Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa

Relazione finale

Fractional Brownian Motion and Hurst Parameter Estimation in Stochastic Volatility Models

Relatrice: **Prof.ssa Giorgia Callegaro**
Dipartimento di Matematica

Candidato: **Sofia Smorgoni**
Matricola: **1068143**

Anno Accademico 2016/2017

Contents

Introduction

When Bachelier presented his doctoral thesis in Mathematics at the University of Paris in 1900, he brought to life the idea that financial assets could be described through collections of random variables. Since that time, stochastic processes have played a leading role in the modelling of prices and returns on assets, bonds and stocks, derivatives and cash. Many models were built upon the idea that prices evolve continuously in time following a Brownian motion, a probabilistic process initially widely used to describe physical phenomena and later adopted to deal with financial economics problems. Black and Scholes model and stochastic volatility models for option pricing were those of greater impact.
Alternative approaches were later introduced, with the aim on filling the gaps of the Brownian motions modelling. In particular, it was observed empirically that correlations between observations that were far apart in time decayed to zero at a slower rate than one would expect from independent data or data following classic models. As a result, self-similar and stationary processes were introduced as valid building blocks for new models and long-range dependence became a rapidly developing subject. Particular importance was attributed to the fractional Brownian motion process, whose characteristics perfectly fitted the new long-memory approach. The peculiarity of the model, in comparison to Brownian motion, consisted in its dependence on the Hurst parameter, whose value determined the short-term and long-term behaviour of the process.

The idea behind this project is then to present the most relevant stochastic processes and to study the application of these processes to financial modelling under a statistical point of view. The central focus is on fractional Brownian motion and the Hurst parameter estimation, whose significance has increased in the last years.

In particular the first chapter introduces the concept of stochastic process and describes white noises and random walks, as well as Martingale and Markov properties, concluding with a comprehensive introduction to Brownian motion. Then, the second chapter focuses on fractional Brownian motion, on its long-range dependence and self-similarity properties. The third chapter presents the notion of financial modelling and the most important result obtained in the last five decades. More precisely, Black

and Scholes model, stochastic volatility models, fractional stochastic volatility models and rough fractional stochastic volatility models are introduced. The fourth chapter provides an introduction on both heuristic and maximum likelihood estimation methods for the Hurst parameter H, together with an example of estimation on real data. Finally, in Appendix A some important definitions can be found and in Appendix B the R code used in the project is reported.

Chapter 1

Stochastic Processes

1.1 Definitions and characteristics

Definition 1.1. *A **time series** is a sequence of numerical data points, each associated to a specific instance of time or time interval.*

Most commonly, a time series is a sequence taken at successive equally spaced points in time, thus it is a sequence of discrete-time data. Time series are usually considered as finite realizations of stochastic processes, since they represent only a limited sample chosen among the infinite possible realizations of the processes.

Definition 1.2. *Let $(\Omega, \mathfrak{F}, P)$ be a probability space where Ω is the sample space, $\mathfrak{F}$ is a $\sigma-$algebra and P is a probability measure. Given a parametric space $\mathcal{T}$, a **stochastic process** is a finite function of $\omega \in \Omega$ and $t \in \mathcal{T}$ such that $X_t : \omega \mapsto \mathbb{R}$ is an $(\mathfrak{F}_t)$-measurable function for every $t \in \mathcal{T}$, where $(\mathfrak{F}_t)_{t\in\mathcal{T}}$ is the filtration[1] on the given probability space.*

The set $\mathcal{T}$, used to index the random variables, is called **index set**. A stochastic process is called **discrete-time stochastic process** if $\mathcal{T}$ has a finite or countable number of elements or **continuous-time stochastic process** if $\mathcal{T}$ is some interval of the real line $\mathbb{R}$.
A stochastic process is usually indicated as $(X_t)_{t\in\mathcal{T}}$, but it can also be written as $\{X(t, \omega) : t \in \mathcal{T}\}$ to highlight that it is a function of two variables: $t \in \mathcal{T}$ that usually represents time, and $\omega \in \Omega$ that indicates a possible state of the world. It is important to notice that:

- for every given $\omega = \omega_0$, $X_\cdot(\omega_0)$ is a function of t called **path** or **trajectory**;
- for every given $t = t_0$, $X_{t_0}(\cdot)$ is a measurable function of $\omega \in \Omega$, hence a random variable.

