginal distribution function $F(x)$. We recall the empirical distribution function,

$$\widehat{F}_n(x) \quad = \quad \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{X_t \leq x\}}.$$

As before, $\mathrm{E}[\widehat{F}_n(x)] = F(x)$ and we can compute the variance of $R_n(x)$ as

$$\begin{aligned} \mathrm{Var}[R_n(x)] &\sim \mathrm{Var}(1_{\{X_0<x\}}) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(\mathbf{1}_{\{X_0\leq x\}}, \mathbf{1}_{\{X_k \leq x\}}) \\ &= F(x)(1-F(x)) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(\mathbf{1}_{\{X_0\leq x\}}, \mathbf{1}_{\{X_k \leq x\}}). \end{aligned}$$

Of course, this computation requires that the infinite sum is well-defined. According to (2.6), this is the case when $\{X_t\}$ is α-mixing with geometric rates. We can also compute the covariance function:

$$\mathrm{Cov}[R_n(x), R_n(y)] = \{F(x \wedge y) - F(x)F(y)\} + 2\sum_{k=1}^{\infty} \mathrm{Cov}(\mathbf{1}_{\{X_0\leq x\}}, \mathbf{1}_{\{X_{t-l} \leq y\}}). \tag{3.6}$$

Because of the different variance and covariance structure, the limiting process changes as compared to the iid case. We have the following result.

Theorem 3. *Assume that $\{X_t\}$ is a strictly stationary α-mixing sequence with geometric rates. Then,*

$$R_n(x) \Rightarrow \tilde{B}_0(x),$$

where $\tilde{B}_0(x)$ is a Gaussian process with covariance given by (3.6).

Chapter 4

Tail empirical process

4.1 Introduction

Assume that $\{X_t\}$ is a strictly stationary regularly varying time series such that its marginal distribution fulfills (1.2). Let u_n be a sequence such that $u_n \to \infty$ and $n\bar{F}(u_n) \to \infty$. We define the **Tail Empirical Distribution** by

$$\widehat{T}_n(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{X_t > u_n x\}} \ .$$

Furthermore, define

$$\widehat{G}_n(x) = \frac{1}{\overline{F}(u_n)} \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}_{\{X_t > u_n x\}} = \frac{\widehat{T}_n(x)}{\overline{F}(u_n)} \ .$$

First, we compute the expectation of $\widehat{G}_n(x)$. We have

$$\mathrm{E}[\widehat{T}_n(x)] = \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}[\mathbf{1}_{\{X_t > u_n x\}}] = P(X_0 > u_n x) = \overline{F}(u_n x) \ ,$$

hence (cf. (1.2)),

$$\lim_{n\to\infty} \mathrm{E}[\widehat{G}_n(x)] = \lim_{n\to\infty} \frac{1}{\overline{F}(u_n)} \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}[\mathbf{1}_{\{X_t > u_n x\}}] = \lim_{n\to\infty} \frac{\overline{F}(u_n x)}{\overline{F}(u_n)} = \psi(x). \tag{4.1}$$

Next, we define the **Tail Empirical Process**:

$$\tau_n(x) = \frac{1}{\sqrt{n\overline{F}(u_n)}} \sum_{t=1}^{n} \left(\mathbf{1}_{\{X_t > u_n x\}} - \overline{F}(u_n x) \right) . \tag{4.2}$$

Note that $\mathrm{E}[\tau_n(x)] = 0$.

4.2 Tail Empirical Process for iid random variables

First, we evaluate the limiting variance for the tail empirical process based on iid random variables. We have

$$\begin{aligned}\mathrm{Var}[\sqrt{n}\widehat{T}_n(x)] &= \mathrm{Var}\left[\frac{\sqrt{n}}{n}\sum_{t=1}^{n}\mathbf{1}_{\{X_t>u_nx\}}\right] = \frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}[\mathbf{1}_{\{X_t>u_nx\}}] \\ &= \mathrm{Var}[\mathbf{1}_{\{X_0>u_nx\}}] = \overline{F}(u_nx)(1-\overline{F}(u_nx))\ .\end{aligned}$$

Therefore

$$\mathrm{Var}\left[\sqrt{n\overline{F}(u_n)}\widehat{G}_n(x)\right] = \frac{1}{\overline{F}(u_n)}\overline{F}(u_nx)(1-\overline{F}(u_nx))\ .$$

Hence, by (1.2),

$$\lim_{n\to\infty}\mathrm{Var}\left[\sqrt{n\overline{F}(u_n)}\widehat{G}_n(x)\right] = \psi(x)\ . \tag{4.3}$$

Expression (4.3) indicates that $\lim_{n\to\infty}\mathrm{Var}(\tau_n(x)) = \psi(x)$. This *suggests* the following result.

.

Lemma 3.1. *Assume that $\{X_t\}$ is an iid regularly varying sequence such that (1.2) holds. Then for each fixed x we have $\tau_n(x)\to\mathcal{N}(0,\psi(x))$ in distribution as $n\to\infty$.*

Proof. We check the central limit theorem for $\tau_n(x)$ by using the Lindeberg conditions, see Section 2.4. Let $Z_{n,t}(x)$ be the random process defined by

$$Z_{n,t}(x) = \frac{1}{\sqrt{n\overline{F}(u_n)}}(\mathbf{1}_{\{X_t>u_nx\}} - \overline{F}(u_nx)), \tag{4.4}$$

so that

$$\tau_n(x) = \sum_{t=1}^{n} Z_{n,t}(x). \tag{4.5}$$

We have already calculated the limiting variance for $\tau_n(x)$. It remains to evaluate the second Lindeberg condition. We have for $1/p+1/q=1$,

$$\begin{aligned}\sum_{t=1}^{n}\mathrm{E}\left[Z_{n,t}^2(x)\mathbf{1}_{\{Z_{n,t}>\epsilon\}}\right] &\le \sum_{t=1}^{n}\left(\mathrm{E}[Z_{n,t}^{2p}(x)]\right)^{\frac{1}{p}}\left(\mathrm{E}[\mathbf{1}_{\{Z_{n,t}>\epsilon\}}]\right)^{\frac{1}{q}} \\ &\le \sum_{t=1}^{n}\left(\mathrm{E}[Z_{n,t}^{2p}(x)]\right)^{\frac{1}{p}}\left(P[Z_{n,t}>\epsilon]\right)^{\frac{1}{q}}\ .\end{aligned}$$

