High quantile estimation for some Stochastic Volatility models

Ling Luo

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

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

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

Abstract

In this thesis we consider estimation of the tail index for heavy tailed stochastic volatility models with long memory. We prove a central limit theorem for a Hill estimator. In particular, it is shown that neither the rate of convergence nor the asymptotic variance is affected by long memory. The theoretical findings are verified by simulation studies.

Acknowledgements

I would like to thank my supervisors, Rafal Kulik and Mahmoud Zarepour, for their friendly guidance, inspiring discussions and unending patience throughout my research project. Specially, I would also like to thank my husband, Hao, for his understanding and support and my lovely son, Travis, for bringing happiness into my life.

Contents

List of Figures

List of Tables

Introduction

Motivation

Typical financial data have the following behavior:

- they are uncorrelated,
- they have strong dependence in squares,
- they are heavy-tailed.

Notice that the autocovariance function may not be defined if the variance is infinite.

To illustrate better, we consider the daily exchange rates between US Dollar and Swiss Franc for the period 2000-2010.

Figure 1 shows that the time series $\{P_i, i = 1, \dots, n\}$ seems to be nonstationary, it has a decreasing trend. Consider the log-differences, i.e. $S_i = \log(P_i/P_{i-1})$, $i = 2, \dots, n$.

The time series plot of S_i, $i = 2, \dots, n$, seems to be stationary. There is no correlation among different lags, hence the log-difference series $\{S_i, i = 2, \dots, n\}$ is **uncorrelated**. Figure 2, lower panel depicts that significant autocorrelations exist in S_i^2, $i = 2, \dots, n$. Therefore the log-difference series is **not independent**. The QQ-plot shows that S_i has heavier tails than the normal distribution.

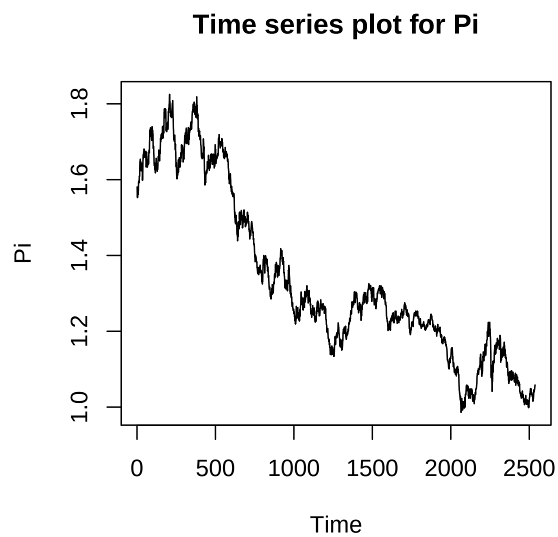

Figure 1: Exchange rates

The natural question is about modeling the time series so that it can capture the above features. Furthermore, we would like to estimate relevant parameters, in particular the so-called tail index in such models.

Structure of the thesis

In Chapter 1 we give a brief overview of the standard time series, long memory and stochastic volatility (SV) models. We will illustrate that only SV models can capture the above mentioned properties. In particular, a pure SV process and EGARCH (Exponential Generalized Autoregressive Conditionally Heteroscedastic) process can have both dependence and heavy-tailed features. These processes are defined in Section 1.5.1 and appropriate tail and dependence properties are discussed. In particular we introduce a notion of the tail index. Furthermore, in Section 1.5.2 we collect some limit theorems for long memory models. We use tools such as Hermite polynomials to establish our results.

In Chapter 2, we establish asymptotic theory for high quantile estimation of EGARCH and SV processes. The core problem resolved in this paper is to study the asymptotic behavior of a tail empirical process with deterministic levels and obtain the corresponding behavior of a tail empirical process with random levels. We conclude in Theorem 2.1.1 that the long memory influences the behaviour of the tail empirical process with deterministic levels, however, this is not the case for random levels; see Section 2.2. We use the latter result to obtain asymptotic normality of the Hill estimator, the most popular estimator of the tail index.

In Chapter 3, we study the numerical performance of three tail index estimators: Peak over Threshold (POT) estimator, Hill estimator and Pickands estimator. Note that the Hill estimator can be thought of as the POT estimator based on random levels. Therefore, the theoretical results in Chapter 2 suggest that the Hill estimator should be robust against long memory, contrary to the standard POT estimator, which is based on deterministic levels. We illustrate that under the SV model with different dependence and index, the dependence does not influence the performance of the Hill estimator.

Contribution

Convergence of the tail empirical processes was studied in Rootzen (2009) in a weakly dependent case and in Kulik and Soulier (2011) for pure long memory stochastic volatility models. Our main contribution is the extension of the latter result to so-called EGARCH models. This extension requires a careful application of the martingale central limit theorem; see Section 2.1. Furthermore, we conduct simulations which confirm our theoretical findings. Similar work has been done in McElroy and

Jach (2011) or Mikosch and Rezapour (2011); in both cases for a pure stochastic volatility model.

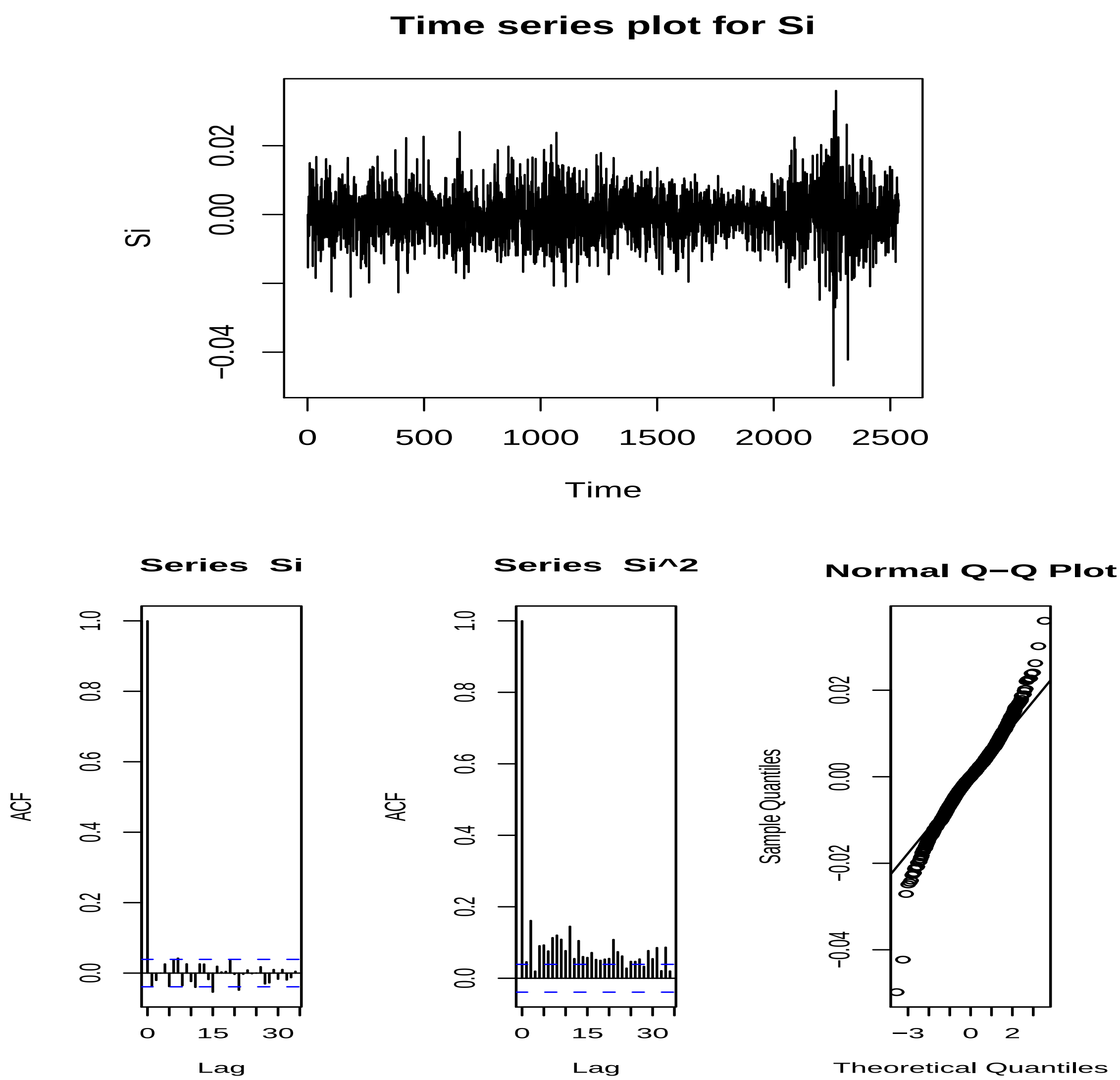

Figure 2: Graph of log-differences: time series, autocovariance function, autocovariance function of squares, QQ-plot

Chapter 1

Time Series Models

Let $\{X_i, -\infty < i < \infty\}$ be a stationary sequence with a finite variance. We will use the following notation:

$$\gamma_X(h) = \mathrm{Cov}(X_i, X_{i+h}) = \mathrm{Cov}(X_0, X_h) ,$$

$$\rho_X(h) = \mathrm{Cor}(X_0, X_h) = \frac{\gamma_X(h)}{\mathrm{Var}(X)} .$$

1.1 White noise

Let $\{Z_i, -\infty < i < \infty\}$ be a sequence of independent standard normal random variables. Of course, these random variables, as well as their squares are uncorrelated. To illustrate this, we simulate 1000 observations from a i.i.d. $N(0,1)$ process.

Figure 1.1 shows 1000 simulated values, the sequence seems to be stationary. Furthermore, the figure shows the corresponding autocorrelation function, the values $\rho_Z(h)$ are close to 0 for all $h > 0$. It indicates that the series $\{Z_i, -\infty < i < \infty\}$ is **uncorrelated**. Also, the squares are **uncorrelated**. We will write $\{Z_i, -\infty < i < \infty\} \sim IID(0,1)$.

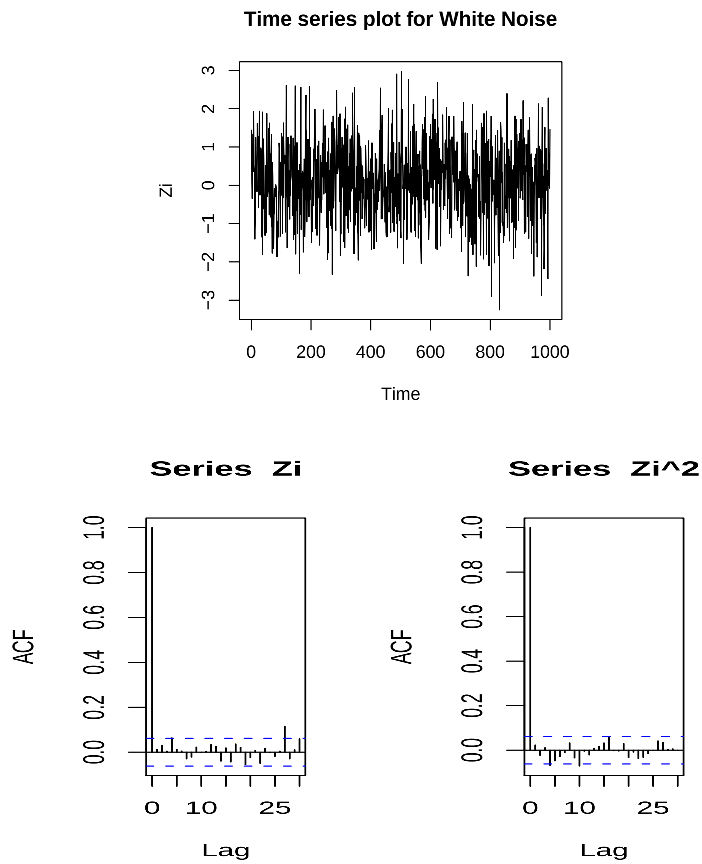

Figure 1.1: Simulated white noise and its ACF

1.2 ARMA models

Definition 1.2.1 *We say that* $\{X_i, -\infty < i < \infty\}$ *is an* ***ARMA(p,q)*** *process if* $\{X_i, -\infty < i < \infty\}$ *is stationary and if for every* i*,*

$$X_i - \phi_1 X_{i-1} - \cdots - \phi_p X_{i-p} = Z_i + \theta_1 Z_{i-1} + \cdots + \theta_q Z_{i-q}, \qquad (1.2.1)$$

where $\{Z_i, -\infty < i < \infty\} \sim \text{White Noise}(0, \sigma^2)$ *and the polynomials*

$$\phi_p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta_q(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

have no common factors.

We may write

$$\phi_p(B)X_i = \theta_q(B)Z_i, \tag{1.2.2}$$

where B *is the backward shift operator* ($B^j X_i = X_{i-j}, j = 0, \pm 1, \ldots$).

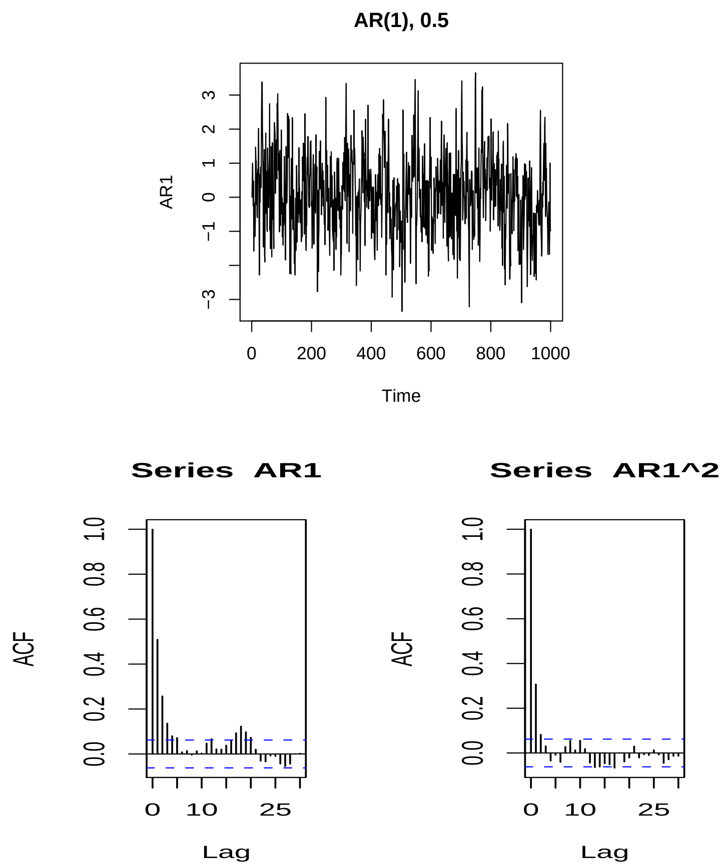

Figure 1.2: Simulated AR(1) and its ACF

We simulate 1000 observations from an AR(1)=ARMA(1,0) process. Figure 1.2 shows that the AR(1) series is stationary. The ACF decays exponentially to zero.

There is **correlation** in the process for different lags h.

Theoretically, we will show that the covariance function is not equal to 0.