[1] For an overview on filtrations see Appendix B.

Definition 1.3. *An **increment** of a stochastic process is the difference between two random variables X_t and X_s of the same stochastic process, with $t > s$.*

For a stochastic process with an index set that can be interpreted as time, an increment indicates how much the stochastic process changes over a certain time period. For example, let t_1 and t_2 be two elements of the index set $\mathcal{T}$ such that $t_2 > t_1$. Then $X_{t_2} - X_{t_1}$ is a random variable called increment.

To satisfactorily describe a stochastic process it would be necessary to specify the joint probability distribution of $(X_{t_1}, ..., X_{t_n})$ for every set of instances $t_1, ..., t_n$ and for every n. This procedure is usually complicated and it is preferable to define the process through the first moments of the random variable X_t.
Let $(X_t)_{t\in\mathcal{T}}$ be a stochastic process. Its mean, variance and autocovariance can be defined as functions of the random variable X_t and time t:

- mean function $\mu_t = \mathbb{E}(X_t)$;
- variance function $\sigma_t^2 = Var(X_t) = \mathbb{E}(X_t - \mu_t)^2$;
- autocovariance function $\gamma_{t,s} = Cov(X_t, X_s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$.

Furthermore, to facilitate the interpretation of the autocovariance function, the following definition can be considered:

Definition 1.4. *The **autocorrelation function** - ACF - is the scale-independent version of the autocovariance function and it is defined as*

$$\rho_{t,s} = Corr(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} = \frac{\gamma_{t,s}}{\sigma_t \sigma_s}$$

where γ is the autocovariance function and σ the variance function of the process.

It is now possible to introduce the concept of stationarity. A stochastic process is stationary if its mean and variance do not show systematic swings and if its dynamics does not show periodic variations, or in other words if it evolves smoothly rather than by abrupt changes.

Definition 1.5. *The process $(X_t)_{t\in\mathcal{T}}$ is **strictly stationary** if*

$$(Y_{t_1}, ..., Y_{t_n}) \overset{d}{=} (Y_{t_1+\tau}, ..., Y_{t_n+\tau}) \quad \forall n \in \mathbb{N}, \forall \tau \in \mathbb{R}^+$$

where $\overset{d}{=}$ denotes equality in distribution, i.e. the two processes are governed by the same probability measure and their joint probability density functions are equivalent.

These conditions impose constraints on the whole distribution of the process and therefore on the moments of every order. An alternative to the notion of strict stationary process is weak or second order stationarity, which only limits the first two moments of the distribution.

Definition 1.6. *The process* $(X_t)_{t\in\mathcal{T}}$ *is* ***weakly stationary*** *if*

$$\begin{aligned}\mathbb{E}(X_t) =&\mu \quad \forall t \in \mathcal{T}\\ Cov(X_t, X_{t+k}) = \gamma_{t,t+k} =&\gamma_k \quad \forall t \in \mathcal{T}, k = 0, \pm 1, ..\end{aligned}$$

Definition (1.6) implies that the mean of the process is constant and, specifically, time-invariant, and that the autocovariance function depends only on k, called lag.
Since the correlation between two variables X_t and X_{t+k} is often due to the correlation these variables have with $X_{t+1}, .., X_{t+k-1}$, it is useful to consider the partial autocorrelation function as a valid alternative for measuring connection between variables.

Definition 1.7. *The* ***partial autocorrelation function*** *- PACF - measures the correlation between* Y_t *and* Y_{t-k} *after their linear dependence on the intervening random variables* $Y_{t-1}, ..., Y_{t-h+1}$ *has been removed. More precisely,*

$$P_k = Corr(X_t, X_{t+k}|X_{t+1}, .., X_{t+k-1}).$$

As described in Wei (2006), the PACF can be computed considering a regression model in which the dependent variable X_{t+k} is regressed on the k lagged variables $X_{t+k-1}, X_{t+k-2}, .., X_t$. Through recursive substitutions, one can find a system of k linear equations which can be solved applying the Cramer's rule, to obtain

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

where $|\cdot|$ represents the determinant of a matrix.
The notion of weak stationary is usually considered sufficient to conduct an analysis on many types of processes. In general, the following property holds, as explained and proved in Di Fonzo et al. (2005):