Take $p=2$ and $q=2$. By Hölder inequality,

$$\begin{aligned}\left(\mathrm{E}[Z_{n,t}^4(x)]\right)^{\frac{1}{2}} &= \left(\frac{1}{\left(\sqrt{n\overline{F}(u_n)}\right)^4}\mathrm{E}[(\mathbf{1}_{\{Z_{n,t}>u_nx\}}-\overline{F}(u_nx))^4]\right)^{\frac{1}{2}} \\ &= \frac{1}{n\overline{F}(u_n)}\left(\mathrm{E}[(\mathbf{1}_{\{Z_{n,t}>u_nx\}}-\overline{F}(u_nx))^4]\right)^{\frac{1}{2}} \le \frac{C}{n\overline{F}(u_n)}\left(\mathrm{E}[(\mathbf{1}_{\{Z_{n,t}>u_nx\}})^4]\right)^{\frac{1}{2}} \\ &\le \frac{C}{n\overline{F}(u_n)}\left(\mathrm{E}[\mathbf{1}_{\{Z_{n,t}>u_nx\}}]\right)^{\frac{1}{2}} \le \frac{C}{n\overline{F}(u_n)}\left(\overline{F}(u_nx)\right)^{\frac{1}{2}} .\end{aligned}$$

[1] Therefore

$$\begin{aligned}\sum_{t=1}^n \mathrm{E}\left[Z_{n,t}^2(x)\mathbf{1}_{\{Z_{n,t}>\epsilon\}}\right] &\le \sum_{t=1}^n \frac{C}{n\overline{F}(u_n)}(\overline{F}(u_nx))^{\frac{1}{2}}\left(P(Z_{n,t}(x)>\epsilon)\right)^{\frac{1}{2}} \\ &\le C\frac{\overline{F}^{\frac{1}{2}}(u_nx)}{\overline{F}^{\frac{1}{2}}(u_n)}\frac{\left(\frac{\mathrm{E}[Z_{n,t}^2(x)]}{\epsilon^2}\right)^{\frac{1}{2}}}{\overline{F}^{\frac{1}{2}}(u_n)} \\ &\le C\frac{\overline{F}^{\frac{1}{2}}(u_nx)}{\overline{F}^{\frac{1}{2}}(u_n)}\frac{\left(\frac{\overline{F}(u_nx)}{n\overline{F}(u_n)}(1-\overline{F}(u_nx))\right)^{\frac{1}{2}}}{\epsilon\overline{F}^{\frac{1}{2}}(u_n)} .\end{aligned}$$

Hence

$$\lim_{n\to\infty}\sum_{t=1}^n \mathrm{E}\left[Z_{n,t}^2(x)\mathbf{1}_{\{Z_{n,t}>\epsilon\}}\right] \le C\lim_{n\to\infty}\psi^{\frac{1}{2}}(x)\frac{1}{\epsilon}\frac{\psi^{\frac{1}{2}}(x)}{\sqrt{n\overline{F}(u_n)}} \le \frac{1}{\epsilon}\lim_{n\to\infty}\frac{\psi(x)}{\sqrt{n\overline{F}(u_n)}} = 0.$$

This finishes the proof. □

We extend the above result to functional convergence.

Theorem 4. *Assume that $\{X_t\}$ is an iid regularly varying sequence such that (1.2) holds. Then,*

$$\tau_n(x) \Rightarrow B(\psi(x)),$$

where $B(x)$ is a Brownian motion on $\boldsymbol{R}_+$ and $\Rightarrow$ denotes weak convergence in $D(-\infty,\infty)$.

Proof. We will prove that $\tau_n(x)$ defined in (4.5) is tight. Let $x_1<x_2$. Define

$$U_t = \frac{1}{\sqrt{n\bar{F}(u_n)}}\left\{\mathbf{1}_{\{u_nx_1<X_t\le u_nx_2\}} - \mathrm{E}[\mathbf{1}_{\{u_nx_1<X_0\le u_nx_2\}}]\right\} = \frac{1}{\sqrt{n\bar{F}(u_n)}}V_t .$$

[1]for any random variable x with mean μ, $E[(X-\mu)^k]\le C_kEX^k$, where C_k is constant.

Then $\tau_n(x_2) - \tau_n(x_1) = \sum_{t=1}^n U_t$. We compute

$$\begin{aligned}
\mathrm{E}|\tau_n(x_2) - \tau_n(x_1)|^4 &= \frac{1}{(n\bar{F}(u_n))^2} \sum_{t=1}^n \mathrm{E}\left[|V_t|^4\right] + \frac{1}{(n\bar{F}(u_n))^2} \sum_{1\le t\neq s\le n} \mathrm{E}[V_t^2 V_s^2] \\
&\le \frac{C}{n} \frac{P(u_n x_1 \le X_0 \le u_n x_2)}{\bar{F}(u_n)} + \left(\frac{P(u_n x_1 \le X_0 \le u_n x_2)}{\bar{F}(u_n)}\right)^2 \\
&= \frac{C}{n}(g_n(x_2) - g_n(x_1)) + (g_n(x_2) - g_n(x_1))^2 ,
\end{aligned}$$

where

$$g_n(s) = \frac{\mathbb{P}(X_0 > u_n s)}{\mathbb{P}(X_0 > u_n)} .$$

The sequence g_n converges to ψ. Hence, we verified that (2.9) holds. The process is tight. □