Example 1.2.2 Consider AR(1) model: $X_i = \phi X_{i-1} + Z_i$, where $|\phi| < 1$, and Z_i, $i = 1, \ldots, n$, is a white noise (i.e, the sequence of uncorrelated random variables, with mean zero and variance σ_Z^2). Then

$$X_i = \phi X_{i-1} + Z_i = \sum_{j=0}^{\infty} \phi^j Z_{i-j}$$

which is a linear process with ACF

$$\gamma_X(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma_Z^2 = \sigma_Z^2 \phi^h \frac{1}{1-\phi^2} = C\phi^h, \qquad h > 0.$$

This means that the AR(1) process is **correlated**. □

We simulate 1000 observations from an MA(1)=ARMA(0,1) process. Figure 1.3 shows that the MA(1) series is stationary. The ACF is almost zero after lag 1 of the process and it shows that the series is **correlated**. Again, the squares of the time series are **correlated**.

Example 1.2.3 Consider MA(1) process

$$X_i = Z_i + \theta Z_{i-1},$$

where $\{Z_i, -\infty < i < \infty\} \sim IID(0, \sigma^2)$. We have

$$\begin{aligned} \gamma_X(h) &= E(X_i X_{i-h}) - 0 = E[(Z_i + \theta Z_{i-1})(Z_{i-h} + \theta Z_{i-h-1})] \\ &= E(Z_i Z_{i-h} + \theta Z_i Z_{i-h-1} + \theta Z_{i-1} Z_{i-h} + \theta^2 Z_{i-1} Z_{i-h-1}). \end{aligned}$$

When $h = 0$,

$$\begin{aligned} \gamma_X(0) &= E(Z_i^2 + \theta Z_i Z_{i-1} + \theta Z_i Z_{i-1} + \theta^2 Z_{i-1}^2) \\ &= \sigma^2 + \theta^2 \sigma^2 = \sigma^2(1 + \theta^2) \end{aligned}$$

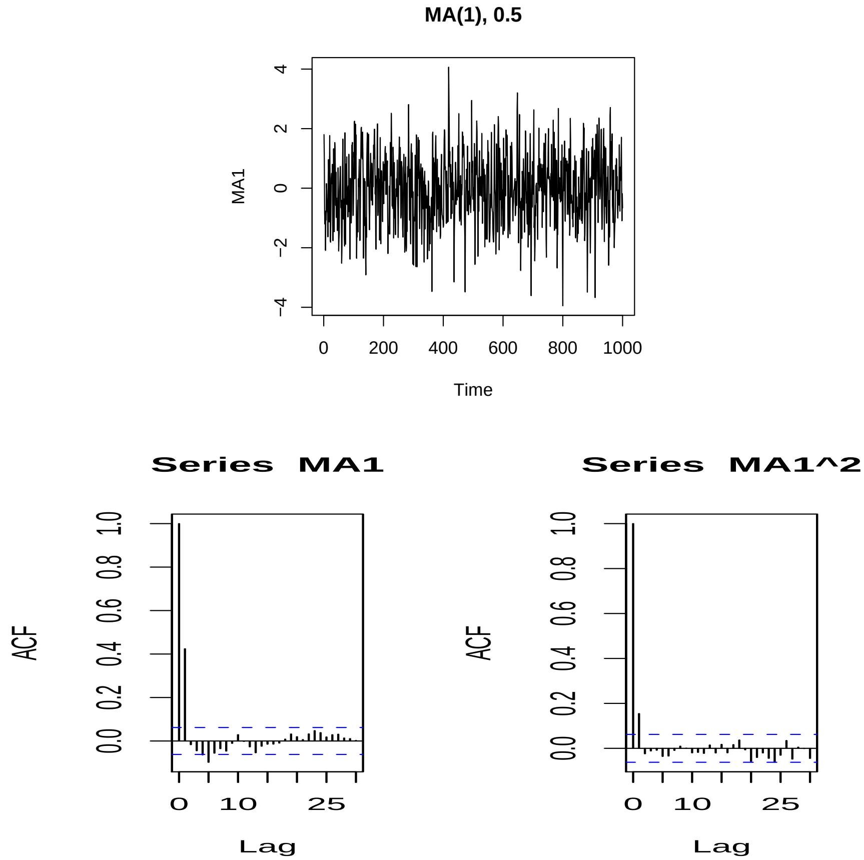

Figure 1.3: Simulated MA(1) and its ACF

When $h = 1$,

$$\begin{aligned}\gamma_X(1) &= E(Z_i Z_{i-1} + \theta Z_i Z_{i-2} + \theta Z_{i-1}^2 + \theta^2 Z_{i-1} Z_{i-2}) \\ &= \theta E(Z_{i-1}^2) = \theta\sigma^2\end{aligned}$$

When $h = -1$,

$$\begin{aligned}\gamma_X(-1) &= E(Z_i Z_{i+1} + \theta Z_i^2 + \theta Z_{i-1} Z_{i+1} + \theta^2 Z_{i-1} Z_i) \\ &= \theta E(Z_i^2) = \theta\sigma^2\end{aligned}$$

When $h = 2$,

$$\gamma_X(2) \quad = \quad E(Z_i Z_{i-2} + \theta Z_i Z_{t-3} + \theta Z_{i-1} Z_{i-2} + \theta^2 Z_{i-1} Z_{t-3}) = 0.$$

Similarly,

$$\gamma_X(h) = 0 \qquad \text{if } h > 2.$$

Thus we have,

$$\gamma_X(h) = \begin{cases} \sigma^2(1+\theta^2) & \text{if } h = 0; \\ \sigma^2\theta & \text{if } h = \pm 1; \\ 0 & \text{if } |h| > 1. \end{cases}$$

□

In summary, ARMA models can not capture the following properties of financial data: the original time series is uncorrelated, but their squares are correlated.

1.3 ARFIMA model

Consider a process $\{X_i, -\infty < i < \infty\}$ defined as

$$\phi_p(B)(1-B)^d X_i = \theta_q(B) Z_i, \qquad \{Z_i, -\infty < i < \infty\} \sim IID(0, \sigma_Z^2), \tag{1.3.1}$$

where d is a rational number and B is the backward shift operator;

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

are respectively the autoregressive and the moving average characteristic polynomials, satisfying

$$\phi(z) \neq 0 \text{ and } \theta(z) \neq 0 \text{ for all z such that } |z| \leq 1.$$

For simplicity, to find the range of d for the process (1.3.1), we consider the case, where $\phi_p(B) = 1$ and $\theta_q(B) = 1$,

$$(1-B)^d X_i = Z_i \tag{1.3.2}$$

If the process (1.3.2) is stationary, then we may write it as (see Brockwell and Davis 2002)

$$X_i = (1-B)^{-d} Z_i. \tag{1.3.3}$$

The operator $(1-B)^{-d}$ in equation (1.3.3) is defined by the binomial expansion

$$\begin{aligned} (1-B)^{-d} &= \sum_{j=0}^{\infty} \binom{-d}{j} (-B)^j = \sum_{j=0}^{\infty} \binom{-d}{j} (-1)^j B^j \\ &= \sum_{j=0}^{\infty} \psi_j B^j \ , \end{aligned}$$

where $\psi_j = \binom{-d}{j}(-1)^j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$. Therefore X_i in equation (1.3.3) is

$$X_i = (1-B)^{-d} Z_i = \sum_{j=0}^{\infty} \psi_j B^j Z_i = \sum_{j=0}^{\infty} \psi_j Z_{i-j}. \tag{1.3.4}$$

Lemma 1.3.1 *We have*

$$\psi_j \approx \frac{1}{\Gamma(d) j^{1-d}}.$$

Proof: Using Stirling's formula

$$\Gamma(x) \approx \sqrt{2\pi} e^{-x} x^{x+1/2} \text{ as } x \to \infty,$$

we have

$$\begin{aligned} \psi_j &= \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \approx \frac{\sqrt{2\pi} e^{-(j+d)} (j+d)^{j+d-1/2}}{\sqrt{2\pi} e^{-(j+1)} (j+1)^{j+1-1/2} \Gamma(d)} \\ &= \frac{1}{\Gamma(d)} \frac{e^{-d+1} (j+d)^{j+d-\frac{1}{2}}}{(j+1)^{j+\frac{1}{2}}}. \end{aligned}$$

Now, we prove that

$$\frac{1}{\Gamma(d)} \frac{e^{-d+1} (j+d)^{j+d-\frac{1}{2}}}{(j+1)^{j+\frac{1}{2}}} \approx \frac{1}{\Gamma(d) j^{1-d}},$$

equivalently,

$$\lim_{j\to\infty} j^{1-d} \frac{e^{-d+1} (j+d)^{j+d-\frac{1}{2}}}{(j+1)^{j+\frac{1}{2}}} = 1.$$

$$
\begin{aligned}
LHS &= e^{-d+1}\lim_{j\to\infty}\frac{j^{1-d}(j+d)^{j+d-\frac{1}{2}}}{(j+1)^{j+\frac{1}{2}}}\\
&= e^{-d+1}\lim_{j\to\infty} j^{1-d}\left(\frac{j+d}{j+1}\right)^{j+1/2}(j+d)^{d-1}\\
&= e^{-d+1}\lim_{j\to\infty}\left(\frac{j}{j+d}\right)^{1-d}\left(\frac{j+d}{j+1}\right)^{j+1/2}\\
&= e^{-d+1}\lim_{j\to\infty}\left(\frac{j}{j+d}\right)^{1-d}\left(\frac{j+d}{j}\right)^{j+1/2}\left(\frac{j}{j+1}\right)^{j+1/2},
\end{aligned}
$$

where $\left(\frac{j}{j+d}\right)^{1-d}\to 1$, $\left(\frac{j+d}{j}\right)^{j}\to e^d$ and $\left(\frac{j}{j+1}\right)^{j}\to e^{-1}$ as $j\to\infty$, so

$$
e^{-d+1}\lim_{j\to\infty}\left(\frac{j}{j+d}\right)^{1-d}\left(\frac{j+d}{j+1}\right)^{j+1/2} = e^{-d+1}e^{d-1} = 1 = RHS.
$$

□

Since

$$
\psi_j \approx \frac{1}{\Gamma(d)j^{1-d}}
$$

we have

$$
\sum_{j=0}^{\infty}\psi_j^2 \approx \frac{1}{(\Gamma(d))^2}\sum_{j=0}^{\infty}\frac{1}{j^{2(1-d)}},
$$

which is convergent if $2(1-d)>1$, i.e., $d<0.5$, therefore $\{\psi_j\}$ is square summable if and only if $d<0.5$ and in this paper we are interested in $d>0$. **The process in (1.3.1) with $0<d<0.5$ is denoted the ARFIMA(p,d,q) (AutoRegressive Fractionally Integrated Moving Average) model and the process X_i in (1.3.2) with $0<d<0.5$ is called the fractionally integrated noise.**

The linear representation (1.3.4) with the coefficients $\psi_j = \frac{1}{\Gamma(d)j^{1-d}}$ yields the following behaviour of ACF (Beran 1994, p. 63):

Lemma 1.3.2 *Consider the stationary ARFIMA(0,d,0) process with $d\in(0,1/2)$. Then*

$$
\rho_X(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)}k^{2d-1}, \qquad k\to\infty.
$$

In summary, the ARFIMA model itself has slowly decaying correlation. ARFIMA models cannot be used directly to model returns, but they will be used as a building block for stochastic volatility models.

1.4 Gaussian long memory sequences

1.4.1 Definition of long memory

There are different definitions of long memory. The most common form is given below (Beran 1994, p. 42):

Definition 1.4.1 *A stationary process $\{X_i, -\infty < i < \infty\}$ is said to be a short memory process if its autocorrelation function (ACF) $\rho_X(k)$ is absolutely summable:*

$$\sum_{k=0}^{\infty} |\rho_X(k)| < \infty. \tag{1.4.1}$$

On the other hand, a stationary process is said to have a long memory if ACF is not absolutely summable, i.e.

$$\sum_{k=-\infty}^{\infty} |\rho_X(k)| = \infty.$$

Example 1.4.2 If

$$|\rho_X(k)| \sim ck^{2d-1}, \qquad \text{as } k \to \infty, \tag{1.4.2}$$

where $d \in (0, 1/2)$, then ACF is not summable. This type of behaviour of ACF is usually assumed for purpose of limit theorems.

Example 1.4.3 Consider the AR(1) model of Example 1.2.2. If $|\phi| < 1$, then

$$\sum_{h=0}^{\infty} |\gamma_X(h)| = C \sum_{h=0}^{\infty} |\phi^h| < \infty.$$

Thus, the AR(1) model has short-memory.

Example 1.4.4 Consider the ARFIMA(0,d,0) model with $d \in (0, 1/2)$. Then

$$\sum_{h=0}^{\infty} |\gamma_X(h)| = \infty.$$

Thus, the ARFIMA(0,d,0) model has long-memory.

1.4.2 Gaussian sequences and Hermite polynomials

Assume that $\{X_i, -\infty < i < \infty\}$ is a stationary Gaussian process with unit variance and covariance

$$Cov(X_i, X_j) = \gamma_X(i-j) = |i-j|^{2d-1} L(|i-j|), \tag{1.4.3}$$

where $d \in (0, 1/2)$ and L is a slowly varying function at infinity.

We will assume for simplicity $L(i) = 1$, since if $d \in (0, 1/2)$,

$$\sum_{j=0}^{\infty} Cov(X_1, X_j) = \sum_{j=0}^{\infty} \gamma_X(j) = \sum_{j=0}^{\infty} j^{2d-1} = +\infty.$$

Thus, the sequence has long memory.

A crucial tool in studying long memory Gaussian processes are Hermite polynomials.

Definition 1.4.5 *For any $i = 0, 1, 2, \ldots$, the Hermite polynomials H_i are defined by the formula*

$$H_i(x) = (-1)^i \exp\left(\frac{x^2}{2}\right) \frac{d^i}{dx^i} \exp\left(\frac{-x^2}{2}\right).$$

Note that $\{H_i, -\infty < i < \infty\}$ is an orthogonal basis for Gaussian random variables, that is, for a standard normal random variable X with a density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2})$$

we have

$$< H_i(X), H_j(X) >= \int_{-\infty}^{\infty} H_i(x) H_j(x) \phi(x) dx$$

$$= \quad Cov(H_i(X), H_j(X)) = E(H_i(X)H_j(X)) = i! \qquad \text{if } i = j, \tag{1.4.4}$$

$$< H_i(X), H_j(X) >= 0, \qquad \text{if } i \neq j, \tag{1.4.5}$$

where

$$< f, g >= \int_{\infty}^{\infty} f(x)g(x)\phi(x)dx.$$

It also shows that the $Var(H_i(X)) = i!$ and $H_i(X), H_j(X)$ are uncorrelated if $i \neq j$, however, $H_i(X), H_j(X)$ are not independent.