Proposition 1.1. *It can be shown that if* $(X_t)_{t\in\mathcal{T}}$ *is a strictly stationary process, then it is also weakly stationary if and only if* $Var(X_t) < \infty$*. Furthermore, if* $(X_t)_{t\in\mathcal{T}}$ *is a Gaussian process, then the notions of strict and weak stationary are equivalent.*

Given a sample of an unknown process, its characteristics have to be estimated based on the observed time series $(y_t)_{t=1}^n$. An estimator for a characteristic T is a function of the sample and will be denoted by $\hat{T}$. If the process is stationary and ergodic, i.e. its statistical properties can be deduced from a single and sufficiently

long sample of the process, then one can estimate its moments consistently, as follows. A consistent and unbiased estimator for the mean μ of the process is

$$\widehat{\mu} = \frac{1}{n}\sum_{t=1}^{n} y_t$$

and a consistent estimate for the autocovariance of the process can be

$$\widehat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \hat{\mu})(y_{t+k} - \hat{\mu}) \tag{1.1}$$

It can be proved that

$$\mathbb{E}(\widehat{\gamma}_k) = \gamma_k - \frac{k}{n}\gamma_k - \frac{n-k}{n}Var(\widehat{\mu})$$

hence $\widehat{\gamma}_k$ is biased. In order that $\widehat{\gamma}_k$ can be considered a valid estimator, it is necessary for n to be enough big and for k to be much smaller that n. From (1.1), the estimator for σ^2 can be obtained by imposing $k = 0$:

$$\widehat{\sigma}^2 = \widehat{\gamma}_0 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \widehat{\mu})^2.$$

Consequently an estimate for the autocorrelation function can be formulated as

$$\widehat{\rho}_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0},$$

which is biased but consistent. In particular, $Var(\widehat{\rho}_k)$ can be approximated with

$$Var(\widehat{\rho}_k) \simeq \frac{1}{n}\sum_{i=-\infty}^{\infty}\left(\rho_i^2 + \rho_{i-k}\rho_{i+k} + 2\rho_i^2\rho_k^2 - 4\rho_i\rho_k\rho_{i-k}\right).$$

Moreover, if $\rho_k \simeq 0$ for $k \geq q$, the Bartlett's approximation is valid

$$Var(\widehat{\rho}_k) \simeq \frac{1}{n}\sum_{i=-q}^{q}\rho_i^2 = \frac{1}{n}\left(1 + 2\sum_{i=1}^{q}\right).$$

More details and all the proves can be found in Wei (2006).

1.2 Fundamental stochastic processes

There are numerous types of stochastic processes, each one with different characteristics and applied in different fields. Two of the most important ones, white noise and random walk, will be discussed in this section.

1.2.1 White noise

One of the most important continuous stochastic processes is white noise, as it can be considered a building block for many other stationary processes.

Definition 1.8. *A process* $(\varepsilon_t)_{t\in\mathcal{T}}$ *is a* ***white noise*** *if*

$$\begin{aligned} \mathbb{E}(\varepsilon_t) &= 0 \quad \forall t \in \mathbb{R}^+ \\ Var(\varepsilon_t) &= \sigma^2 < \infty \quad \forall t \in \mathbb{R}^+ \\ Cov(\varepsilon_t, \varepsilon_s) &= 0 \quad \forall t \neq s \end{aligned}$$

Hence a white noise process consists of a sequence of serially uncorrelated random variables with zero mean and finite variance, and is often indicated as $\varepsilon_t \sim WN(0, \sigma^2)$. The autocorrelation function can be easily calculated and is clearly equal to

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, .. \end{cases}$$

as shown in Figure 1.2.

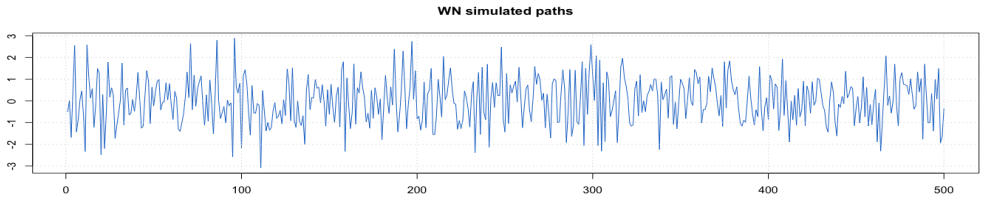

Figure 1.1: Simulated White Noise process.