4.3 Tail Empirical distribution with strictly stationary process

Assume that $\{X_t\}$ is a strictly stationary regularly varying sequence such that (1.2) holds. Furthermore, we assume that it is α-mixing with geometric rates. We recall the Tail Empirical Distribution:

$$\widehat{T}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{X_t > u_n x\}},$$

$$\widehat{G}_n(x) = \frac{\widehat{T}_n(x)}{\overline{F}(u_n)},$$

and

$$\mathrm{E}[\widehat{T}_n(x)] = \overline{F}(u_n x).$$

Therefore (cf (4.1), (4.4))

$$\mathrm{Var}\left[\sqrt{n\overline{F}(u_n)}\widehat{G}_n(x)\right] \sim \psi(x) + 2\sum_{t=2}^n \mathrm{Cov}(Z_{n,1}(x), Z_{n,t}(x)) ,$$

given that the latter series is summable, As in the situation of standard empirical processes, this is the case when the sequence $\{X_t\}$ is α-mixing, with geometric rates.

Set

$$k = k_n = n\bar{F}(u_n) . \tag{4.6}$$

Theorem 5. *Assume that $\{X_t\}$ is a strictly stationary sequence. Suppose that there exist sequences $r = r_n$, u_n, $m = m_n$, $q = q_n$ such that $r + m + q = o(n)$ and*

- $u_n \to \infty$, $\frac{n}{k}\mathbb{P}(X_0 > u_n x) \to \psi(x)$;
- $r = o(m)$;
- $\{X_t\}$ *is α-mixing such that* $\frac{qr}{k}\sum_{i=m}^{\infty}\alpha_i \to 0$;
- *There exists a function $\gamma(x)$ such that* $n\sum_{t=2}^{m}\operatorname{Cov}(Z_{n,1}(x), Z_{n,t}(x)) \to \gamma(x)$ *as* $n \to \infty$;
- $q\alpha_r \to 0$.

Then

$$\tau_n(x) = \sum_{t=1}^{n} Z_{n,t}(x) \to \mathcal{N}(0, \psi(x) + 2\gamma(x)).$$

We note that m and r will play a role of the size of large and small blocks, respectively, whereas q will be the number of small blocks.

4.3.1 Sketch of a proof of the Theorem 5

In order to prove the theorem, we will use a blocking technique. Let $X_1, \ldots, X_n$ be a sample from a strictly stationary sequence of random variables. We split it in large blocks of size m each and small blocks of size r each, such that there are q large blocks and q small blocks, $n = q(r+m)$. Large blocks will be denoted by I_j, $j = 1, \ldots, q$, small blocks will be denoted by $\widetilde{I}_j$, $j = 1, \ldots, q$. Recall that

$$Z_{n,t}(x) = \frac{1}{\sqrt{k}}\{\mathbf{1}_{\{X_t > u_n x\}} - \overline{F}(u_n x)\}.$$

Furthermore, define

$$\tau_n(x; B) = \sum_{t \in B} Z_{n,t}(x),$$

with $B \subseteq \{1, \ldots, n\}$.

Example 5.1. $\tau_n(x) = \tau_n(x; \{1, \ldots, n\}) = \sum_{t=1}^{n} Z_{n,t}(x)$.

Step 1: Show that small blocks can be removed. Let $\widetilde{I}_j \subseteq \{1, \ldots, n\}$ be the jth small block of size r. Then

$$\mathrm{Var}\left[\sum_{j=1}^{q} \tau_n(x, \widetilde{I}_j)\right] = q\mathrm{Var}\left[\tau_n(x, \widetilde{I}_1)\right] + 2q \sum_{j=2}^{q} \mathrm{Cov}\left[\tau_n(x, \widetilde{I}_1), \tau_n(x, \widetilde{I}_j)\right].$$

We want to show that $\mathrm{Var}\left[\sum_{j=1}^{q} \tau_n(x, \widetilde{I}_j)\right] \to 0$ as $n \to \infty$. This means that small blocks can be removed. We look at the first term on the right hand side:

$$q\mathrm{Var}\left[\tau_n(x, \widetilde{I}_1)\right] = q\left(r\mathrm{Var}\left[Z_{0,n}(x)\right] + 2r \sum_{j=2}^{r} \mathrm{Cov}\left[Z_{n,1}(x), Z_{n,t}(x)\right]\right). \qquad (4.7)$$

Now

$$\begin{aligned}\mathrm{Var}[Z_{0,n}(x)] &= \frac{1}{k}\mathrm{Var}[\mathbf{1}_{\{X_0 > u_n x\}}] \\ &\sim \frac{1}{k}\mathbb{P}(X_0 > u_n x) = \frac{n}{k}\mathbb{P}(X_0 > u_n x)\frac{1}{n} \sim \psi(x)\frac{1}{n} \qquad \text{as } n \to \infty.\end{aligned}$$

Hence, if $qr/n \to 0$ then the first term in (4.7) is negligible.