The first 6 Hermite polynomials are:

$$H_0(x) = 1, H_1(x) = x, H_2(x) = x^2 - 1,$$

$$H_3(x) = x^3 - 3x, H_4(x) = x^4 - 6x^2 + 3, H_5(x) = x^5 - 10x^3 + 15x.$$

Lemma 1.4.6 (Arcones (1994)) *Let $a_1, \dots, a_k$ be real numbers such that $a_1^2 + \cdots + a_k^2 = 1$. Then*

$$H_q\left(\sum_{j=1}^k a_j x_j\right) = \sum_{q_1+\cdots+q_k=q} \frac{q!}{q_1! \cdots q_k!} \prod_{j=1}^k a_j^{q_j} H_{q_j}(x_j).$$

Lemma 1.4.7 *Let X, Y be a pair of jointly standard normal random variables with covariance $\gamma = cov(X, Y)$ and suppose that f is a measurable transformation with Hermite expansion*

$$f(x) = \sum_{i=0}^{\infty} \frac{a_i}{i!} H_i(x),$$

where $H_i(x)$ are Hermite polynomials and

$$a_i =< f(X), H_i(X) >= E(f(X)H_i(X)) = \int f(x)H_i(x)\phi(x)dx.$$

Then

$$cov(H_i(X), H_i(Y)) = i!\gamma^i, \tag{1.4.6}$$

$$cov(H_i(X), H_j(Y)) = 0, \qquad \textit{for } i \neq j, \tag{1.4.7}$$

$$cov(f(X), f(Y)) = \sum_{i=0}^{\infty} \frac{a_i^2}{i!} \gamma^i. \tag{1.4.8}$$

Proof of lemma 1.4.7: We can write $Y = \gamma X + \sqrt{1-\gamma^2} Z$, where Z is $N(0,1)$ and is independent of X. Using Lemma 1.4.6,

$$\begin{aligned} Cov(H_i(X), H_i(Y)) &= E(H_i(X)H_i(Y)) = E(H_i(X)H_i(\gamma X + \sqrt{1-\gamma^2} Z)) \\ &= E(H_i(X) \sum_{q_1+q_2=i} \frac{i!}{q_1!q_2!} \gamma^{q_1} \sqrt{1-\gamma^2}^{q_2} H_{q_1}(X) H_{q_2}(Z)) \\ &= \sum_{q_1+q_2=i} \frac{i!}{q_1!q_2!} \gamma^{q_1} \sqrt{1-\gamma^2}^{q_2} E(H_i(X)H_{q_1}(X)) E(H_{q_2}(Z)) \\ &= \gamma^i E(H_i(X)H_i(X)) = \gamma^i i!, \end{aligned}$$

where the last equality follows from equations 1.4.4 and 1.4.5 that $q_1 = i$ and $q_2 = 0$ in order to have $q_1 + q_2 = i$. Therefore (recall that $Y = \gamma X + \sqrt{1-\gamma^2} Z$),

$$\begin{aligned} Cov(f(X), f(Y)) &= E(f(X)f(Y)) = E\left(\sum_{i=0}^{\infty} \frac{a_i}{i!} H_i(X) \sum_{j=0}^{\infty} \frac{a_j}{j!} H_j(Y) \right) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \gamma^j \frac{a_i a_j}{i! j!} E(H_i(X)H_j(X)) = \sum_{i=0}^{\infty} \frac{a_i^2}{(i!)^2} \gamma^i i! = \sum_{i=0}^{\infty} \frac{a_i^2}{i!} \gamma^i. \end{aligned}$$

□

1.5 Stochastic Volatility models

Example 1.5.1 Let $\{Z_i, -\infty < i < \infty\}$ be i.i.d. $N(0,1)$ and $\{X_i, -\infty < i < \infty\}$ be an $AR(1)$ model, i.e. $X_i = \phi X_{i-1} + U_i$, where $\{U_i, -\infty < i < \infty\} \sim IID(0, \sigma_U^2)$. We assume that the sequences $\{X_i, -\infty < i < \infty\}$ and $\{Z_i, -\infty < i < \infty\}$ are mutually independent. We define

$$Y_i = X_i Z_i \,, \qquad -\infty < i < \infty.$$

We simulate 1000 observations from Y_i, $i = 1, \ldots, 1000$.

Figure 1.4 shows that the composite series is stationary. The ACF is almost zero for all lags $h > 0$ which shows that the series is **uncorrelated**. The squares are

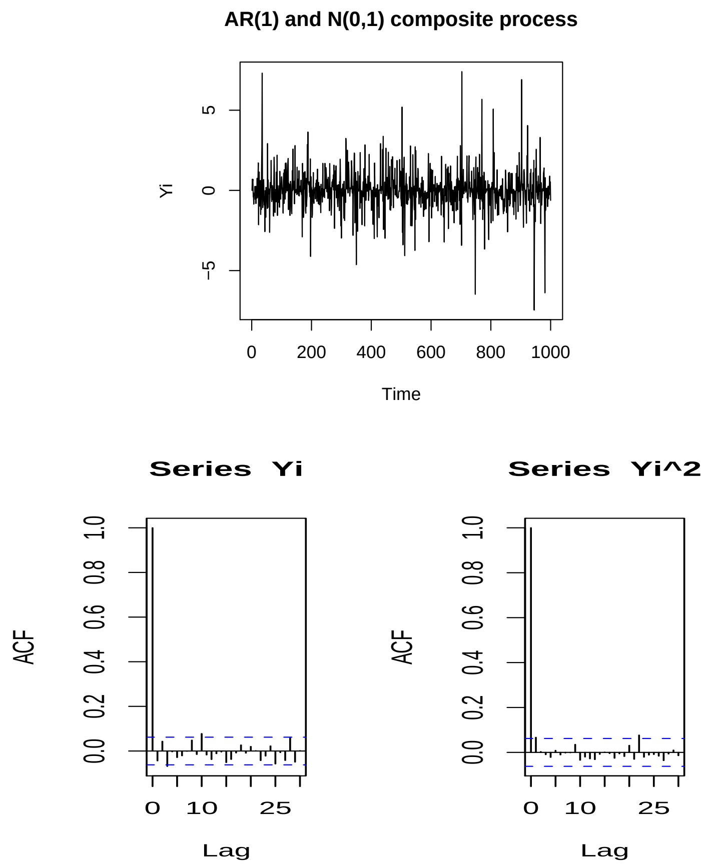

Figure 1.4: Simulated SV model and its ACF

correlated, we conclude that the sequence $\{Y_i, -\infty < i < \infty\}$ is **not independent**. Indeed, for $h \neq 0$

$$\begin{aligned} Cov(Y_i, Y_{i+h}) &= E(Y_i Y_{i+h}) = E(X_i Z_i X_{i+h} Z_{i+h}) \\ &= E(X_i X_{i+h}) E(Z_i) E(Z_{i+h}) = 0 \ , \end{aligned}$$

which means that the series $\{Y_i, -\infty < i < \infty\}$ is **uncorrelated**. On the other hand

$$\begin{aligned} Cov(Y_i^2, Y_{i+h}^2) &= E(Y_i^2 Y_{i+h}^2) - E(Y_i^2) E(Y_{i+h}^2) \\ &= E(Z_i^2 X_i^2 Z_{i+h}^2 X_{i+h}^2) - E(Z_i^2 X_i^2) E(Z_{i+h}^2 X_{i+h}^2) \end{aligned}$$

$$
\begin{aligned}
&= E(Z_i^2 Z_{i+h}^2)E(X_i^2 X_{i+h}^2) - E(Z_i^2)E(X_i^2)E(Z_{i+h}^2)E(X_{i+h}^2) \\
&= \{E(Z_i^2)E(Z_{i+h}^2)\}Cov(X_i^2, X_{i+h}^2) = (\mathrm{Var} Z_1)^2 Cov(X_i^2, X_{i+h}^2)
\end{aligned}
$$

where, in particular,

$$
\begin{aligned}
Cov(X_i^2, X_{i+1}^2) &= Cov(X_i^2, (\phi X_i + U_{i+1})^2) = Cov(X_i^2, \phi^2 X_i^2 + U_{i+1}^2 + 2\phi X_i U_{i+1}) \\
&= \phi^2 Cov(X_i^2, X_i^2) + Cov(X_i^2, U_{i+1}^2) + 2\phi Cov(X_i^2, X_i U_{i+1}) = \phi^2 \mathrm{Var}(X_i^2) \neq 0.
\end{aligned}
$$

Above we used the fact that X_i and U_{i+1} are independent. A similar computation can be carried out for a general $h \geq 1$. We conclude that that $\{X_i^2, -\infty < i < \infty\}$ is **correlated**. Thus, $Cov(Y_i^2, Y_{i+h}^2) \neq 0$, which means that $\{Y_i^2, -\infty < i < \infty\}$ is **correlated** as well. In conclusion,

$$
\begin{cases}
Cov(Y_i, Y_{i+h}) = 0 \text{ for } h \neq 0 & (Y_i \text{ uncorrelated}) \\
Cov(Y_i^2, Y_{i+h}^2) \neq 0 & (Y_i^2 \text{ correlated})
\end{cases}
$$

so $\{Y_i, -\infty < i < \infty\}$ is not an independent sequence and inherits covariance structure from $\{X_i^2, i \geq 0\}$.

The above example suggests the following definition of a stochastic volatility model.

Definition 1.5.2 *Without loss of generality, we will assume that $\sigma(\cdot)$ is a positive function. Let $\{X_i, -\infty < i < \infty\}$ be a stationary sequence, and $\{Z_i, -\infty < i < \infty\}$ be a sequence of i.i.d. random variables such that $E(Z_i) = 0$. Define a stochastic volatility (SV) process $\{Y_i, -\infty < i < \infty\}$ by*

$$
Y_i = \sigma(X_i) Z_i \; . \tag{1.5.1}
$$

Remark 1.5.3 For theoretical purposes, there is no need to assume that σ is positive. However, in financial applications $\sigma(X_i)$ plays a role of a conditional variance. Therefore, we keep this assumption throughout the thesis.

We note that in the above definition we do not impose any assumption about independence between the sequences $\{X_i, -\infty < i < \infty\}$ and $\{Z_i, -\infty < i < \infty\}$. However, in practice we need to consider certain models in order to be able to work with such processes. We shall assume that $\{X_i, -\infty < i < \infty\}$ is a Gaussian linear process (infinite order moving average)

$$X_i = \sum_{k=1}^{\infty} c_k \eta_{i-k}, \qquad -\infty < i < \infty,$$

where $\{\eta_i, -\infty < i < \infty\}$ is a sequence of independent standard normal random variables and $\sum_{k=1}^{\infty} c_k^2 = 1$. In particular, $\{X_i, -\infty < i < \infty\}$ are dependent standard normal random variables.

Two special cases which we are going to deal with are:

- Pure Stochastic Volatility (SV) model; where $\{\eta_i, -\infty < i < \infty\}$ and $\{Z_i, -\infty < i < \infty\}$, are independent.
- EGARCH model; where $\{(\eta_i, Z_i), -\infty < i < \infty\}$ is a sequence of i.i.d. random vectors. Thus, for fixed i, Z_i and X_i are independent. Of course, the EGARCH model includes SV.

In what follows, we will write "stochastic volatility" (SV) model for all models described by the Definition 1.5.2. We will write "pure SV" for the model described above, where we have independence between the sequences $\{X_i, -\infty < i < \infty\}$ and $\{Z_i, -\infty < i < \infty\}$.

The pure SV and EGARCH models are flexible enough to model both long memory and heavy tails. Indeed, as we argued in Example 1.5.1, the squares of both models inherit the dependence structure of $\{X_i, -\infty < i < \infty\}$. At the same time, we can choose $\{Z_i, -\infty < i < \infty\}$ to be heavy tailed. Let $F_Z(x) = P(Z \leq x)$ be the marginal cumulative distribution of the noise sequence $\{Z_i, -\infty < i < \infty\}$.

Definition 1.5.4 *A function L is called slowly varying at infinity if for all $c > 0$,*

$$\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1$$

Example 1.5.5 If we let $L(x) = x$, then $\frac{L(cx)}{L(x)} = \frac{cx}{x} = c$, thus $L(x) = x$ is not slowly varying.

Example 1.5.6 If we let $L(x) = \log x$, then $\frac{\log(cx)}{\log x} = \frac{\log c}{\log x} + \frac{\log x}{\log x} \to 1$, as $x \to \infty$. Thus, $L(x) = \log x$ is slowly varying.

We will assume that for some $\alpha \in (1, \infty)$, the tail of Z is

$$\overline{F}_Z(x) = P(Z > x) = x^{-\alpha} L(x), \qquad x \to \infty \tag{1.5.2}$$

where L is a slowly varying function. The parameter α is called the tail index. In particular, for each $c > 0$,

$$\lim_{x\to\infty} \frac{P(Z > cx)}{P(Z > x)} = \lim_{x\to\infty} \frac{(cx)^{-\alpha} L(cx)}{x^{-\alpha} L(x)} = c^{-\alpha}. \tag{1.5.3}$$

In fact, $\bar{F}_Z(x)$ is called regularly varying at infinity with index $-\alpha$ (see Resnick 2007).

Lemma 1.5.7 (Breiman Lemma, Resnick 2007, p. 231) *Define $Y_i = \sigma(X_i)Z_i$, assume that for fixed i the random variables X_i and Z_i are independent and that (1.5.2) holds. If $\mathrm{E}[\sigma^{\alpha+\epsilon}(X_i)] < \infty$, for some $\epsilon > 0$, then*

$$P(Y_i > x) \sim \mathrm{E}[\sigma^{\alpha}(X_i)] P(Z_i > x).$$

Breiman Lemma shows that the tail behaviour of Z is inherited by Y.

1.5.1 Long Memory in Stochastic Volatility

Example 1.5.1 gives us some idea how to introduce long memory in stochastic volatility. The following example shows how the long memory structure of $\sigma^2(X_i)$ is inherited by Y_i^2.

Example 1.5.8 Let $\{Z_i, i \geq 0\}$ be $IID(0, \sigma_Z^2)$, and let $\{X_i, t \geq 1\}$ be a long memory Gaussian sequence as in Section 1.4. Assume that $\{Z_i, -\infty < i < \infty\}$ and $\{X_i, -\infty < i < \infty\}$ are independent. Define $Y_i = \sigma(X_i)Z_i$. Then for $h \neq 0$,

$$E(Y_i) = E(Z_i\sigma(X_i)) = E(Z_i)E(\sigma(X_i)) = 0,$$

$$\begin{aligned} Cov(Y_i, Y_{i+h}) &= E(Y_iY_{i+h}) - E(Y_i)E(Y_{i+h}) = E(Y_iY_{i+h}) - 0 \\ &= E(Z_iZ_{i+h})E(\sigma(X_i)\sigma(X_{i+h})) = 0E(\sigma(X_i)\sigma(X_{i+h})) = 0 \end{aligned}$$

$$Cov(Y_i^2, Y_{i+h}^2) = (\mathrm{Var}Z_1)^2 Cov(\sigma^2(X_i), \sigma^2(X_{i+h}))$$

which shows that the SV model is uncorrelated, but Y_i^2 are correlated, hence the sequence $\{Y_i, -\infty < i < \infty\}$ is **not independent** and inherits covariance structure from $\{\sigma^2(X_i), -\infty < i < \infty\}$.