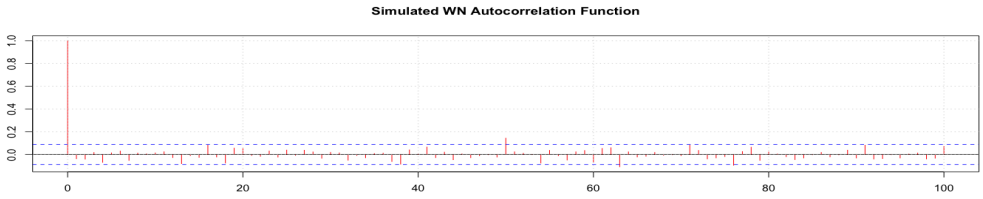

Figure 1.2: ACF for simulated WN process.

Remark 1.1. Technically, a process with a finite variance whose observations are independent and identically distributed is a white noise process but a white noise process is not necessarily generated from an i.i.d. random variable, since the ε_t are not necessarily identically distributed or independent.

If the random variables ε_t are normally distributed, the process is called Gaussian white noise.

1.2.2 Random walk

A random walk is a mathematical object that describes a path consisting of a succession of random steps on some mathematical space. For example, the path traced by a molecule as it travels in a liquid or a gas, the price of a stock and the financial status of a gambler can all be approximated by random walk models, even though they may not be truly random in reality.
Various types of random walks are of interest, which can differ in several ways. The term itself most often refers to a special category of Markov chains or Markov processes, but many time-dependent processes are referred to as random walks, with a modifier indicating their specific properties.

Definition 1.9. *A **random walk** is a process $(X_n)_{n\in\mathbb{N}}$ where the current value of the variable is defined as the sum of the past value and an error term represented by a white noise. Algebraically a random walk is described as follows:*

$$\begin{aligned} X_n &= X_{n-1} + \varepsilon_n \\ X_0 &= \mu \\ \varepsilon_n &\sim WN(0, \sigma_\varepsilon^2). \end{aligned}$$

Through recursive substitutions, X_n can be represented as a sum of white noises at different times

$$X_n = \mu + \sum_{i=1}^{n} \varepsilon_i.$$

Furthermore, it can be easily shown that the mean of a random walk process is constant and equal to

$$\begin{aligned} \mathbb{E}(X_n) &= \mathbb{E}\left(\mu + \sum_{i=1}^{n} \varepsilon_i\right) \\ &= \mu + \sum_{i=1}^{n} \mathbb{E}(\varepsilon_i) = \mu, \end{aligned}$$

but its variance is not, as described in the following passages

$$\begin{aligned} Var(X_n) &= Var\left(\mu + \sum_{i=1}^{n} \varepsilon_i\right) \\ &= \sum_{i=1}^{n} Var(\varepsilon_i) = t\sigma_\varepsilon^2 \end{aligned}$$

Therefore a random walk process is non stationary, since its variance increases with t. The process can be made stationary through an appropriate transformation, called first differentiation, which consists in computing the difference between every observation and the one immediately preceding.

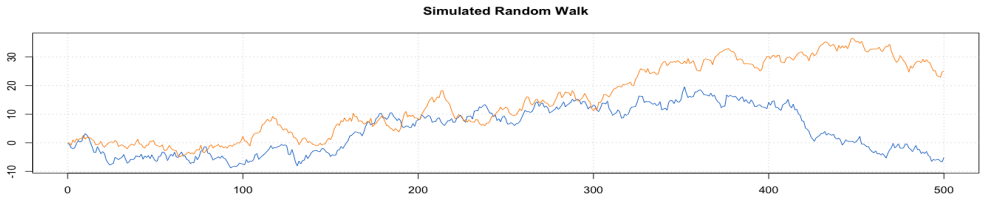

Figure 1.3: Simulated Random Walk processes.

Assuming $\mu = 0$, the autocorrelation function of a random walk can be computed as

$$\begin{aligned}\rho_k &= \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)Var(X_{t+k})}} \\ &= \frac{\mathbb{E}((\varepsilon_1 + .. + \varepsilon_t)(\varepsilon_{k+1}, .., \varepsilon_{t+k}))}{\sqrt{t\sigma_\varepsilon^2 \cdot (t+k)\sigma_\varepsilon^2}} \\ &= \frac{t\sigma_\varepsilon^2}{\sigma_\varepsilon^2\sqrt{t(t+k)}} \\ &= \sqrt{\frac{t}{t+k}}.\end{aligned}$$

Hence it can be inferred that

$$\rho_k = \sqrt{\frac{t}{t+k}} \underset{t\to\infty}{\longrightarrow} 1$$

which indicates that the process has infinite memory, as shown in Figure 1.4.