Next we show that $q \sum_{j=2}^{q} \mathrm{Cov}\left[\tau_n(x; \widetilde{I}_1), \tau_n(x; \widetilde{I}_j)\right] \to 0$ as $n \to \infty$. Since the mixing sequence α_j can be chosen to be nonincreasing (without loss of generality), we have:

$$\begin{aligned} q\sum_{j=2}^{q} \mathrm{Cov}\left[\tau_n(x; \widetilde{I}_1), \tau_n(x; \widetilde{I}_j)\right] &\leq q \sum_{j=2}^{q} \mathrm{E}\left[\mid \tau_n(x; \widetilde{I}_1) \times \tau_n(x; \widetilde{I}_j) \mid\right] \\ &= q \sum_{l=2}^{q} \mathrm{E}\left[\mid \sum_{s \in \widetilde{I}_1} Z_{n,s}(x) \times \sum_{t \in \widetilde{I}_l} Z_{n,t}(x) \mid\right] \\ &\leq q \sum_{l=2}^{q} \sum_{s \in \widetilde{I}_1} \sum_{t \in \widetilde{I}_l} \mathrm{E}\left[\mid Z_{n,s}(x) \times Z_{n,t}(x) \mid\right] \\ &= \frac{q}{k} \sum_{l=2}^{q} \sum_{i \in \widetilde{I}_1} \sum_{j \in \widetilde{I}_l} \alpha_{j-i} \leq \frac{q}{k} \sum_{l=2}^{q} \sum_{i \in \widetilde{I}_1} \sum_{j \in \widetilde{I}_l} \alpha_j \\ &= \frac{qr}{k} \sum_{l=2}^{q} \sum_{j \in \widetilde{I}_l} \alpha_j = \frac{qr}{k} \sum_{i=m}^{\infty} \alpha_i \to 0\,. \end{aligned}$$

A similar argument yields negligibility of the second term of (4.7). This means that small blocks can be removed. Hence, the limit of $\tau_n(x) = \tau_n(x, \{1, \ldots, n\})$, is the same as of $\sum_{j=1}^{q} \tau_n(x; I_j)$.

Step 2: The result of the first step is that the limiting distribution of $\tau_n(x) = \sum_{t=1}^n Z_{n,t}(x)$ is the same as to $\sum_{j=1}^q \tau_n(x, I_j)$.

Let $\tau_n^*(x, I_j)$, $j = 1 \ldots, q$, be random variables with the same distribution as $\tau_n(x, I_j)$, but $\tau_m^*(x, I_1), \tau_m^*(x, I_2), \ldots, \tau_m^*(x, I_n)$ are independent. Then

$$|\mathrm{E}\left[\prod_{j=1}^q e^{it\tau_n(x,I_j)}\right] - \mathrm{E}\left[\prod_{j=1}^q e^{it\tau_n^*(x,I_j)}\right]| \le |\mathrm{E}\left[\prod_{j=1}^q e^{it\tau_n(x,I_j)}\right] - \prod_{j=1}^q \mathrm{E}[e^{it\tau_n^*(x,I_j)}]| \le 4q\alpha_r ,$$

see [5, p. 31]. Hence, the limiting distribution of $\sum_{j=1}^q \tau_n(x, I_j)$ is the same as the limiting distribution of $\sum_{j=1}^q \tau_n^*(x, I_j)$.

Step 3: Hence, we look at $\sum_{j=1}^q \tau_n^*(x, I_j)$, the sum of independent random variables with mean 0 and the same distribution. We have to check what is the variance of $\tau_n^*(x, I_1)$:

$$\mathrm{Var}\left[\tau_n^*(x, I_1)\right] = m\mathrm{Var}(Z_{0,n}(x)) + 2m \sum_{j=2}^m \mathrm{Cov}(Z_{n,1}(x), Z_{n,t}(x)).$$

Therefore, as $n \to \infty$,

$$\begin{aligned}
\mathrm{Var}\left[\sum_{j=1}^q \tau_n^*(x, I_j)\right] &= \sum_{j=1}^q \mathrm{Var}\left[\tau_n^*(x, I_j)\right] \\
&= qm\mathrm{Var}\left[Z_{0,n}(x)\right] + 2qm \sum_{j=2}^m \mathrm{Cov}\left[Z_{n,1}(x), Z_{n,t}(x)\right] \\
&= \frac{qm}{k}\mathbb{P}(X_0 > u_n x) + 2qm \sum_{j=2}^m \mathrm{Cov}\left[Z_{n,1}(x), Z_{n,t}(x)\right] \\
&\sim \frac{n}{k}\mathbb{P}(X_0 > u_n x) + 2qm \sum_{j=2}^m \mathrm{Cov}\left[Z_{n,1}(x), Z_{n,t}(x)\right] \\
&\sim \psi(x) + 2\gamma(x).
\end{aligned}$$

Hence,

$$\mathrm{Var}\left[\sum_{j=1}^q \tau_n^*(x, I_j)\right] \sim \psi(x) + 2\gamma(x).$$

Step 4: Verification of the second Lindeberg condition is more technical and involves computation of the fourth moments. We skip it, we refer to [8].

Chapter 5

Estimation of Limiting Conditional Probabilities

Let $\{X_t\}$ be a stationary sequence. A goal of this section is to estimate

$$\mathbb{P}(X_h > y \mid X_0 > y) \tag{5.1}$$

when y is large. We will consider $h = 1$ only. Then, the conditional probability is written as

$$\frac{\mathbb{P}(X_0 > y, X_1 > y)}{\mathbb{P}(X_0 > y)}.$$

If we have observations $X_1, \ldots, X_{n+1}$, an empirical estimator of this probability is

$$\frac{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i > y, X_{i+1} > y\}}}{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i > y\}}}.$$

However, we are interested in large values of y. In order to do this, recall that in (4.6) we set $k = n\bar{F}(u_n)$. Then

$$\begin{aligned} \mathrm{E}\left[\frac{1}{k}\sum_{i=1}^n \mathbf{1}_{\{X_i > u_n, X_{i+1} > u_n\}}\right] &= \frac{1}{k} n \mathbb{P}(X_0 > u_n, X_1 > u_n) = \frac{\mathbb{P}(X_0 > u_n, X_1 > u_n)}{k/n} \\ &\sim \frac{\mathbb{P}(X_0 > u_n, X_1 > u_n)}{\mathbb{P}(X_0 > u_n)} \sim \mathbb{P}(X_1 > u_n | X_0 > u_n). \end{aligned}$$