Take in particular $\sigma(x) = e^x$. Then using Lemma 1.4.7 with $f(x) = \sigma^2(x) = e^{2x}$ and Equation (1.4.3) for γ_X,

$$\begin{aligned} \mathrm{Cov}(\sigma^2(X_0), \sigma^2(X_h)) &= \mathrm{Cov}(f(X_0), f(X_h)) = \sum_{i=1}^{\infty} \frac{a_i^2}{(i!)^2} i! \gamma_X^i(h) \\ &= \sum_{i=1}^{\infty} \frac{a_i^2}{i!} \left(h^{(2d-1)}\right)^i \sim \frac{a_1^2}{1!}\gamma_X(h) \sim a_1^2 h^{(2d-1)} \end{aligned} \quad (1.5.4)$$

where

$$a_1 = E(\exp^{2X} X) \neq 0.$$

In other words, the covariance of $\sigma^2(X_i)$, up to a constant, is the same as that of $\{X_i, -\infty < i < \infty\}$.

Lemma 1.5.9 *Let $\{X_i, -\infty < i < \infty\}$ be a stationary LRD Gaussian process as defined in Section 1.4, with covariance defined in (1.4.3). If $d \in (0, 1/2)$, then*

$$Var(\bar{X}_n) \sim 2n^{2d-1} \frac{1}{2d(2d+1)} \sim \frac{C}{n^{(1-2d)}}.$$

The above result is standard in the long memory literature, see for example page 6 in Beran (1994).

For a comparison with iid or short memory sequences notice that $Var(\bar{X}_n) \sim \frac{C}{n}$.

1.5.2 Limit theorems under long memory

Recall the classical central limit theorem. If $\{X_i, -\infty < i < \infty\}$ are iid with mean zero and unit variance, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_i \xrightarrow{d} N(0,1),$$

The situation changes completely if long memory is involved. See Chapter 3 in Beran (1994).

Lemma 1.5.10 *Let $\{X_i, -\infty < i < \infty\}$ be a stationary LRD Gaussian process with mean zero and unit variance, as defined in Section 1.4, with covariance defined in (1.4.3). If $d \in (0, 1/2)$, then*

$$\frac{1}{n^{1/2+d}\sqrt{\frac{1}{d(2d+1)}}} \sum_{t=1}^{n} X_i \xrightarrow{d} N(0,1).$$

From equation (1.5.4), $Cov(f(X_0), f(X_i)) \sim a_1^2 \gamma_X(h)$ and

$$Var\left(\sum_{i=1}^{n} f(X_i)\right) \sim a_1^2 Var\left(\sum_{i=1}^{n} X_i\right).$$

It suggests the following lemma. For details see Chapter 3 in Beran (1994).

Lemma 1.5.11 *If f has the Hermite expansion given in Lemma 1.4.7, then the limiting behaviour of*

$$\sum_{t=1}^{n} \{f(X_i) - E(f(X_i))\}$$

is the same as that of $a_1 \sum_{t=1}^{n} X_i$, where $a_1 = E(f(X)X) \neq 0$:

$$\frac{1}{n^{1/2+d}\sqrt{\frac{1}{d(2d+1)}}} \sum_{t=1}^{n} \{f(X_i) - E(f(X_i))\} \xrightarrow{d} a_1 N(0,1) \equiv N(0, a_1^2).$$

Note: If $a_1 = 0$, the limiting behaviour is more complicated and it is beyond the scope of this thesis.

Chapter 2

High quantiles estimation for long memory stochastic volatility models

In this section we establish asymptotic theory for high quantile estimation of EGARCH and SV processes. Specifically, in section 2.1 we will establish the asymptotic behavior of a tail empirical process with deterministic levels. In section 2.2 we will obtain a corresponding behavior of a tail empirical process with random levels. The latter result will be used in section 2.3 to obtain asymptotic normality of the Hill estimator of the tail index.

Recall the Definition 1.5.2 of the EGARCH process. Let u_n be a sequence of constants such that:

- $u_n \to \infty$;
- $n\bar{F}(u_n) \to \infty$.

The associated conditional tail distribution function is defined for $x > 0$:

$$T_n(x) = P(Y > u_n(1+x) | Y > u_n) = \frac{\overline{F}(u_n(1+x))}{\overline{F}(u_n)}.$$

Under the assumption (1.5.3), we have

$$T_n(x) \to (1+x)^{-\alpha} = T(x) \text{ as } u_n \to \infty. \tag{2.0.1}$$

We also need the following assumption.

Assumption (A): If $i \neq j$, then for any sets A and B,

$$P(Y_i \in u_n A, Y_j \in u_n B) \sim C_{i,j} P(Y_i \in u_n A) P(Y_j \in u_n B), \qquad \text{as } n \to \infty,$$

where $u_n A = \{u_n x : x \in A\}$. Also, $\sup_{i,j} C_{i,j} \le C$.

This condition holds for example when Y_i is the SV model. It is also verified for EGARCH model, when $Z_i = Z_i' \Psi_1(\eta_i) + \Psi_2(\eta_i)$, where $\{Z_i', i \ge 1\}$ is the sequence of i.i.d. random variables with the same distribution as Z_i and $\{Z_i', i \ge 1\}$ is independent of $\{\eta_i, -\infty < i < \infty\}$.

Example 2.0.12 If $Y_i = \sigma(X_i) Z_i$ is a pure SV model, by using Breiman's Lemma 1.5.7, we have

$$\begin{aligned} P(Y_i > u_n y, Y_j > u_n y) &= E[P(Y_i > u_n y, Y_j > u_n y | X_i, X_j)] \\ &\sim E[\sigma^\alpha(X_i)\sigma^\alpha(X_j)] P(Z_i > u_n y) P(Z_j > u_n y) \\ &\sim \frac{E[\sigma^\alpha(X_i)\sigma^\alpha(X_j)]}{E[\sigma^\alpha(X_i)] E[\sigma^\alpha(X_j)]} P(Y_i > u_n y) P(Y_j > u_n y), \end{aligned}$$

so $C_{i,j} = \frac{E[\sigma^\alpha(X_i)\sigma^\alpha(X_j)]}{E[\sigma^\alpha(X_i)]E[\sigma^\alpha(X_j)]}$. In particular, if $\sigma(x) = e^x$ and using $(X_i + X_j) \sim N(0, 2 + 2\rho_{j-i})$, so $C_i, j = \frac{E[e^{\alpha(X_i+X_j)}]}{E[e^{\alpha X_i}]E[e^{\alpha X_j}]} = \frac{e^{\frac{\alpha^2}{2}(2+2\rho_{(j-i)})}}{e^{\frac{\alpha^2}{2}} e^{\frac{\alpha^2}{2}}}$. Since $|\ \rho \ | \le 1$, we have $C_{i,j} \le e^{\alpha^2} = C$.

2.1 Tail empirical process with deterministic levels

Define the empirical tail distribution function

$$\tilde{T}_n(s) = \frac{1}{n\overline{F}(u_n)} \sum_{i=1}^{n} 1_{\{Y_i > u_n(1+s)\}}, \tag{2.1.1}$$

and the tail empirical process

$$e_n(s) = \widetilde{T}_n(s) - T_n(s), \qquad s > 0.$$

Note that

$$E[\widetilde{T}_n(s)] = \frac{n}{n\overline{F}(u_n)} P(Y > u_n(1+s)) = \frac{\overline{F}(u_n(1+s))}{\overline{F}(u_n)} = T_n(s),$$

therefore, $T_n(s)$ is the proper centering. Our main result of this section is the following theorem.

Theorem 2.1.1 *Let $Y_i = \sigma(X_i)Z_i$ be the EGARCH model as in the definition 1.5.2. Assume that the tail of Z is regularly varying with index $-\alpha$ (see (1.5.2)) and $E[\sigma^\alpha(X)] < \infty$. Furthermore, assume that X_i is a long memory Gaussian process with covariance given by (1.4.2). Assume that (A) holds.*

- *If $n^{2d}\bar{F}(u_n) \to 0$, then*

$$\sqrt{n\bar{F}(u_n)}e_n(s) \Rightarrow W(T(s)),$$

  where W is a standard Brownian motion.

- *If $n^{2d}\bar{F}(u_n) \to \infty$, then*

$$\frac{1}{n^{d-1/2}\sqrt{\frac{1}{d(2d+1)}}} e_n(s) \Rightarrow T(s)\frac{E(X_i\sigma^\alpha(X_i))}{E(\sigma^\alpha(X_i))}N(0,1),$$

where $\Rightarrow$ denotes weak convergence in $D[0,\infty)$.

Remark 2.1.2 If $n^{2d}\bar{F}(u_n) \to c \in (0,\infty)$, then the limiting process is the sum of two processes which appear in Theorem 2.1.1, however, their joint distribution may be complicated.

Remark 2.1.3 We note that in the definition of $\tilde{T}_n(\cdot)$ we use the unknown quantity $\bar{F}(u_n)$. Consequently, the result for Theorem 2.1.1 is not applicable in practice. However, this result is used to conclude a limiting behaviour of a ”practical” tail empirical process considered in the next section.

Remark 2.1.4 The space $D[0,\infty)$ is defined in the Appendix.

The proof of this theorem will be divided into several steps. Decompose the tail empirical process into

$$\begin{aligned} e_n(s) &= \tilde{T}_n(s) - T_n(s) \\ &= \frac{1}{n\bar{F}(u_n)}\sum_{i=1}^{n}(1_{\{Y_i>u_n(1+s)\}} - E[1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1}]) \\ &\quad + \frac{1}{n\bar{F}(u_n)}\sum_{i=1}^{n}E[1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1}] - T_n(s) =: M_n(s) + R_n(s), \end{aligned}$$

where $\mathcal{F}_i = \sigma(\eta_i, \eta_{i-1}, \cdots, Z_i, Z_{i-1}, \cdots)$. Note that X_i is $\mathcal{F}_{i-1}$-measurable, since $X_i = \sum_{k=1}^{\infty} c_k\eta_{i-k}$ and Z_i is independent of $\mathcal{F}_{i-1}$.

In what follows, we will show that

$$\sqrt{n\bar{F}(u_n)}M_n(s) \Rightarrow W(T(s)).$$

This follows from Lemma 2.1.6 (finite dimensional convergence) and Lemma 2.1.8 (tightness).

Furthermore, in Lemma 2.1.9 we will show that

$$\frac{1}{n^{d-1/2}\sqrt{\frac{1}{d(2d+1)}}}R_n(s) \Rightarrow T(s)\frac{E(X_i\sigma^{\alpha}(X_i))}{E(\sigma^{\alpha}(X_i))}N(0,1)$$

From these two expressions, the theorem follows.

Note that $M_n(s)$, $n \geq 1$, is a martingale. Indeed if we let $V_i = 1_{\{Y_i>u_n(1+s)\}}$ and $M_i = \frac{1}{\sqrt{n\bar{F}(u_n)}}(V_i - E[V_i|\mathcal{F}_{i-1}])$, then

$$\begin{aligned} E(M_i|\mathcal{F}_{i-1}) &= \frac{1}{\sqrt{n\bar{F}(u_n)}}E\{(V_i - E[V_i|\mathcal{F}_{i-1}])|\mathcal{F}_{i-1}\} \\ &= \frac{1}{\sqrt{n\bar{F}(u_n)}}\{E(V_i|\mathcal{F}_{i-1}) - E[E(V_i|\mathcal{F}_{i-1})|\mathcal{F}_{i-1}]\} \\ &= \frac{1}{\sqrt{n\bar{F}(u_n)}}\{E(V_i|\mathcal{F}_{i-1}) - E(V_i|\mathcal{F}_{i-1})\} = 0. \end{aligned}$$

2.1.1 Weak convergence of the martingale part M_n

First, we will evaluate the variance of the martingale part (see Lemma 2.1.5). Then, we will prove a central limit theorem for the martingale part (see Lemma 2.1.6). That convergence is easily generalized to a finite dimensional convergence. Furthermore, we will show that the martingale part is tight (see Lemma 2.1.8).

Lemma 2.1.5 *Under the conditions of Theorem 2.1.1, we have*

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n\overline{F}(u_n)}\sum_{i=1}^{n}\{1_{\{Y_i>u_n(1+s)\}}-E(1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1})\}\right)\\
\sim\ & \frac{(1+s)^{-\alpha}}{n\overline{F}(u_n)}\sim\frac{(1+s)^{-\alpha}}{nE[\sigma^{\alpha}(X_i)]\overline{F}_Z(u_n)}.
\end{aligned}$$

Proof: Let $V_i=1_{\{Y_i>u_n(1+s)\}}$, then using the martingale property

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n\overline{F}(u_n)}\sum_{i=1}^{n}\{1_{\{Y_i>u_n(1+s)\}}-E(1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1})\}\right)\\
=\ & \frac{1}{n^2\overline{F}^2(u_n)}\sum_{i=1}^{n}\mathrm{Var}[V_i-E(V_i|\mathcal{F}_{i-1})]=\frac{n}{n^2\overline{F}^2(u_n)}\mathrm{Var}[V_i-E(V_i|\mathcal{F}_{i-1})]\\
=\ & \frac{1}{n\overline{F}^2(u_n)}\mathrm{Var}[V_i-E(V_i|\mathcal{F}_{i-1})]=\frac{1}{n\overline{F}^2(u_n)}\{E(V_i-E(V_i|\mathcal{F}_{i-1}))^2-0\}\\
=\ & \frac{1}{n\overline{F}^2(u_n)}\{E(V_i^2)+E[(E(V_i|\mathcal{F}_{i-1}))^2]-2E[V_iE(V_i|\mathcal{F}_{i-1})]\}. \qquad (2.1.2)
\end{aligned}$$

Thus, we split $\mathrm{Var}[V_i-E(V_i|\mathcal{F}_{i-1})]$ into 3 terms; the first term $A=E(V_i^2)$, the second one $B=E[(E(V_i|\mathcal{F}_{i-1}))^2]$ and the third term $C=E[V_iE(V_i|\mathcal{F}_{i-1})]$.

For the first term, we have

$$\begin{aligned}
A &= E(V_i^2)=E(V_i)=E[1_{\{Y_i>u_n(1+s)\}}]\\
&= P(Y_i>u_n(1+s))=\overline{F}(u_n(1+s)).
\end{aligned}$$

By Breiman's Lemma 1.5.7,

$$A\sim E[\sigma^{\alpha}(X_i)]\overline{F}_Z(u_n(1+s))\text{ as }u_n\to\infty,$$

and by (1.5.3),

$$\lim_{u_n\to\infty}\frac{P(Z>u_n(1+s))}{P(Z>u_n)}=(1+s)^{-\alpha}.$$

Thus,

$$A\sim E[\sigma^\alpha(X_i)](1+s)^{-\alpha}\overline{F}_Z(u_n)\sim(1+s)^{-\alpha}\overline{F}(u_n) \text{ as } u_n\to\infty. \tag{2.1.3}$$

For the second term, Using (1.5.3) we have,

$$\begin{aligned}\lim_{u_n\to\infty}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)} &= \lim_{u_n\to\infty}\frac{E(1_{\{\sigma(X_i)Z_i>u_n(1+s)\}}|X_i)}{P(Z>u_n)}\\ &= \lim_{u_n\to\infty}\frac{P(Z_i>\frac{u_n(1+s)}{\sigma(X_i)})}{P(Z>u_n)}=(1+s)^{-\alpha}\sigma^\alpha(X_i).\end{aligned} \tag{2.1.4}$$

Thus,

$$\begin{aligned}\lim_{u_n\to\infty}\frac{B}{P^2(Z>u_n)} &= \lim_{u_n\to\infty}E\left[\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\right]\\ &= E\left[\lim_{u_n\to\infty}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\lim_{u_n\to\infty}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\right]\\ &= E[(1+s)^{-2\alpha}\sigma^{2\alpha}(X_i)]=(1+s)^{-2\alpha}E[\sigma^{2\alpha}(X_i)].\end{aligned}$$

Consequently,

$$B\sim(1+s)^{-2\alpha}E[\sigma^{2\alpha}(X_i)]\overline{F}_Z^2(u_n) \text{ as } u_n\to\infty. \tag{2.1.5}$$

Here and in the sequel, we are allowed to interchange the limit with integration since $E[V_i\mid\mathcal{F}_{i-1}]$ is bounded.