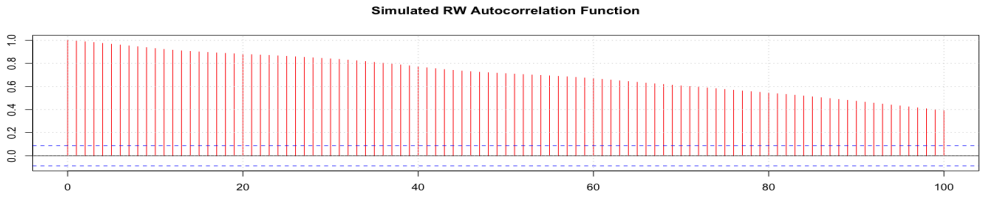

Figure 1.4: ACF for simulated RW process.

1.3 Martingale

A martingale is a model of fair game where the knowledge of past events never helps predicting the mean of the future winnings and only the present event matters.
In particular, a martingale is a stochastic process for which, at a particular time in the realized sequence, the expectation of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values. More precisely, the following definition is given

Definition 1.10. *A random process $(X_t)_{t\in\mathcal{T}}$ is a **martingale** with respect to the filtration $(\mathfrak{F}_t)_{t\in\mathcal{T}}$ if*

- $(X_t)_{t\in\mathcal{T}}$ *is adapted to* $(\mathfrak{F}_t)_{t\in\mathcal{T}}$*;*
- $E(|X_t|) < \infty \quad \forall t$*;*
- $E(X_s|\mathfrak{F}_t) = X_t \quad \forall s \geq t$.

*Furthermore a process satisfying the inequality $E(X_s|\mathfrak{F}_t) \leq X_t \ \forall s \geq t$, is called a **supermartingale**, and a process satisfying $E(X_s|\mathfrak{F}_t) \geq X_t \ \forall s \geq t$ is called a **submartingale**.*

Note that the martingale property is always given with respect to a filtration. The first condition says that we can observe the value X_t at time t, and the second condition is just a technical condition. The really important condition is the third one, which says that the expectation of a future value of X, given the information available today, equals today's observed value of X, or, in other words, that a martingale has no systematic drift.

Moreover, in probability theory, a **martingale difference sequence** - MDS - is a martingale for which the expectation, with respect to the past, is zero. By construction, this implies that if Y_t is a martingale, then $X_t = Y_t - Y_{t-1}$ is an MDS. In particular, $(X_t)_{t\in\mathcal{T}}$, is a conditionally homoskedastic martingale difference sequence if $\mathbb{E}(X_t|\mathfrak{F}_{t-1}) = 0$ and $Var(X_t|\mathfrak{F}_{t-1}) = \sigma^2$. It is a conditionally heteroskedastic martingale difference sequence if $\mathbb{E}(X_t|\mathfrak{F}_{t-1}) = 0$ and $Var(X_t|\mathfrak{F}_{t-1}) = \sigma_t^2$.
MDSs are useful components in probability theory, as they imply less restrictive conditions on the memory of sequences than independence.

1.4 Markov processes

In probability theory and related fields, a Markov process, named after the Russian mathematician Andrey Markov, is a stochastic process characterized by the 'memorylessness' propriety. Loosely speaking, a process satisfies this property if one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history, hence independently from such history.

Definition 1.11. *A stochastic process is called **Markovian** if it has the Markov property.*

Proposition 1.2. *Given a probability space $(\Omega, \mathfrak{F}, P)$ where $\mathfrak{F}$ is a σ-algebra endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$, and given a bounded Borel function f [2], a stochastic process*

[2] A Borel function is a function that is Borel measurable, i.e. the inverse image of any open set in its codomain is a Borel set of its domain.

has the ***Markov property*** *if*

$$P(f(X_t) \in A|\mathfrak{F}_s) = P(f(X_t) \in A|X_s) \quad \forall t > s \geq 0 \quad \forall A \in \mathscr{B}(\mathbb{R})$$

Remark 1.2. Note that Proposition (1.2) shows how the probability distribution of X_s given the information available at time $t < s$, which is embodied in the σ-algebra $\mathfrak{F}_t$, is equal to the probability distribution of X_s given the value taken by X_t.