Hence, this suggests the following estimator of the limiting conditional probability in (5.1)

$$\frac{1}{k}\sum_{i=1}^n \mathbf{1}_{\{X_i > u_n, X_{i+1} > u_n\}}. \tag{5.2}$$

In the next chapter we will explain the role of k and u_n. We note that k and u_n are linked by the unknown function F, so the estimator is not practical at the moment. Define

$$\widetilde{Z}_{n,t}(x) = \frac{1}{\sqrt{k}} \left(\mathbf{1}_{\{X_t > u_n x, X_{t+1} > u_n x\}} - \mathbb{P}(X_0 > u_n x, X_1 > u_n x) \right),$$

and $\widetilde{\tau}_n(x; B) = \sum_{t \in B} \widetilde{Z}_{t,n}(x)$. Note that $\mathrm{E}[\widetilde{Z}_{t,n}(x)] = 0$ for all $x \in \mathbf{R}$.

Theorem 6. *Assume that $\{X_t\}$ is a strictly stationary sequence. Suppose that there exist sequences $r = r_n$, u_n, $m = m_n$, $q = q_n$ such that $r + m + q = o(n)$ and*

- $u_n \to \infty$, $\frac{n}{k} \mathbb{P}(X_0 > u_n x, X_1 > u_n x) \sim \widetilde{\psi}(x)$;
- $r = o(m)$ *(Hence* $\frac{rq}{n} \to 0$*)*;
- $\{X_t\}$ *is* α*-mixing such that* $\frac{qr}{k} \sum_{i=m}^{\infty} \alpha_i \to 0$;
- *There exists a function* $\widetilde{\gamma}(x)$ *such that* $n \sum_{t=2}^{m} \mathrm{Cov}(\widetilde{Z}_{n,1}(x), \widetilde{Z}_{n,t}(x)) \to \widetilde{\gamma}(x)$ *as* $n \to \infty$;
- $q\alpha_r \to 0$.

Then

$$\widetilde{\tau}_n(x) = \sum_{t=1}^{n} \widetilde{Z}_{t,n}(x) \to \mathcal{N}(0, \widetilde{\psi}(x) + 2\widetilde{\gamma}(x)) .$$

Proof. The proof mimics the proof of Theorem 5.
Step 1: Small blocks are negligible.

$$\mathrm{Var}\left[\sum_{j=1}^{q} \widetilde{\tau}_n(x; \widetilde{I}_j) \right] = q\mathrm{Var}\left[\widetilde{\tau}_n(x, \widetilde{I}_1) \right] + 2q \sum_{j=2}^{q} \mathrm{Cov}\left[\widetilde{\tau}_n(x, \widetilde{I}_1), \widetilde{\tau}_n(x, \widetilde{I}_j) \right].$$

We compute as $n \to \infty$,

$$\mathrm{Var}\left[\widetilde{Z}_{n,0}(x) \right] = \frac{1}{k} \mathbb{P}(X_0 > u_0 x, X_1 > u_n x) = \frac{n}{k} \mathbb{P}(X_0 > u_n x, X_1 > u_n x) \frac{1}{n} \sim \frac{1}{n} \widetilde{\psi}(x) .$$

If $x = 1$, then $\widetilde{\psi}(x) = \lim_{n \to \infty} \frac{n}{k} \mathbb{P}(X_0 > u_n, X_1 > u_n) = \lim_{n \to \infty} \mathbb{P}(X_1 > u_n | X_0 > u_n)$.

Moreover

$$
\begin{aligned}
q\sum_{j=2}^{q}\operatorname{Cov}\left[\tau_n(x;\widetilde{I}_1),\tau_n(x;\widetilde{I}_j)\right] &\le q\sum_{j=2}^{q}\operatorname{E}\left[\mid \tau_n(x;\widetilde{I}_1)\times\tau_n(x;\widetilde{I}_j)\mid\right]\\
&= q\sum_{l=2}^{q}\operatorname{E}\left[\mid \sum_{s\in\widetilde{I}_1}\tilde{Z}_{n,s}(x)\times\sum_{t\in\widetilde{I}_l}\tilde{Z}_{n,t}(x)\mid\right]\\
&\le q\sum_{l=2}^{q}\sum_{s\in\widetilde{I}_1}\sum_{t\in\widetilde{I}_l}\operatorname{E}\left[\mid \tilde{Z}_{n,s}(x)\times\tilde{Z}_{n,t}(x)\mid\right]\\
&= \frac{q}{k}\sum_{l=2}^{q}\sum_{i\in\widetilde{I}_1}\sum_{j\in\widetilde{I}_l}\alpha_{j-i-i}\le\frac{q}{k}\sum_{l=2}^{q}\sum_{i\in\widetilde{I}_1}\sum_{j\in\widetilde{I}_l}\alpha_j\\
&= \frac{qr}{k}\sum_{l=2}^{q}\sum_{j\in\widetilde{I}_l}\alpha_j=\frac{qr}{k}\sum_{i=m}^{\infty}\alpha_i\to 0.
\end{aligned}
$$

□

Step 2: We consider $\widetilde{\tau}_n(x)=\sum_{t=1}^{n}\widetilde{Z}_{t,n}(x)$ and $\sum_{j=1}^{q}\widetilde{\tau}_n(x,\widetilde{I}_j)$.

Let $\widetilde{\tau}_n^*(x,\widetilde{I}_j)$, $j=1,\ldots,q$ be independent random variables, with the same distribution as $\widetilde{\tau}_n(x,\widetilde{I}_j)$. From Step 2 in the proof of Theorem 5, we conclude that $\sum_{j=1}^{q}\widetilde{\tau}_n(x,\widetilde{I}_j)$ has the same the limiting distribution as $\sum_{j=1}^{q}\widetilde{\tau}_n^*(x,\widetilde{I}_j)$.