For the third term, using (1.5.3) and (2.1.4), we have,

$$\begin{aligned}&\lim_{u_n\to\infty}\frac{C}{P^2(Z>u_n)}=\lim_{u_n\to\infty}\frac{E[V_iE(V_i|\mathcal{F}_{i-1})]}{P^2(Z>u_n)}\\ &= \lim_{u_n\to\infty}E\left(\frac{V_i}{P(Z>u_n)}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\right)=E\left(\lim_{u_n\to\infty}\frac{V_i}{P(Z>u_n)}\lim_{u_n\to\infty}\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\right)\\ &= E\left(\lim_{u_n\to\infty}\frac{V_i}{P(Z>u_n)}(1+s)^{-\alpha}\sigma^\alpha(X_i)\right)=\lim_{u_n\to\infty}E\left(\frac{V_i}{P(Z>u_n)}(1+s)^{-\alpha}\sigma^\alpha(X_i)\right)\\ &= (1+s)^{-\alpha}\lim_{u_n\to\infty}E\left(\frac{V_i}{P(Z>u_n)}\sigma^\alpha(X_i)\right).\end{aligned}$$

Furthermore,

$$\begin{aligned}
& (1+s)^{-\alpha} \lim_{u_n\to\infty} E\left(\frac{V_i}{P(Z>u_n)}\sigma^\alpha(X_i)\right) \\
= & \ (1+s)^{-\alpha} \lim_{u_n\to\infty} E\left(E[\frac{V_i}{P(Z>u_n)}\sigma^\alpha(X_i)|X_i]\right) \\
= & \ (1+s)^{-\alpha} \lim_{u_n\to\infty} E\left(\sigma^\alpha(X_i)E[\frac{V_i}{P(Z>u_n)}|X_i]\right) \\
= & \ (1+s)^{-\alpha} \lim_{u_n\to\infty} E\left(\sigma^\alpha(X_i)E[\frac{1_{\{\sigma(X_i)Z_i>u_n(1+s)\}}}{P(Z>u_n)}|X_i]\right) \\
= & \ (1+s)^{-\alpha} \lim_{u_n\to\infty} E\left(\sigma^\alpha(X_i)\frac{P(Z_i>\frac{u_n(1+s)}{\sigma(X_i)}|X_i)}{P(Z>u_n)}\right) \\
= & \ (1+s)^{-\alpha}E[\sigma^\alpha(X_i)(1+s)^{-\alpha}\sigma^\alpha(X_i)] = (1+s)^{-2\alpha}E[\sigma^{2\alpha}(X_i)]
\end{aligned}$$

Thus,

$$C \sim (1+s)^{-2\alpha}E[\sigma^{2\alpha}(X_i)]\overline{F}_Z^2(u_n) \text{ as } u_n\to\infty. \tag{2.1.6}$$

After comparing the results (2.1.3), (2.1.5) and (2.1.6), we find that A dominates B and C, so terms B and C are negligible. Consequently,

$$\mathrm{Var}[V_i - E(V_i|\mathcal{F}_{i-1})] \sim A \sim E[\sigma^\alpha(X_i)](1+s)^{-\alpha}\overline{F}_Z(u_n), \tag{2.1.7}$$

and

$$\begin{aligned}
\mathrm{Var}\sum_{i=1}^n[V_i - E(V_i|\mathcal{F}_{i-1})] &= \sum_{i=1}^n \mathrm{Var}[V_i - E(V_i|\mathcal{F}_{i-1})] = n\mathrm{Var}[V_i - E(V_i|\mathcal{F}_{i-1})] \\
&\sim nE[\sigma^\alpha(X_i)](1+s)^{-\alpha}\overline{F}_Z(u_n) \sim n(1+s)^{-\alpha}\overline{F}(u_n).
\end{aligned}$$

By (2.1.2), (2.1.3) and (2.1.7), we get

$$\mathrm{Var}\left(\frac{1}{n\overline{F}(u_n)}\sum_{i=1}^n\{1_{\{Y_i>u_n(1+s)\}} - E(1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1})\}\right) \sim \frac{(1+s)^{-\alpha}}{n\overline{F}(u_n)}.$$

$\square$

Since $\sum_{i=1}^n[V_i - E(V_i|\mathcal{F}_{i-1})]$ is a martingale, it suggests the following result.

Lemma 2.1.6 *Under the conditions of Theorem 2.1.1 we have*

$$\sqrt{n\overline{F}(u_n)}M_n(s) = \frac{1}{\sqrt{n\overline{F}(u_n)}}\sum_{i=1}^{n}[V_i - E(V_i|\mathcal{F}_{i-1})] \xrightarrow{d} N(0,(1+s)^{-\alpha}).$$

To prove the above lemma, we recall the following martingale CLT (see Hall and Heyde (1980)).

Lemma 2.1.7 (Martingale CLT) *Let* $((X_{ni}, A_{n,i}) : n \geq 1; i = 1, \cdots, k_n)$ *be a martingale difference array; that is, for each fixed* n, X_{ni} *is a martingale difference sequence. If*

- $\sum_{i=1}^{k_n} E(X_{ni}^2|\mathrm{A}_{n,i-1}) \xrightarrow{p} \sigma^2$, *and*
- $\sum_{i=1}^{k_n} E[X_{ni}^2 \mathrm{I}(|X_{ni}| > \epsilon)|\mathrm{A}_{n,i-1}] \xrightarrow{p} 0$ *for each* $\epsilon > 0$

then

$$Z_n = \sum_{i=1}^{k_n} X_{ni} \xrightarrow{d} Z \sim N(0,\sigma^2).$$

We apply the above theorem to the scaled martingale part: $Z_n = \sqrt{n\overline{F}(u_n)}M_n(s)$ and $X_{ni} = M_i = \frac{1}{\sqrt{n\overline{F}(u_n)}}[V_i - E(V_i|\mathcal{F}_{i-1})]$ and $A_{n,i} = \mathcal{F}_i$. Let $\overline{M}_i = V_i - E(V_i|\mathcal{F}_{i-1})$, so that

$$M_i^2 \;=\; \frac{1}{n\overline{F}(u_n)}[V_i - E(V_i|\mathcal{F}_{i-1})]^2 = \frac{1}{n\overline{F}(u_n)}\overline{M}_i^2.$$

Then

$$\begin{aligned}
&E(\overline{M}_i^2|\mathcal{F}_{i-1}) \\
&= E[(V_i - E(V_i|\mathcal{F}_{i-1}))^2|\mathcal{F}_{i-1}] \\
&= E[[V_i^2 + (E(V_i|\mathcal{F}_{i-1}))^2 - 2V_iE(V_i|\mathcal{F}_{i-1})]|\mathcal{F}_{i-1})] \\
&= E(V_i^2|\mathcal{F}_{i-1}) + E[(E(V_i|\mathcal{F}_{i-1}))^2|\mathcal{F}_{i-1}] - 2E[V_iE(V_i|\mathcal{F}_{i-1})|\mathcal{F}_{i-1}],
\end{aligned}$$

where

$$E(V_i^2|\mathcal{F}_{i-1}) = E(V_i|\mathcal{F}_{i-1}),$$

$$E[(E(V_i|\mathcal{F}_{i-1}))^2|\mathcal{F}_{i-1}] = (E(V_i|\mathcal{F}_{i-1}))^2,$$

$$E[V_i E(V_i|\mathcal{F}_{i-1})|\mathcal{F}_{i-1}] = E(V_i|\mathcal{F}_{i-1})E(V_i|\mathcal{F}_{i-1}) = (E(V_i|\mathcal{F}_{i-1}))^2.$$

Thus,

$$E(\overline{M}_i^2|\mathcal{F}_{i-1}) = E(V_i|\mathcal{F}_{i-1}) - (E(V_i|\mathcal{F}_{i-1}))^2, \tag{2.1.8}$$

and recall that,

$$\lim_{u_n\to\infty} \frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)} = \lim_{u_n\to\infty} \frac{P(Z_i > \frac{u_n(1+s)}{\sigma(X_i)})}{P(Z>u_n)} = (1+s)^{-\alpha}\sigma^\alpha(X_i).$$

We compute that

$$\begin{aligned} \frac{E(\overline{M}_i^2|\mathcal{F}_{i-1})}{P(Z>u_n)} &= \frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)} - \left(\frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)}\right)^2 P(Z>u_n) \\ &\sim \frac{E(V_i|\mathcal{F}_{i-1})}{P(Z>u_n)} \sim (1+s)^{-\alpha}\sigma^\alpha(X_i). \end{aligned}$$

Thus,

$$E(\overline{M}_i^2|\mathcal{F}_{i-1}) \sim (1+s)^{-\alpha}\sigma^\alpha(X_i)\overline{F}_Z(u_n), \tag{2.1.9}$$

consequently, using (2.1.9) and Breiman's Lemma (1.5.7)

$$\begin{aligned} \sum_{i=1}^n E(M_i^2|\mathcal{F}_{i-1}) &= \frac{1}{n\overline{F}(u_n)} \sum_{i=1}^n E(\overline{M}_i^2|\mathcal{F}_{i-1}) \sim \frac{(1+s)^{-\alpha}\overline{F}_Z(u_n)}{n\overline{F}(u_n)} \sum_{i=1}^n (\sigma^\alpha(X_i)) \\ &\sim \frac{(1+s)^\alpha}{n} \sum_{i=1}^n (\sigma^\alpha(X_i)) \frac{\overline{F}_Z(u_n)}{E(\sigma^\alpha(X_i))\overline{F}_Z(u_n)} \xrightarrow{P} (1+s)^{-\alpha}. \end{aligned}$$

where $\lim_{n\to\infty} \frac{\sum_{i=1}^n (\sigma^\alpha(X_i))}{n} = E(\sigma^\alpha(X_i))$ follows directly from the Ergodic Theorem and the first condition in the martingale CLT is verified. We note that the sequence X_i is ergodic, since $X_i = \varphi(\varepsilon_i, \varepsilon_{i-1}, \ldots,)$, where ε_i are i.i.d. normal random variables.

$\square$

For the second condition,

$$\begin{aligned} & E\left(\frac{\overline{M}_i^2}{n\overline{F}(u_n)}1_{\{|\overline{M}_i|>\varepsilon\sqrt{n\overline{F}(u_n)}\}}|\mathcal{F}_{i-1}\right) \\ = & \frac{1}{n\overline{F}(u_n)}E\left([V_i - E(V_i|\mathcal{F}_{i-1})]^2 1_{\{|\overline{M}_i|>\varepsilon\sqrt{n\overline{F}(u_n)}\}}|\mathcal{F}_{i-1}\right) = 0 \end{aligned}$$

since $|\overline{M}_i|$ is bounded and ≤ 2 and $n\overline{F}(u_n) \to \infty$ as $n \to \infty$, the indicator function returns a value 0 for n sufficiently large. Thus, the second condition in the martingale CLT is

$$\sum_{i=1}^{n} E\left(M_i^2 1_{\{|M_i|>\varepsilon\}}|\mathcal{F}_{i-1}\right) \xrightarrow{P} 0.$$

$\square$

Recall that

$$\begin{aligned} M_n(s) &= \frac{1}{n\overline{F}_Z(u_n)}\sum_{i=1}^{n}\left(1_{\{Y_i>u_n(1+s)\}} - E[1_{\{Y_i>u_n(1+s)\}}|\mathcal{F}_{i-1}]\right) \\ &= \frac{1}{n\overline{F}(u_n)}\sum_{i=1}^{n}(V_i(s) - E(V_i(s)|\mathcal{F}_{i-1})) \end{aligned}$$

Lemma 2.1.8 *Define $\widetilde{M}_n(s) = \sqrt{n\overline{F}(u_n)}M_n(s)$. Under the conditions of Theorem 2.1.1, the process $\widetilde{M}_n(s)$ is tight.*

Proof:

Using Theorem 13.5 in Billingsley (1968), in order to check tightness we have to verify the following condition: for $s_1 \leq s_2 \leq s_3$, we have to bound

$$E[|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^2|\widetilde{M}_n(s_3) - \widetilde{M}_n(s_2)|^2]. \tag{2.1.10}$$

By Hölder's inequality,

$$\begin{aligned} & E[|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^2|\widetilde{M}_n(s_3) - \widetilde{M}_n(s_2)|^2] \\ \leq & \left[E\left(|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^4\right)\right]^{\frac{1}{2}}\left[E\left(|\widetilde{M}_n(s_3) - \widetilde{M}_n(s_2)|^4\right)\right]^{\frac{1}{2}}, \end{aligned}$$

where

$$
\begin{aligned}
& E(|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^4) \\
= \;& \frac{1}{(n\overline{F}(u_n))^2} E\left(\sum_{i=1}^{n} [V_i(s_2) - V_i(s_1)] - [E(V_i(s_2)|\mathcal{F}_{i-1}) - E(V_i(s_1)|\mathcal{F}_{i-1})]\right)^4.
\end{aligned}
$$

If we denote $V_i(s) - E(V_i(s)|\mathcal{F}_{i-1}) = X_i(s)$ and $X_i(s_2) - X_i(s_1) = \widetilde{X}_i$, then,

$$
[V_i(s_2) - V_i(s_1)] - [E(V_i(s_2)|\mathcal{F}_{i-1}) - E(V_i(s_1)|\mathcal{F}_{i-1})] = X_i(s_2) - X_i(s_1) = \widetilde{X}_i.
$$

then,

$$
\begin{aligned}
& E(|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^4) \\
= \;& \frac{1}{(n\overline{F}(u_n))^2} E(\sum_{i=1}^{n} \widetilde{X}_i)^4 \leq \frac{1}{(n\overline{F}(u_n))^2} E(\sum_{i=1}^{n} \widetilde{X}_i^2)^2 \\
= \;& \frac{1}{(n\overline{F}(u_n))^2} E(\sum_{i=1}^{n} \widetilde{X}_i^4 + \sum_{i,j=1, i\neq j}^{n} (\widetilde{X}_i^2 \widetilde{X}_j^2)) \\
= \;& \frac{1}{(n\overline{F}(u_n))^2} (\sum_{i=1}^{n} E\widetilde{X}_i^4 + \sum_{i,j=1, i\neq j}^{n} E(\widetilde{X}_i^2 \widetilde{X}_j^2)) \\
\leq \;& \frac{1}{(n\overline{F}(u_n))^2} \left(nE(\widetilde{X}_1^4) + \sum_{i,j=1, i\neq j}^{n} E(\widetilde{X}_i^2 \widetilde{X}_j^2) \right) \qquad \text{(Stationarity)}.
\end{aligned}
$$