In other words, the behaviour of the process in the future is stochastically independent of its behaviour in the past, given the current state of the process.

1.5 Brownian motion

Brownian movement is the name given to the irregular movement of pollen, suspended in water, observed by the botanist Robert Brown. In 1828, while looking through a microscope at particles trapped in pollen grains in water, he noticed that the particles moved through the water, but he was not able to determine the mechanisms that caused this motion. Albert Einstein published a paper in 1905 that explained in precise detail how the motion that Brown had observed was a result of the pollen being moved by individual water molecules. This random movement results in a dispersal or diffusion of the pollen in the water.
The range of application of Brownian motion as defined here goes far beyond the study of microscopic particles in suspension and includes modelling of stock prices, of thermal noise in electrical circuits, of certain limiting behaviour in queueing and inventory systems, and of random perturbations in a variety of other physical, biological, economic and management systems.

Definition 1.12. *A real valued stochastic process $(B_t)_{t\geq 0}$ on a probability space $(\Omega, \mathfrak{F}, P)$ is called a* ***Brownian motion*** *or Wiener process, with starting point $x \in \mathbb{R}$, if the following holds:*

- *B(0)=x;*
- *the process has independent increments, i.e. for $0 \leq t_1 \leq t_2 \leq ... \leq t_n$ 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, i.e. $B_{t+h} - B_t \sim N(0, h)$;*
- *almost surely, the function $t \mapsto B_t$ is continuous i.e. B_t has continuous trajectories.*

Moreover $(B_t)_{t\geq 0}$ is a ***standard Brownian motion*** *if x=0.*

Remark 1.3. The second condition is a Markov property saying that, conditional on the present value B_t, any past information of the process B_j with $j < t$ is irrelevant to the future B_{t+h} with $h > 0$.

In other words, a Weiner process is a special stochastic process with zero drift and variance proportional to the length of the time interval. This means that the rate of change in expectation is zero and the rate of change in variance is 1.

Remark 1.4. An important property of the Brownian motion is that its paths are not differentiable almost surely. In other words, for a standard Brownian motion B_t it can be shown that dB_t/dt does not exist for all elements of Ω. As a result, the usual integration in calculus can not be used to handle integrals involving a standard Brownian motion. An alternative approach is the introduction of Itô's calculus, which will be presented e.g. in Theorem (1.1).

In practice, the mean and variance of a stochastic process can evolve over time in a more complicated manner. Hence, a further generalization of the stochastic process is needed. To this aim, we need to consider the **generalized Brownian motion** in which the expectation has a drift rate μ and the rate of variance change is σ^2.
Let $(X_t)_{t\geq 0}$ be such a process. Its local dynamics can be approximated by a stochastic differential equation - SDE - of the following type:

$$dX_t = \mu dt + \sigma dB_t$$

where B is a Brownian motion. In literature, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}$ are referred to as the **drift** and **volatility** parameters of the generalized Brownian motion $(X_t)_{t\geq 0}$ and the process itself is called a **diffusion**.
The drift and volatility parameters are time invariant, but if one extends the model by allowing μ and σ to be functions of the stochastic process X, then an **Itô drift-diffusion process** is generated, such that

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t.$$

This process is known as a stochastic diffusion function, with $\mu(t, x) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ and $\sigma(t, x) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ being the drift and diffusion functions.
This process plays an important role in mathematical finance and it can be written as

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s, \tag{1.2}$$

or, equivalently, as

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \tag{1.3}$$

$$X_0 = a \tag{1.4}$$

where (1.3) represents a stochastic differential of X_t and (1.4) the initial condition

of the process. Note that Equation (1.2) has two parts: a deterministic integral and a stochastic integral. The first one can be solved via Riemann integration theory, whereas the second is an Itô integral. The following formula has to be considered when dealing with Itô integrals:

Theorem 1.1 (**Itô's formula**). *Assume that the process $(X_t)_{t\in\mathcal{T}}$ is a generalized Brownian motion that satisfies the following stochastic differential equation*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t,$$

where μ and σ are adapted processes, and let f be a $\mathcal{C}^{1,2}$-function. Define the process Z by $Z(t) = f(t, X_t)$. Then Z has a stochastic differential given by

$$df(t, X_t) = \left[\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial X_t^2}\right]dt + \sigma\frac{\partial f}{\partial X_t}dB_t$$

where B is a Brownian motion.