Step 3: We compute now the variance of independent random variables with the same distribution (as in Step 3 in the proof of Theorem 5):

$$
\operatorname{Var}\left[\sum_{j=1}^{q}\widetilde{\tau}_n^*(x,\widetilde{I}_j)\right]\sim\widetilde{\psi}(x)+2\widetilde{\gamma}(x).
$$

Conclusion:

$$
\widetilde{\tau}_n(x)=\sum_{t=1}^{n}\widetilde{Z}_{n,t}(x)\to\mathcal{N}(0,\widetilde{\psi}(x)+2\widetilde{\gamma}(x))\ .
$$

□

Chapter 6

Simulation Studies

6.1 Estimation of the tail index

Our basic assumption is that the marginal distribution of the time series is heavy tailed, that is $P(|X_0| > x) = x^{-\alpha}L(x)$, where L is slowly varying and $\alpha > 0$ is the tail index. First, we will look at estimation of the tail index using the Hill estimator $\widehat{\alpha}_{Hill}$.

Let $X_1, \ldots, X_n$ be iid Pareto random variables, with the density $f(x) = C\alpha x^{-\alpha-1}$ for $x > b$ and $C > 0$. We evaluate the constant C as

$$\int_b^{+\infty} f(x) \;=\; \int_b^{+\infty} C\alpha x^{-\alpha-1} = \left[C\alpha\frac{1}{-\alpha}x^{-\alpha}\right]_b^{+\infty} = Cb^{-\alpha} = 1 \Rightarrow C = b^{\alpha}.$$

Based on observations $x_1, \ldots, x_n$, we compute the likelihood and the log-likelihood as follows:

$$\begin{aligned}
L(\alpha) &= \alpha^n b^{n\alpha} \prod_{i=1}^{n} x_i^{-\alpha-1}, \\
\log(L(\alpha)) &= n\log\alpha + n\alpha\log b - (\alpha+1)\sum_{t=1}^{n}\log x_t, \\
\frac{\partial \log(L(\alpha))}{\partial\alpha} &= \frac{n}{\alpha} + n\log b - \sum_{t=1}^{n}\log x_t, \\
\frac{n}{\alpha} &= \sum_{t=1}^{n}(\log x_t - \log b), \\
\widehat{\alpha} &= \frac{n}{\sum_{t=1}^{n}\log(\frac{x_t}{b})}, \\
\frac{1}{\widehat{\alpha}} &= \frac{\sum_{t=1}^{n}\log(\frac{x_t}{b})}{n}.
\end{aligned}$$

In general α is a characteristic of the tail, and so we want now to use k-order statistics. We order data as $X_{1,n} \leq \ldots \leq X_{n,n}$. This leads to the following Hill estimator:

$$\frac{1}{\widehat{\alpha}_{Hill}} = \frac{\sum_{t=1}^{k} \log(\frac{X_{n,n-t+1}}{X_{n,n-k}})}{k} \,. \tag{6.1}$$

In practice, one uses the Hill plot, that is one plots $\widehat{\alpha}_{Hill}$ (that depends on k) against different values of k and one looks for a *stability region*. Below, we estimate the tail index for AR(1) models with Pareto innovations and different values of α. We note that it is usually very hard to find a good estimate of α.

Figure 6.1: AR(1) model with $\rho = 0.5$ and $\alpha = 1$. There is a small stability region when one uses about $k = 100$ order statistics, yielding $\widehat{\alpha}_{Hill} \approx 1$.

Figure 6.2: AR(1) model with $\rho = 0.7$ and $\alpha = 2$. There is a small stability region when one uses about $k = 100$ order statistics, yielding $\widehat{\alpha}_{Hill} \approx 2$.

Figure 6.3: AR(1) model with $\rho = 0.9$ and $\alpha = 4$. There is a small stability region when one uses about $k = 100$ order statistics, yielding $\widehat{\alpha}_{Hill} \approx 3$.

6.2 Estimation of the limiting conditional probabilities

We use (5.2), that is

$$\hat{p}_k := \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\{X_i > u_n, X_{i+1} > u_n\}} \ ,$$

to estimate limiting conditional probabilities. Since the level $u_n \to \infty$ has to be determined, as in case of the Hill estimator we replace u_n with $X_{n,n-k}$. Then, as in case of the Hill estimator, we produce an extremogram plot, that is we plot $\hat{p}_k$ against different values of k and we try to identify a stability region.

6.2.1 AR(1) model with $\rho = 0.9$ and $\alpha = 4$

We consider for AR(1) model with different values of ρ and α. We note that for $\rho > 0$,

$$\lim_{x \to \infty} P(X_1 > x \mid X_0 > x) = \rho^{\alpha} \ . \tag{6.2}$$

We perform some numerical studies for AR(1) model $X_t = \rho X_{t-1} + Z_t$ with $\rho = 0.9$ and Pareto innovations with the tail index $\alpha = 4$.

- On Figure 6.4 we show one realization of the estimator $\hat{p}_k$ plotted for different values of k. The values of the estimator are displayed against the values of the order statistics (the largest statistics displayed at the right part of the picture). Hence, the graph is constructed in a different way as compared to the Hill estimator. In the case of the Hill estimator, on the horizontal line we display number k of extremes used in the computation of the Hill estimator (from $k = 1$ - only one largest extremes, to $k = n$), whereas here we plot against the actual value of order statistics, from smallest to largest.

- Large variability of the estimator is displayed on Figure 6.5, which is based on 10 repetitions.

- On Figure 6.6 we show boxplots based on 1000 Monte Carlo repetitions, using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations. There is large variability when k is small; on the other hand there is small variability but large bias when k is large.

Figure 6.4: Extremogram for AR(1) model with $\rho = 0.9$ and $\alpha = 4$. The dotted horizontal line denotes the true value ρ^{α}.