The first inequality is the standard inequality for martingales. (Hall, Heyde 1980) Furthermore,

$$
E(\widetilde{X}_1^4) \leq CE(V_1(s_2) - V_1(s_1))^4 = CP(u_n(1+s_1) < Y < u_n(1+s_2))
$$

for some constant C and note that

$$
\lim_{n\to\infty} \frac{P(u_n(1+s_1) < Y < u_n(1+s_2))}{\bar{F}(u_n)} = T(s_1) - T(s_2) \leq T(s_1) - T(s_3).
$$

Furthermore, if assumption A holds,

$$
E(\widetilde{X}_i^2 \widetilde{X}_j^2) \leq CE[V_i(s_2) - V_i(s_1)]^2 [V_j(s_2) - V_j(s_1)]^2
$$

$$
\begin{aligned}
&= CP(u_n(1+s_1) < Y_i < u_n(1+s_2), u_n(1+s_1) < Y_j < u_n(1+s_2)) \\
&\sim C_{i,j}P(u_n(1+s_1) < Y_i < u_n(1+s_2)) \times P(u_n(1+s_1) < Y_j < u_n(1+s_2)),
\end{aligned}
$$

and so

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{C_{i,j}P(u_n(1+s_1) < Y_i < u_n(1+s_2)) * P(u_n(1+s_1) < Y_j < u_n(1+s_2))}{\bar{F}(u_n)\bar{F}(u_n)} \\
&= C_{i,j}(T(s_1)-T(s_2))(T(s_1)-T(s_2)) = C_{i,j}(T(s_1)-T(s_2))^2 \le C_{i,j}(T(s_1)-T(s_3))^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E(|\widetilde{M}_n(s_2) - \widetilde{M}_n(s_1)|^4) \\
&\le \frac{1}{n\bar{F}(u_n)}(T(s_1)-T(s_3)) + \frac{1}{n^2}\left(\sum_{i,j=1,i\neq j}^{n} C_{i,j}\right)(T(s_1)-T(s_3))^2 \\
&= \frac{1}{n\bar{F}(u_n)}(T(s_1)-T(s_3)) + C\left(\frac{n(n-1)}{n^2}\right)(T(s_1)-T(s_3))^2 \\
&\le \frac{1}{n\bar{F}(u_n)}(T(s_1)-T(s_3)) + C(T(s_1)-T(s_3))^2.
\end{aligned}
$$

Analogously,

$$
E(|\widetilde{M}_n(s_3) - \widetilde{M}_n(s_2)|^4) \le \frac{1}{n\bar{F}(u_n)}(T(s_1)-T(s_3)) + C(T(s_1)-T(s_3))^2.
$$

Thus,

$$
E[|\widetilde{M}_n(s_1)-\widetilde{M}_n(s_2)|^2|\widetilde{M}_n(s_2)-\widetilde{M}_n(s_3)|^2] \le \frac{1}{n\bar{F}(u_n)}(T(s_1)-T(s_3))+C(T(s_1)-T(s_3))^2.
$$

This guarantees tightness. □

2.1.2 Weak convergence of long memory part R_n

The term R_n is treated using Lemma 1.5.11. Note that

$$
R_n(s) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{G}(X_i)
$$

where

$$
\widetilde{G}(X_i) = G(X_i) - E(G(X_i))
$$

and

$$G(X_i) = \frac{1}{\overline{F}(u_n)} E[1_{\{Y_i > u_n(1+s)\}} | \mathcal{F}_{i-1}] = \frac{1}{\overline{F}(u_n)} E[1_{\{\sigma(X_i) Z_i > u_n(1+s)\}} | \mathcal{F}_{i-1}].$$

Note that function G depends on n and s. However, $R_n(s)$ has a form similar to the expression in Lemma 1.5.11.

We have to compute

$$\begin{aligned}
a_1 &= E(G(X_i)X_i) = E\{\frac{1}{\overline{F}(u_n)} E[1_{\{\sigma(X_i)Z_i > u_n(1+s)\}} | \mathcal{F}_{i-1}] X_i\} \\
&= \frac{1}{\overline{F}(u_n)} E\{E[1_{\{\sigma(X_i)Z_i > u_n(1+s)\}} X_i | \mathcal{F}_{i-1}]\} = \frac{1}{\overline{F}(u_n)} E(1_{\{\sigma(X_i)Z_i > u_n(1+s)\}} X_i) \\
&= E\left(X_i \frac{1_{\{\sigma(X_i)Z_i > u_n(1+s)\}}}{\overline{F}(u_n)} \right) = E\left(E[X_i \frac{1_{\{\sigma(X_i)Z_i > u_n(1+s)\}}}{\overline{F}(u_n)} | X_i] \right) \\
&= E\left(X_i E[\frac{1_{\{\sigma(X_i)Z_i > u_n(1+s)\}}}{\overline{F}(u_n)} | X_i] \right) = E\left(X_i \frac{P(Z_i > \frac{u_n(1+s)}{\sigma(X_i)} | X_i)}{\overline{F}(u_n)} \right) \\
&\sim E\left(X_i \frac{(1+s)^{-\alpha} \sigma^\alpha(X_i)}{E(\sigma^\alpha(X_i))} \right) = (1+s)^{-\alpha} \frac{E(X_i \sigma^\alpha(X_i))}{E(\sigma^\alpha(X_i))}.
\end{aligned}$$

This suggests the following limiting result. The proof is not given here, see Kulik and Soulier (2011). We note that their proof for the long memory part $R_n(s)$ does not involve SV assumption and thus it holds for EGARCH model as well.

Lemma 2.1.9 *Under the conditions of Theorem 2.1.1,*

$$\frac{1}{n^{d-1/2}\sqrt{\frac{1}{d(2d+1)}}} R_n(s) \Rightarrow T(s) \frac{E(X_i \sigma^\alpha(X_i))}{E(\sigma^\alpha(X_i))} N(0,1)$$

We note that the limiting process is degenerate, i.e. it is a standard normal random variable, multiplied by a deterministic function.

2.2 Tail empirical process with random levels

Recall the definition (2.1.1). We replace there $n\bar{F}(u_n)$ with k_n and u_n with $Y_{n-k_n:n}$, the (k_n+1)th largest observation, in order to define the following process:

$$\hat{T}_n(s) = \frac{1}{k_n}\sum_{j=1}^{n} 1_{\{Y_j > Y_{n-k_n:n}(1+s)\}},$$

$$T(s) = (1+s)^{-\frac{1}{\gamma}}, \text{ where } \gamma = \frac{1}{\alpha},$$

$$\hat{e}_n^*(s) = \hat{T}_n(s) - T(s).$$

It is assumed that $k_n \to \infty$ and as $n \to \infty$.

The process $\hat{e}_n^*(s)$ is the modification of the process e_n considered in Section 2.1: instead of the deterministic level u_n, we have a random level. We will show below that this process has different convergence properties than e_n.

The statement of the Theorem 2.1.1 can be re-written as

$$a_n e_n(s) = a_n(\tilde{T}_n(s) - T_n(s)) \Rightarrow \Psi(s),$$

where the scaling a_n and the limiting process $\Psi(s)$ have two different forms. We will argue below that

$$a_n \hat{e}_n^*(s) \Rightarrow \Psi(s) - T(s)\Psi(0).$$

First, we have to replace the centering $T_n(s)$ with $T(s)$. This is allowed under the so-called second order regular variation condition (see Kulik and Soulier (2011) for details). In particular, if $\bar{F}(x) = x^{-\alpha} + x^{-\beta}$, $x > 1$, $\beta > \alpha$, then the second order regular variation holds.

Lemma 2.2.1 *From Theorem 2.1.1 and under second order regular variation,*

- *In the short memory case, if* $a_n = \sqrt{n\bar{F}(u_n)} = \sqrt{k_n}$, $\Psi(s) = W(T(s))$ *then,*

$$\sqrt{k_n}\hat{e}_n^*(s) \Rightarrow W(T(s)) - T(s)W(1) = B(T(s)), \tag{2.2.1}$$

where $B(\cdot)$ *is the Brownian bridge.*

- *In the long memory case, if* $a_n = \text{const}.n^{1/2-d}$, $\Psi(s) = \frac{E[X\sigma^\alpha(X)]}{E[\sigma^\alpha(X)]}T(s)N(0,1)$, *then,*

$$a_n\hat{e}_n^*(s) \Rightarrow \frac{E[X\sigma^\alpha(X)]}{E[\sigma^\alpha(X)]}T(s)N(0,1) - \frac{E[X\sigma^\alpha(X)]}{E[\sigma^\alpha(X)]}T(s)N(0,1) = 0.$$

Therefore, in the case of random levels, the long memory scaling is too big. It is possible to show that (2.2.1) holds for arbitrary $d \in (0, 1/2)$, see Kulik and Soulier (2011).

Generally, for any choice of $d \in (0, 1/2)$, we will have

$$\sqrt{k_n}\hat{e}_n^*(s) \Rightarrow B(T(s)). \tag{2.2.2}$$

Lemma 2.2.2 (Vervaat's Lemma) *Assume that* s_n *is a sequence of constants, and* $x_n(\cdot)$, $\rho(\cdot)$ *are deterministic functions and* $x_n^{\leftarrow}(\cdot)$, $\rho^{\leftarrow}(\cdot)$ *are the generated inverse function, If* $s_n(x_n(s) - \rho(s)) \to Y(s)$, *then*

$$s_n(x_n^{\leftarrow}(s) - \rho^{\leftarrow}(s)) \to -(\rho^{\leftarrow}(s))'Y(\rho^{\leftarrow}(s))$$

Proof of lemma 2.2.1: Define $\delta_n = \frac{Y_{n-k_n:n} - u_n}{u_n}$. We have

$$\tilde{T}_n(s + \delta_n(1+s)) = \hat{T}_n(s),$$

and

$$\hat{e}_n^*(s) = \tilde{T}_n(s + \delta_n(1+s)) - T(s) = e_n(s + \delta_n(1+s)) + T_n(s + \delta_n(1+s)) - T(s).$$

Under second order regular variation we can replace T_n with T. Thus, heuristically,

$$\hat{e}_n^*(s) \approx e_n(s + \delta_n(1+s)) + T(s + \delta_n(1+s)) - T(s).$$

The first term in the Taylor's expansion of $T(s+\delta_n(1+s)) - T(s)$ is

$$T'(s)\delta_n(1+s) = -\alpha T(s)\delta_n = T'(0)T(s)\delta_n.$$

Thus, we expect

$$\hat{e}_n^*(s) \approx \hat{e}_n(s+\delta(1+s)) + T'(0)T(s)\delta_n. \tag{2.2.3}$$

Now, we want to show that

$$\sqrt{k_n}T'(0)\delta_n \xrightarrow{d} -W(1).$$

Using Vervaat's Lemma 2.2.2, if we take

$$x_n(s) = \tilde{T}_n(s), \rho(s) = T(s) \text{ and } s_n = \sqrt{k_n} = \sqrt{n\bar{F}(u_n)}$$

and we have from Theorem 2.1.1, for the weakly dependence case,

$$\sqrt{k_n}(\tilde{T}_n(s) - T(s)) \xrightarrow{d} \Psi(s) = W(T(s)).$$

Then

$$\sqrt{k_n}(\tilde{T}_n^{\leftarrow}(s) - T^{\leftarrow}(s)) \xrightarrow{d} -(T^{\leftarrow}(s))'W(TT^{\leftarrow}(s)) = -(T^{\leftarrow}(s))'W(s).$$

Take $s=1$, we have

$$\sqrt{k_n}(\tilde{T}_n^{\leftarrow}(1) - T^{\leftarrow}(1)) \xrightarrow{d} -(T^{\leftarrow}(1))'W(1).$$

Note that $T^{\leftarrow}(s) = s^{-\gamma} - 1, (T^{\leftarrow}(s))' = -\gamma s^{-\gamma-1}$, Thus $(T^{\leftarrow}(1))' = -\gamma = -1/\alpha$. We conclude

$$\sqrt{k_n}(\tilde{T}_n^{\leftarrow}(1) - T^{\leftarrow}(1)) \xrightarrow{d} 1/\alpha W(1)$$

and $\alpha = -T'(0)$. Equivalently,

$$\sqrt{k_n}T'(0)(\tilde{T}_n^{\leftarrow}(1) - T^{\leftarrow}(1)) \xrightarrow{d} -W(1).$$

Note also that

$$\tilde{T}_n^{\leftarrow}(1) = \frac{Y_{n-k_n:n}}{u_n} - 1 = \delta_n$$

and $T^{\leftarrow}(1) = 0$. Thus, we conclude

$$\sqrt{k_n} T'(0)\delta_n \stackrel{d}{\to} -W(1),$$

which also ensures that $\delta_n \to 0$ as $\sqrt{k_n} \to \infty$. Therefore,

$$\sqrt{k_n}\hat{e}_n^*(s) \approx \sqrt{k_n} e_n(s) + \sqrt{k_n} T'(0) T(s)\delta_n \approx \Psi(s) - T(s)\Psi(0).$$

and the proof is complete. $\square$

2.3 Asymptotic normality of the Hill estimator

Using results from Section 2.2, in this section we establish asymptotic normality of Hill estimator based on a stationary EGARCH model. Let $Y_{1:n} \le Y_{2:n} \le \ldots \le Y_{n-1:n} \le Y_{n:n}$ be the increasing order statistics of a stationary sequence $Y_1, \ldots, Y_n$. The Hill estimator is defined as

$$H_{k_n,n} = \hat{\gamma}_n = \frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{Y_{n-j+1:n}}{Y_{n-k_n:n}}.$$

The tail empirical distribution $\hat{T}_n(s)$ and the tail empirical process $\hat{e}_n^*(s)$ play a crucial role in the study the Hill estimator. Indeed:

Lemma 2.3.1 *We have*

$$\alpha^{-1} = \gamma = \int_0^\infty \frac{T(s)}{1+s}\, ds, \tag{2.3.1}$$

$$\hat{\gamma}_n = \int_0^\infty \frac{\hat{T}_n(s)}{1+s}\, ds \tag{2.3.2}$$

and

$$\int_0^\infty \frac{B(T(s))}{1+s}\, ds \stackrel{d}{=} \gamma \int_0^1 \frac{B(t)}{t}\, dt, \tag{2.3.3}$$

where B is a standard Brownian bridge.