Figure 6.5: 10 Extremogram repetitions AR(1) model with $\rho = 0.9$ and $\alpha = 4$. The dotted horizontal line denotes the true value ρ^{α}.

Figure 6.6: Monte Carlo estimation of the limiting conditional probabilities using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations, $\rho = 0.9$ and $\alpha = 4$.

6.2.2 AR(1) model with $\rho = 0.7$ and $\alpha = 2$

We perform some numerical studies for AR(1) model $X_t = \rho X_{t-1} + Z_t$ with $\rho = 0.7$ and Pareto innovations with the tail index $\alpha = 2$.

- On Figure 6.7 we show one realization of the estimator $\hat{p}_k$ plotted for different values of k. The values of the estimator are displayed against the value of order statistics (the largest order statistics displayed at the right part of the picture). Hence, the graph is constructed in a different way as compared to the Hill estimator.

- Large variability of the estimator is displayed on Figure 6.8, which is based on 10 repetitions.

- On Figure 6.9 we show boxplots based on 1000 Monte Carlo repetitions, using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations. There is large variability when k is small; on the other hand there is small variability but large bias when k is large.

Figure 6.7: Extremogram for AR(1) model with $\rho = 0.7$ and $\alpha = 2$. The dotted horizontal line denotes the true value ρ^α.

Figure 6.8: 10 Extremogram repetitions AR(1) model with $\rho = 0.7$ and $\alpha = 2$. The dotted horizontal line denotes the true value ρ^α.

6.2.3 AR(1) model with $\rho = 0.5$ and $\alpha = 1$

We perform some numerical studies for AR(1) model $X_t = \rho X_{t-1} + Z_t$ with $\rho = 0.5$ and Pareto innovations with the tail index $\alpha = 1$.

- On Figure 6.10 we show one realization of the estimator $\hat{p}_k$ plotted for different

Figure 6.9: Monte Carlo estimation of the limiting conditional probabilities using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations, $\rho = 0.7$ and $\alpha = 2$.

values of k. The values of the estimator are displayed against the value of order statistics (the largest order statistics displayed at the right part of the picture). Hence, the graph is constructed in a different way as compared to the Hill estimator.

- Large variability of the estimator is displayed on Figure 6.11, which is based on 10 repetitions.
- On Figure 6.12 we show boxplots based on 1000 Monte Carlo repetitions, using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations. There is large variability when k is small; on the other hand there is small variability but large bias when k is large.

Figure 6.10: Extremogram for AR(1) model with $\rho = 0.5$ and $\alpha = 1$. The dotted horizontal line denotes the true value ρ^α.

Figure 6.11: 10 Extremogram repetitions AR(1) model with $\rho = 0.5$ and $\alpha = 1$. The dotted horizontal line denotes the true value ρ^α.

6.2.4 GARCH(1,1)

We perform some numerical studies for GARCH(1,1) model $X_t = \sigma_t \varepsilon_t$ with $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1} + \beta_1 \varepsilon_{t-1}$ with normal innovations ε_t. Note that X_t are heavy tailed, however, the tail index cannot be written explicitly. Furthermore, the limiting conditional probability cannot be written explicitly.

Figure 6.12: Monte Carlo estimation of the limiting conditional probabilities using $k = 5\%, 10\%, 15\%, 20\%$ and 25% of observations, $\rho = 0.5$ and $\alpha = 1$, there is large variability when k is small, and there is small variability but a large bias when k is large.

- On Figure 6.14 we show one realization of the estimator $\hat{p}_k$ plotted for different values of k. The values of the estimator are displayed against the value of order statistics (the largest statistics displayed at the right part of the picture). Hence, the graph is constructed in a different way as compared to the Hill estimator.
- On Figure 6.13 we show GARCH(1,1) model with $\alpha_0 = 0$, $\alpha_1 = 0.9$ and $\beta_1 = 0.8$. There is a small stability region when one uses about $k = 100$ order statistics, yielding $\widehat{\alpha}_{Hill} \approx 1.4$.

Figure 6.13: GARCH(1,1)

Figure 6.14: Extremogram for GARCH(1,1) model with $\alpha_0 = 0$, $\alpha_1 = 0.9$ and $\beta_1 = 0.8$.

Chapter 7

Conclusions

The goal of this thesis was to estimate the limiting conditional probabilities defined in (1.4). We stated and proved the central limit theorem for the corresponding empirical estimator (see Theorem 6). The estimator involves an unknown sequence u_n and in Section 6.2 we replaced u_n with an extreme order statistics $X_{n,n-k}$. To prove formally the central limit theorem for the estimator based on order statistics, one needs to establish first the functional convergence in Theorem 6. This is beyond the scope of this thesis.

Several natural questions arise:

- As evidenced by our simulation studies, the estimator is very unstable. Can we find a better (that is, more stable, or less variable) estimator?

- The choice of $k = k_n$, the number of order statistics used, is crucial. How to do this in practice? One approach is to look for a *stability region* on the graph like Figure 6.4. Can we do this in a more automatic way?

These questions should be the subject of a future research.