Proof: For 1. notice that

$$1. \int_0^\infty \frac{T(s)}{1+s}\,ds = \int_0^\infty \frac{(1+s)^{-\frac{1}{\gamma}}}{1+s}\,ds = \int_0^\infty (1+s)^{-\frac{1}{\gamma}-1}\,ds$$
$$= -\gamma(1+s)^{-\frac{1}{\gamma}}\Big|_0^\infty = \gamma(1+0)^{-\frac{1}{\gamma}} = \gamma.$$

For 2, define $\hat{T}_n^*(s) = \frac{1}{k_n}\sum_{j=1}^n 1_{\{Y_j > Y_{n-k_n:n}(s)\}}$, where $s \geq 1$.

$$\int_0^\infty \frac{\hat{T}_n(s)}{1+s}\,ds = \int_1^\infty \frac{\hat{T}_n^*(s)}{s}\,ds = \int_1^\infty \frac{1}{k_n}\sum_{j=1}^n 1_{\{Y_j > sY_{n-k:n}\}}\frac{1}{s}\,ds$$
$$= \frac{1}{k_n}\sum_{j=1}^n \int_1^\infty 1_{\{Y_j > sY_{n-k_n:n}\}}\frac{1}{s}\,ds. \tag{2.3.4}$$

Note that $1_{\{Y_j > Y_{n-k_n:n}(s)\}} = 1_{\{\frac{Y_j}{Y_{n-k_n:n}} > s\}}$ and $s \geq 1$, Y_j must be bigger than $Y_{n-k_n:n}$. This means that $Y_j = Y_{n-k_n+1:n}$ or $Y_j = Y_{n-k_n+2:n}, \cdots$, or $Y_j = Y_{n:n}$. Therefore, the sum in (2.3.4) becomes

$$\frac{1}{k_n}\sum_{j=1}^n \int_1^{\frac{Y_j}{Y_{n-k_n:n}}} \frac{1}{s}\,ds = \frac{1}{k_n}\sum_{j=1}^{k_n} \int_1^{\frac{Y_{n-j+1:n}}{Y_{n-k_n:n}}} \frac{1}{s}\,ds = \frac{1}{k_n}\sum_{j=1}^{k_n} \log\left(\frac{Y_{n-j+1:n}}{Y_{n-k_n:n}}\right) = \hat{\gamma}_n.$$

For 3, using the substitution rule, by letting $t = T(s) = (1+s)^{-\frac{1}{\gamma}}$, then $dt = (-\frac{1}{\gamma})(1+s)^{-\frac{1}{\gamma}-1}ds$, so $\frac{ds}{1+s} = (-\gamma)(1+s)^{\frac{1}{\gamma}}dt = (-\gamma)\frac{1}{t}dt$. Therefore

$$\int_0^\infty \frac{B(T(s))}{1+s}\,ds = \int_0^1 B(t)(-\gamma)\frac{1}{t}dt = (-\gamma)\int_0^1 \frac{B(t)}{t}\,dt \overset{d}{=} \gamma\int_0^1 \frac{B(t)}{t}\,dt.$$

(Since Brownian motion increments are normally distributed, therefore the integral is symmetric.)

$\square$

From (2.3.1)-(2.3.2) and (2.2.2) we conclude

$$\sqrt{k_n}\,(\hat{\gamma}_n - \gamma) = \sqrt{k_n}\int_0^\infty \frac{\hat{e}_n^*(s)}{1+s}ds \xrightarrow{\text{d}} \int_0^\infty \frac{B(T(S))}{1+s}ds \overset{\text{d}}{=} \gamma\int_0^1 \frac{B(t)}{t}\,dt.$$

We note that the latter integral is a standard normal random variable. Therefore, we obtain the following result.

Recall that $k = k_n \to \infty$ as $n \to \infty$. We note also that k_n is the user-chosen number of extreme observations.

Theorem 2.3.2 *Under the assumptions of Theorem 2.1.1 and the appropriate second-order condition,* $\sqrt{k_n}(\hat{\gamma}_n - \gamma)$ *converges weakly to the centered Gaussian distribution with variance* γ^2.

Chapter 3

Numerical experiments

3.1 Estimators for the tail index α

The tail index α dominates the asymptotic behaviour of a distribution and indicates the heaviness of the tail. There are lots of estimators for α, we particularly focus on the POT estimator, Hill estimator and Pickands estimator.

3.1.1 Peaks Over Threshold (POT) Maximum Likelihood (ML) estimator

The POT method analyzes exceedances over a given high thresholds.
Suppose $\{Y_i, -\infty < i < \infty\}$ is an i.i.d. sequence and u is a threshold. Define the exceedance times $\{\tau_r, r \geq 1\}$ by

$$\tau_1 = \inf\{j \geq 1 : Y_j > u\},$$

$$\tau_2 = \inf\{j \geq \tau_1 : Y_j > u\},$$

$$\vdots$$

$$\tau_r = \inf\{j \geq \tau_{r-1} : Y_j > u\},$$

The sequence $\{Y_{\tau_r}, r \geq 1\}$ are called the exceedances, where r is the number of exceedences over u.

If $\{Y_i, -\infty < i < \infty\}$ is i.i.d. with common distribution F, then $\{Y_{\tau_r}, r \geq 0\}$ is also i.i.d., and for $\overline{F}^{[u]}(x) = P(Y_{\tau_r} > x) = P(Y > x|Y > u)$,

$$\overline{F}^{[u]}(x) = \begin{cases} \frac{P(Y>x,Y>u)}{P(Y>u)} = \frac{P(Y>x)}{P(Y>u)} = \frac{\overline{F}(x)}{\overline{F}(u)} & \text{if } x > u, \\ 1 & \text{if } x < u. \end{cases}$$

Assume that $P(Y > x) = \overline{F}_Y(x) = x^{-\alpha}L(x)$, then for $x > u$,

$$\frac{P(Y > xu)}{P(Y > u)} = \frac{\overline{F}_Y(xu)}{\overline{F}_Y(u)} = \frac{(xu)^{-\alpha}L(ux)}{u^{-\alpha}L(u)} \to x^{-\alpha} \text{ as } u \to \infty.$$

Assume that Y is Pareto, then we have,

$$P(Y > x) = x^{-\alpha}, x > 1,$$

$$\overline{F}^{[u]}(x) = \begin{cases} \frac{x^{-\alpha}}{u^{-\alpha}} = (\frac{x}{u})^{-\alpha} & \text{if } x > u, \\ 1 & \text{if } x < u. \end{cases}$$

The maximum-likelihood estimator of $\frac{1}{\alpha}$ can be found as follows:

Consider likelihood function for $Y_{\tau_1}, \ldots, Y_{\tau_r}$:

$$L = \prod_{i=1}^{r} \alpha \left(\frac{Y_{\tau_i}}{u} \right)^{-(\alpha+1)}$$

Taking logarithms yields,

$$\log L = r \log \alpha - (\alpha + 1) \sum_{i=1}^{r} \log \left(\frac{Y_{\tau_i}}{u} \right)$$

$$\widehat{\alpha}^{-1} = \hat{\gamma}_{r,n}^{(POT)} = \frac{1}{r} \sum_{i=1}^{r} \log \left(\frac{Y_{\tau_i}}{u} \right) \tag{3.1.1}$$

3.1.2 Hill estimator

Consider equation 3.1.1, if we choose $u = Y_{(k_n+1)}$, the $(k+1)$th largest observation and $Y_{(1)} > u, Y_{(2)} > u, \cdots, Y_{(k_n)} > u$, then we have the **Hill Estimator** of $1/\alpha$ based on $k_n + 1$ upper-order statistics:

$$\hat{\gamma}_{k_n,n}^{(Hill)} = \frac{1}{k_n}\sum_{i=1}^{k_n} \log \frac{Y_{(i)}}{Y_{(k_n+1)}} \tag{3.1.2}$$

for $k = 1, \ldots, n-1$.

There is a theoretical formula which gives an optimal choice of k_n. However, that formula is useless for practical purposes. In practice, one looks at "stability region" of the Hill plot and then one picks the biggest possible k_n to have the shortest confidence interval. For example, on Figure 3.3 I would choose $k_n = 600$.

3.1.3 Pickands estimator

Suppose $Z_i, i \geq 1$ is i.i.d. with common distribution F. The Pickands estimator of $\gamma = \frac{1}{\alpha}$ uses differences of quantiles and is based on using three upper-order statistics $Z_{(k)}, Z_{(2k)}, Z_{(4k)}$, from a sample of size n. The estimator is defined as (recall that $k = k_n$ depends on n)

$$\hat{\gamma}_{k,n}^{(Pickands)} = \left(\frac{1}{\log 2}\right) \log\left(\frac{Z_{(k)} - Z_{(2k)}}{Z_{(2k)} - Z_{(4k)}}\right). \tag{3.1.3}$$

Using the result in Resnick(2007), p.93,

$$\frac{Z_{([k/y])} - a_{n/k}}{a_{n/k}} \to \frac{y^\gamma - 1}{\gamma}, 0 \leq y < \infty,$$

where $a_n = n^{1/\alpha}$, we have

$$\begin{aligned}
\frac{Z_{(k)} - Z_{(2k)}}{Z_{(2k)} - Z_{(4k)}} &= \frac{\frac{(Z_{(k)} - a_{n/k})}{a_{n/k}} - \frac{(Z_{(2k)} - a_{n/k})}{a_{n/k}}}{\frac{(Z_{(2k)} - a_{n/k})}{a_{n/k}} - \frac{(Z_{(4k)} - a_{n/k})}{a_{n/k}}} \\
&\xrightarrow{P} \left(\frac{0 - \gamma^{-1}((\frac{1}{2})^\gamma - 1)}{\gamma^{-1}((\frac{1}{2})^\gamma - 1) - \gamma^{-1}((\frac{1}{4})^\gamma - 1)}\right)
\end{aligned}$$

$$= \quad 2^{\gamma}.$$

By the continuous mapping theorem and taking logarithms yields

$$\log\left(\frac{Z_{(k)} - Z_{(2k)}}{Z_{(2k)} - Z_{(4k)}}\right) \xrightarrow{P} \log 2^{\gamma}$$

and dividing by $\log 2$

$$\left(\frac{1}{\log 2}\right)\left(\log \frac{Z_{(k)} - Z_{(2k)}}{Z_{(2k)} - Z_{(4k)}}\right) \xrightarrow{P} \gamma$$

3.2 Simulations: POT method for different choices of threshold u

We simulate $Y_i = e^{0.2X_i} Z_i$, $i = 1, \ldots, n$, where Z_i are independent Pareto with parameter $\alpha = 2$ and X_i is ARFIMA with standard normal innovations and $d = 0.2$. For a sample of size $n = 1000$, we obtain $Y_1, \ldots, Y_{1000}$ and we estimate α using the maximum likelihood estimator based on exceedences over threshold u,

$$\widehat{\alpha}^{-1} = \frac{1}{r}\sum_{i=1}^{r} \log\left(\frac{Y_{\tau_i}}{u}\right)$$

where r = number of exceedences over u.

For the deterministic levels, we choose $u_1 = \max(Y_1, \ldots, Y_{1000})/2$ based on the first sample, and then keep this value for Monte Carlo simulation.

Also in simulation, we could let $u_1 = \sqrt{5}$, since we assumed $\alpha = 2$. Indeed, in a sample of size n, we expect $nP(Y > u_1)$ values which are bigger than u_1. We have

$$nP(Y > u_1) \approx n u_1^{-2} = k_n,$$

so that $u_1 = \sqrt{\frac{n}{k_n}}$. Consequently, for $n = 1000$ and $k_n = 200$, we obtain $u_1 = \sqrt{5}$.

We choose $u = Y_{(n-k_n,n)}$ for random levels. The following simulated results indicate that when $u = Y_{(n-k_n,n)}$, the estimator is better.

- Sample Mean of $1/\hat{\alpha}$ when $u = \max(Y_1, \ldots, Y_{1000})/2$: 0.4535045;
- Sample Mean of $1/\hat{\alpha}$ when $u = Y_{(n-k_n,n)}$: 0.5012772;
- Sample standard deviation of $1/\hat{\alpha}$ when $u = \max(Y_1, \ldots, Y_{1000})/2$: 0.1550521;
- Sample standard deviation of $1/\hat{\alpha}$ when $u = Y_{(n-k_n,n)}$: 0.03536714.

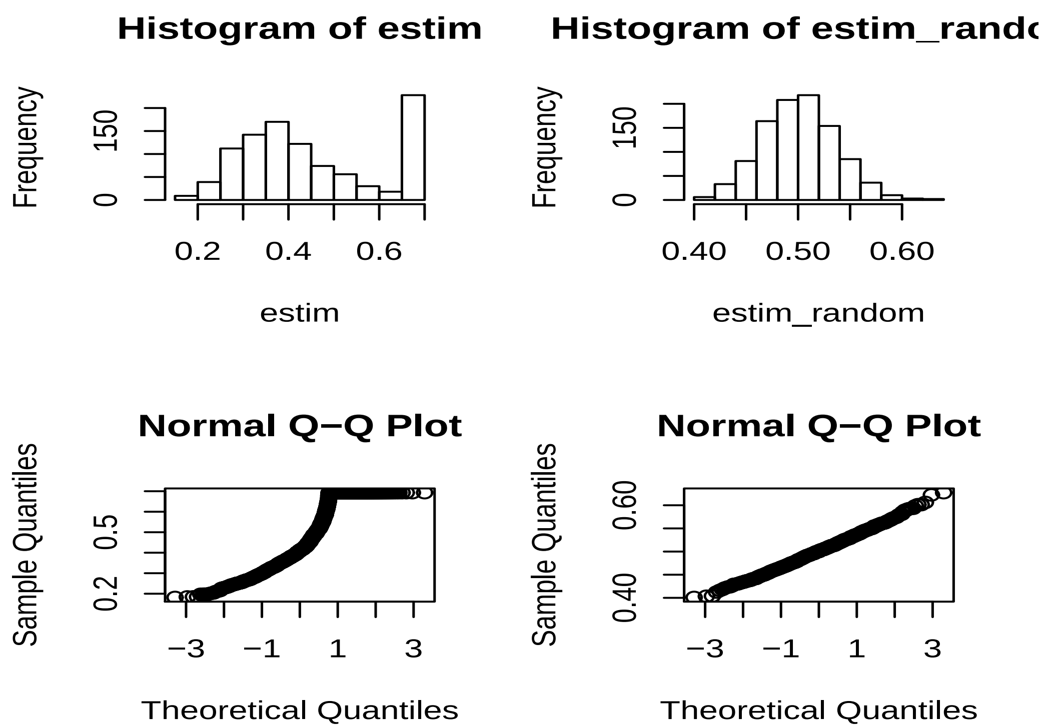

Figure 3.1: Statistics for POT simulation. The graphs illustrate empirical distribution of $\hat{\alpha}_1^{-1}, \ldots, \hat{\alpha}_M^{-1}$ based on $M = 1000$ Monte Carlo runs.

The Figure 3.1 illustrates the above results visually. The empirical distribution of $\hat{\alpha}_1^{-1}, \ldots, \hat{\alpha}_{1000}^{-1}$ at random levels is symmetric and has a mean equal to 0.5012772 which is very close to the true value 0.5. The QQ plot also shows that it is normally distributed. The Histogram and QQ plot of the estimator at deterministic levels are not appropriate.

3.3 Simulations: Hill and Pickands estimators

We simulate the pure SV model: $Y_i = \sigma(X_i)Z_i$, $i = 1, \ldots, n$, where Z_i are independent Pareto with parameter α and X_i is a long memory ARFIMA with standard normal innovations and the dependence parameter $d \in [0, 1/2)$. We assume that $\{X_i, i = 1, \ldots, n\}$ and $\{Z_i, i = 1, \ldots, n\}$ are mutually independent. In particular, we take $n = 1000, K = 200$ and $\sigma(X_i) = e^{\beta X_i}$, then our model becomes $Y_i = e^{\beta X_i} Z_i$, where $0 < \beta < 1$ is the scale parameter.

Table 3.1: Hill estimator simulation; standard deviation $\beta = 0.2$

$1/\alpha$ \ d	0	0.2	0.4	0.45
0.667	Mean=0.6631	Mean=0.6670	Mean=0.6717	Mean=0.6659
	Stdev=0.0664	Stdev=0.0682	Stdev=0.0648	Stdev=0.0648
0.5	Mean=0.5001	Mean=0.5010	Mean=0.4988	Mean=0.5010
	Stdev=0.0506	Stdev=0.0500	Stdev=0.0515	Stdev=0.0503
0.25	Mean=0.2513	Mean=0.2518	Mean=0.2511	Mean=0.2530
	Stdev=0.0249	Stdev=0.0251	Stdev=0.0251	Stdev=0.0246
0.167	Mean=0.1711	Mean=0.1718	Mean=0.1791	Mean=0.1833
	Stdev=0.0174	Stdev=0.0170	Stdev=0.0174	Stdev=0.0188
0.1	Mean=0.1208	Mean=0.1226	Mean=0.1379	Mean=0.1452
	Stdev=0.0114	Stdev=0.0111	Stdev=0.0140	Stdev=0.0165

Table 3.1 shows that dependence of X_i does not influence tail index estimation, unless α is big. Note however, that from a practical point of view, bigger values of α

Table 3.2: Hill estimator simulation; standard deviation $\beta = 1$

d / $1/\alpha$	0	0.2	0.4	0.45
0.667	Mean=0.7138	Mean=0.7242	Mean=0.7796	Mean=0.8042
	Stdev=0.0665	Stdev=0.0675	Stdev=0.0812	Stdev=0.0839
0.5	Mean=0.6045	Mean=0.6185	Mean=0.6908	Mean=0.7273
	Stdev=0.0563	Stdev=0.0597	Stdev=0.0701	Stdev=0.0814
0.25	Mean=0.4979	Mean=0.5169	Mean=0.6092	Mean=0.6535
	Stdev=0.0459	Stdev=0.0479	Stdev=0.0704	Stdev=0.0782
0.167	Mean=0.4853	Mean=0.4992	Mean=0.5949	Mean=0.6412
	Stdev=0.0421	Stdev=0.0459	Stdev=0.0653	Stdev=0.0815
0.1	Mean=0.4761	Mean=0.4956	Mean=0.5942	Mean=0.6401
	Stdev=0.0426	Stdev=0.0453	Stdev=0.0669	Stdev=0.0815

are not interesting.

Table 3.2 shows that if the variability of βX_i increases, then dependence starts to play a role.

Theoretically, there shouldn't be an influence of dependence on the quality of the estimator. The quality of the estimator depends only on α and β. Estimators are good if both α and β are small.

In the next two tables (Table 3.3, Table 3.4) we collect results about Pickand's estimator. Table 3.3 shows that the dependence d does not influence the tail index estimation as well when α is small ($\alpha \leq 2$).

Table 3.4 shows that if β increases, dependence starts to play a role as well.

We can say that the Hill estimator is better than the Pickands estimator after the comparison of numerical results.

Table 3.3: Pickands estimator simulation; standard deviation $\beta = 0.2$

$1/\alpha$ \ d	0	0.2	0.4	0.45
0.667	Mean=0.6550	Mean=0.6617	Mean=0.6718	Mean=0.6659
	Stdev=0.2526	Stdev=0.2542	Stdev=0.2522	Stdev=0.2630
0.5	Mean=0.5026	Mean=0.5085	Mean=0.4979	Mean=0.4958
	Stdev=0.2488	Stdev=0.2447	Stdev=0.2443	Stdev=0.2480
0.25	Mean=0.2133	Mean=0.1984	Mean=0.1581	Mean=0.1510
	Stdev=0.2326	Stdev=0.2373	Stdev=0.2410	Stdev=0.2260
0.167	Mean=0.0265	Mean=0.0235	Mean=0.0045	Mean=-0.0172
	Stdev=0.2317	Stdev=0.2382	Stdev=0.2331	Stdev=0.2318
0.1	Mean=-0.1066	Mean=-0.1215	Mean=-0.1160	Mean=-0.1115
	Stdev=0.2338	Stdev=0.2286	Stdev=0.2248	Stdev=0.2365

3.4 Hill plots

In practice, one of the graphical methods for estimating the tail index α is the Hill plot. What is usually done is to first construct a Hill plot, then to infer the value of α from a stable region in the graph. We draw the Hill plots for the pure SV model: $Y_i = e^{0.2X_i} Z_i$, $i = 1, \ldots, n$, where Z_i are independent Pareto with parameter $\alpha = 2$ and X_i is ARFIMA with standard normal innovations and $d = 0.2$.

Figure 3.2 is a Hill plot for $1,000$ iid observations from the Pareto distribution with $\alpha = 2$. Notice the right side of the graph clearly indicates the correct value of 2. Figure 3.3 generates Hill plots of the pure SV model observations incrementally when dependence varies. The Hill plots are more stable when $200 < K < 600$. Of cause in this example the optimal K could be 600 which minimizes the asymptotic mean square error of Hill estimator.

Table 3.4: Pickands estimator simulation; standard deviation $\beta = 1$

$1/\alpha$ \ d	0	0.2	0.4	0.45
0.667	Mean=0.6196	Mean=0.6182	Mean=0.6802	Mean=0.7229
	Stdev=0.2492	Stdev=0.2413	Stdev=0.2580	Stdev=0.2599
0.5	Mean=0.4760	Mean=0.4873	Mean=0.5686	Mean=0.6081
	Stdev=0.2447	Stdev=0.2507	Stdev=0.2496	Stdev=0.2616
0.25	Mean=0.3391	Mean=0.3688	Mean=0.4756	Mean=0.5349
	Stdev=0.2374	Stdev=0.2326	Stdev=0.2501	Stdev=0.2580
0.167	Mean=0.3163	Mean=0.3318	Mean=0.4572	Mean=0.5221
	Stdev=0.2341	Stdev=0.2407	Stdev=0.2508	Stdev=0.2791
0.1	Mean=0.3070	Mean=0.3301	Mean=0.4522	Mean=0.5315
	Stdev=0.2399	Stdev=0.2424	Stdev=0.2479	Stdev=0.2650

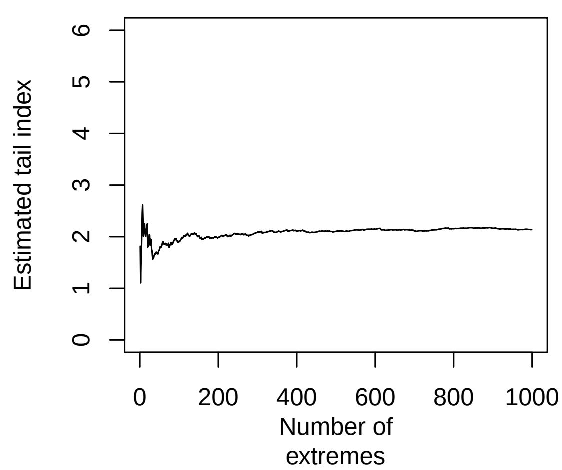

Figure 3.2: Hill plot for iid Pareto

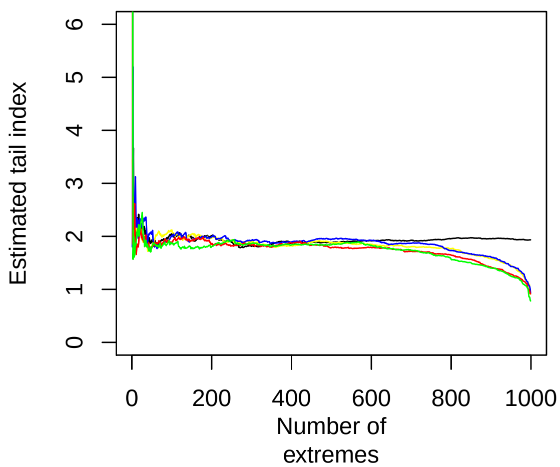

Figure 3.3: Hill plots with different dependence parameters (d=0(yellow),0.2(blue),0.4(red),0.45(green))

Appendix A - $D[0, \infty)$ space

(Resnick, 2007)

We start by considering $D[0,1]$, the space of right-continuous functions on $[0,1)$ that have finite left limits on $(0,1]$. Define time deformation

$$\Lambda = \{\lambda : [0,1] \mapsto [0,1] : \lambda(0) = 0, \lambda(1) = 1, \lambda(\cdot) \text{ is countinuous, strictly increasing}\}.$$

Let $e(t) \in \Lambda$ be the identity transformation and denote the uniform distance between x and y as

$$\|x - y\| := \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

The Skorohod metric $d(x, y)$ between two functions $x, y \in D[0,1]$ is

$$d(x, y) = \inf_{\lambda \in \Lambda} \|\lambda - e\| \vee \|x - y \circ \lambda\|.$$

Let $D[0, \infty)$ be the space of functions $x : [0, \infty) \to R$ that are right continuous with left limits. Denote the restriction of $x \in D[0, \infty)$ to the interval $[0, s]$ by $r_s x(\cdot)$, where

$$r_s x(t) = x(t), 0 \leq t \leq s.$$

Let d_s be the Skorohod metric on $D[0, \infty)$ and define d_∞, the Slorohod metric on $D[0, \infty)$, by

$$d_\infty(x, y) = \int_0^\infty e^{-s}(d_s(r_s x, r_s y) \wedge 1) ds.$$

Appendix B - R codes

R code for POT method

```
# Parameters set-up
library(fracdiff);  # Fracdiff needed to simulate long memory sequence;
n<-1000;    # Sample size;
alpha<-2;   # Tail index;
dep<-0.2;   # Long memory parameter;
stdev<-0.2;     # Standard deviation for long memory sequence;
y<-1:n;

# Simulation of Pareto random variables

rpareto<-function(n,alpha,lower.point=0.1)
    {
    if (alpha<=0)
    stop("alpha must >0.");
    x<-runif(n);
    pareto<-x^(-1/alpha);
    sign<-sample(c(-1,1),n,replace=TRUE);
    return(sign*lower.point*pareto);
```

```
    }


M<-1000;     # Number of Monte Carlo iterations;
estim=1:M; estim_fixed_level=1:M; estim_random=1:M;

# Monte Carlo simulation


for(j in 1:M)
    {
    ParetoSample<-rpareto(n,alpha);
    x<-fracdiff.sim(n,d=dep);   # Long memory sequence;
    dep.part<-exp(x$series*stdev);
    y<-ParetoSample*dep.part;   # SV model;

    u1<-max(y)/2;   # Deterministic level;
    r<-length(y[y>u1]);
    est<-(1/r)*sum(log(y[y>u1]/u1));
    estim[j]=est;

    k<-200; # Number of extremes used in Hill estimator;
    SortedY<-sort(y);
    u<-SortedY[n-k];
    est_random<-(1/k)*sum(log(y[y>u]/u));
    estim_random[j]=est_random;
    }

# Statistics for POT estimators;
```

```
mean(estim); mean(estim_random); sd(estim); sd(estim_random);
par(mfrow=c(2,2))
hist(estim); hist(estim_random); qqnorm(estim);
qqnorm(estim_random);
```

R code for Hill plots

```
Hillalpha<-function(x)
{
    ordered<-rev(sort(abs(x)));
    ordered<-ordered[ordered[]>0];
    n<-length(ordered);
    loggs<-log(ordered);
    hill<-cumsum(loggs[1:(n-1)])/(1:(n-1))-loggs[2:n];
    return(hill);
}

rpareto<-function(n,alpha,lower.point=0.1)
{
    if (alpha<=0);
    stop("alpha must >0.");
    x<-runif(n);
    pareto<-x^(-1/alpha);
    return(lower.point*pareto);
}

library(fracdiff);
```

```
HillPlot<-function(n=1000,alpha=2,stdev=0.2,Dep= c(0,0.2,0.4,0.45))
{
    colors=c("blue","red","green");
    alpha_vec=c(rep(1/alpha,n-1));
    Index=1:(n-1);
    temp2=1;
    delta=sample(c(-1,1),n,replace=TRUE);
    ParetoSample<-delta*rpareto(n,alpha);
    hill.plot.Pareto<-Hillalpha(ParetoSample);
    hill.plot<-matrix(0,length(Dep),n-1);
    temp=1;
    for (dep in Dep)
    {
        X<-fracdiff.sim(n,d=dep);
        dep.part<-exp(X$series*stdev);
        Y<-ParetoSample*dep.part;
        hill.plot[temp,]<-Hillalpha(Y);
        temp=temp+1;
     }
    readline(prompt = "Press <Enter> to continue...");
    print("Hill estimator  Pareto iid");
    par(mfrow=c(2,2));
    for(i in 1:4)
    {
        temp=1;
        plot(Index,1/hill.plot.Pareto,type="l",xlab="Number of extremes",
        ylab="Estimated tail index",ylim=c(0,3*alpha));
        readline(prompt = "Press <Enter> to continue...");
```

```
        print("Hill estimator - SV model. Dependence parameter is");
        print(Dep[i])
        for(dep in Dep[1:i])
        {
            readline(prompt = "Press <Enter> to continue...");
            color=sample(colors,1,replace=TRUE);
            points(Index,1/hill.plot[temp,],type="l",xlab="Number of extremes",
            ylab="Estimated tail index",col=color);
            temp=temp+1;
        }
    }
}
```

Bibliography

[1] Arcones, Miguel A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. Ann. Probab. **22**, no. 4, 2242-2274.

[2] Beran, Jan (1994). *Statistics for Long Memory Processes*. Chapman-Hall/CRC.

[3] Billingsley, Patrick (1968). *Convergence of probability measures*. John Wiley & Sons.

[4] Brockwell, Peter J., Davis, Richard A (2002). Introduction to time series and forecasting. Second edition. Springer Texts in Statistics. Springer-Verlag, New York.

[5] Hall, P.; Heyde, C. C. (1980). Martingale limit theory and its application. Probability and Mathematical Statistics. Academic Press, Inc.

[6] Kulik, Rafał and Soulier, Philippe (2011). The tail empirical process for long memory stochastic volatility sequences. *Stochastic Process. Appl.* **121**, no. 1, 109–134.

[7] McElroy, Tucker and Jach, Agnieszka (2011). Tail Index Estimation in the Presence of Long-Memory Dynamics. *Preprint*

[8] Mikosch, Thomas and Rezapour, Mohsen (2011). Stochastic volatility models with possible extremal clustering. *Preprint*

[9] Palma, Wifredo (2007). Long-memory time series. Theory and methods. Wiley Series in Probability and Statistics.

[10] Resnick, Sidney I. (2007). *Heavy-tail phenomena. Probabilistic and statistical modeling.* Springer Series in Operations Research and Financial Engineering. Springer.

[11] Rootzén, Holger (2009). Weak convergence of the tail empirical process for dependent sequences. Stochastic Process. Appl. 119, no. 2, 468–490.