Chapter 8

R Codes

```
library(tseries)
```

8.1 Preliminary Plots

```
alpha=4;
n=500;
size=n;
phi=0.9;
SimTimeSeries=arima.sim(model=list(ar=c(phi)),n=size,rand.gen=rt,df=alpha);
TimeSeries=abs(SimTimeSeries);
par(mfrow=c(1,2))
plot.ts(TimeSeries,ylab="Time series");
HillPlot(TimeSeries);
print(quantile(TimeSeries,0.95));
q_{95}=quantile(TimeSeries,0.95);
print(quantile(TimeSeries,0.99));
q_{99}=quantile(TimeSeries,0.99);
print(phi^{alpha});
r=50;
alpha.estim=HillIndex(TimeSeries,r);
print(alpha.estim);
```

8.2 Extremogram plot for limiting conditional probabilities

```
par(mfrow=c(1,3);
par(mfrow=c(1,1));
mylag=1;
Extr<-Extremogram(TimeSeries,lag=mylag);
SortedTS<-Extr[,2];
Extremogram.Estimates<-Extr[,1];
range=max(Extremogram.Estimates);
plot(SortedTS,Extremogram.Estimates,type="l",col="blue",
    xlab="Order Statistics", ylab="Estimated Joint Probabilities",
    main="Extremogram",ylim=c(0,1));
lines(SortedTS,c(rep($phi^{mylag\times alpha}$,length(SortedTS))),
type="l",col="blue",lwd=0.3,lty=2)
```

Estimation of the limiting conditional probabilities for the simulated time series

```
alpha=4;
n=500;
size=n;
phi=0.9;
r=50;
mylag=1;
quant_{95}=NULL;
quant_{99}=NULL;
nofrep=10;

 SimTimeSeries=arima.sim(model=list(ar=c(phi)),n=size,rand.gen=rt,df=alpha);
TimeSeries=abs(SimTimeSeries);
q_{95}=quantile(TimeSeries,0.95);
q_{99}=quantile(TimeSeries,0.99);
quant_{95}=c(quant_{95},q_{95});
quant_{99}=c(quant_{99},q_{99});
```

```
alpha.estim=HillIndex(TimeSeries,r);
print(alpha.estim);
Extr<-Extremogram(TimeSeries,lag=mylag);
SortedTS<-Extr[,2];
Extremogram.Estimates<-Extr[,1];
 Extremogram.1=matrix(0,nofrep,length(Extremogram.Estimates));
 Extremogram.2=matrix(0,nofrep,length(Extremogram.Estimates));
Extremogram.1[1,]=SortedTS;
Extremogram.2[1,]=Extremogram.Estimates;

for(i in 2:nofrep){
 SimTimeSeries=arima.sim(model=list(ar=c(phi)),n=size,rand.gen=rt,df=alpha);
TimeSeries=abs(SimTimeSeries);
q_{95}=quantile(TimeSeries,0.95);
q_{99}=quantile(TimeSeries,0.99);
quant_95=c(quant_{95},q_{95});
quant_{99}=c(quant_{99},q_{99});
Extr<-Extremogram(TimeSeries,lag=mylag);
SortedTS<-Extr[,2];
Extremogram.Estimates<-Extr[,1];
Extremogram.1[i,]=SortedTS;
Extremogram.2[i,]=Extremogram.Estimates;
}
colors=c("red","blue","green","brown","yellow","pink",
    "blueviolet","burlywood","grey","darkgoldenrod3");
par(mfrow=c(1,1))
range.1=min(Extremogram.1);
range.2=max(Extremogram.1);
range.3=min(Extremogram.2);
range.4=max(Extremogram.2);
plot(Extremogram.1[1,],Extremogram.2[1,],type="l",col="black",
    xlab="Order Statistics", ylab="Estimated Joint Probabilities",
    main="Extremogram", ylim=c(0,1));
for(i in 2:nofrep)
{
 lines(Extremogram.1[i,],Extremogram.2[i,],type="l",col=colors[i%%10]);
```

```
}
lines(SortedTS,c(rep(phi^{mylag\times alpha},length(SortedTS))),
    type="l",col="black",lwd=0.3,lty=2);
q_{95}=mean(quant_{95});
q_{99}=mean(quant_{99});
abline(v = q_95,lty=3,lwd=0.01);
abline(v = {q_99},lty=3,lwd=0.01);
```

8.3 Monte Carlo estimation of the limiting conditional probabilities

```
alpha=4;
n=500;
size=n;
phi=0.9;
r=50;
SimTimeSeries=arima.sim(model=list(ar=c(phi)),n=size,rand.gen=rt,df=alpha);
TimeSeries=abs(SimTimeSeries);
r=50;
nofrep=100;
mylag=1;
Estimates.Extr.Lag1<-Extremogram.K(TimeSeries,lag=mylag)
 Extremogram.Lag1=matrix(0,nofrep,length(Estimates.Extr.Lag1));
for(i in 1:nofrep){
 SimTimeSeries=arima.sim(model=list(ar=c(phi)),n=size,rand.gen=rt,df=alpha);
TimeSeries=abs(SimTimeSeries);
r=50;
mylag=1;
Estimates.Extr.Lag1<-Extremogram.K(TimeSeries,lag=mylag);
Extremogram.Lag1[i,]=Estimates.Extr.Lag1;
}
par(mfrow=c(1,1));
mylag=1;
boxplot(list(Extremogram.Lag1[,1],Extremogram.Lag1[,2],
```

```
    Extremogram.Lag1[,3],Extremogram.Lag1[,4],Extremogram.Lag1[,5]),
    main="Extremogram Lag 1",ylim=c(0,1));
abline(h=phi^alpha,col="blue",lwd=0.3,lty=2);
```

Bibliography

[1] B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of garch processes. *Stochastic Processes and their Applications*, 99:95–115, 2002.

[2] Patrick Billingsley. *Convergence of probability measures.* New York, Wiley, 1968.

[3] P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting.* Springer, 2001.

[4] R. Davis and S. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability*, 13:179–195, 1985.

[5] Paul Doukhan. *Mixing. Theory and examples.* Springer, 1968.

[6] P. Hall and C. Heyde. *Martingale limit theory and its applications.* New York: Academic Press., 1980.

[7] R. Kulik and P. Soulier. *Heavy Tailed Time Series.* Lecture Notes, 2014.

[8] T. Mikosch R. Davis. The extramogram: A correlation for extreme. *Bernoulli*, 15(4):977–1009, 2009.

[9] S. Resnick. *Heavy Tail Phenomena.* Springer, 2007.