A complete proof is outside the scope of this text, but it can be found in Björk (2009). To briefly describe it, the formula can be derived by forming the Taylor series expansion of the function up to its second derivatives and retaining terms up to first order in the time increment and second order in the Brownian motion increment.
In general, Itô's lemma, or formula, is used to find the differential of a time-dependent function of a stochastic process. Its best known application is in the derivation of the Black-Scholes equation for option values.

1.5.1 Geometric Brownian Motion

A geometric Brownian motion - GBM -, also known as exponential Brownian motion, is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion with drift. It is often used in mathematical finance to model stock prices in the Black-Scholes model. Formally:

Definition 1.13. *Let $\sigma > 0$ and $\mu \in \mathbb{R}$. The **geometric Brownian motion** with drift parameter μ and volatility parameter σ is the solution to the stochastic differential equation*

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

where B is a standard Brownian motion.

When $X_0 > 0$, the application of Itô's lemma to $f(t, X_t) = \ln X_t = Y_t$ gives

$$df(t, X_t) = \left[\frac{\partial f}{\partial t} + \mu X_t\frac{\partial f}{\partial X_t} + \frac{1}{2}\sigma^2 X_t^2\frac{\partial^2 f}{\partial X_t^2}\right]dt + \sigma X_t\frac{\partial f}{\partial X}dB_t$$

Performing further calculations, one can obtain

$$\begin{aligned} dY_t &= \left[\frac{\partial Y_t}{\partial t} + \mu X_t \frac{\partial Y_t}{\partial X_t} + \frac{1}{2}\sigma^2 X_t^2 \frac{\partial^2 Y_t}{\partial X_t^2}\right] dt + \sigma X_t \frac{\partial Y_t}{\partial X} dB_t \\ &= \left[\mu X_t \frac{1}{X_t} - \frac{1}{2}\sigma^2 X_t^2 \frac{1}{X_t^2}\right] dt + \sigma X_t \frac{1}{X_t} dB_t \\ d\ln(X_t) &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t. \end{aligned}$$

The result shows that the logarithm of a GBM X follows a generalized Brownian motion with drift parameter $\left(\mu - \sigma^2/2\right)$ and scale parameter σ. The geometric Brownian motion itself is simply the exponential of the non-differentiated process Y_t

$$X_t = X_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma B_t\right\}$$

The process is clearly always positive, and this is one of the main reasons why it has been so widely adopted to describe financial markets.

Chapter 2

Fractional Brownian Motion

Most real-world problems encountered in financial economics consider Brownian motion as the source of randomness and uncertainty. It is for example used in the pricing of options in the popular Black-Scholes-Merton theory.
Despite its common application in numerous situations, empirical evidence has failed to prove Brownian motion as the source of uncertainty, for two main reasons. First, several studies have shown that asset return distributions observed in financial markets do not follow the Gaussian law, since they tend to have a positive excess kurtosis and heavy tails. Second, time series of return distribution exhibit long-range dependency. To overcome the criticism, several heavy tailed distributions, such as the stable and Laplace, have been evaluated as possible alternatives in the description of return distribution and fractional Brownian motion has been introduced to capture the long-range dependency of financial time series.

2.1 Definition

Fractional Brownian motion, introduced in 1968 by Mandelbrot and Van Ness, is an extension of Brownian motion obtained by adding one parameter, called Hurst parameter, which can take on a value between 0 and 1. The parameter takes is name after the hydrologist Harold Edwin Hurst, who first studied long-range dependence, documenting and mathematically describing the dependence properties of the water level of the river Nile.

Definition 2.1. *Let $H \in (0, 1]$. A **fractional Brownian motion** - fBm - with Hurst parameter H is a continuous Gaussian process $B^H = (B_t^H)_{t \geq 0}$ such that:*

- $B_0^H = 0$;
- $\mathbb{E}(B_t^H) = 0 \quad \forall t \in \mathbb{R}^+$;
- $\mathbb{E}(B_t^H B_s^H) = \frac{1}{2}(s^{2H} + t^{2H} - |t - s|^{2H}) \quad \forall s, t \in \mathbb{R}^+$.

Remark 2.1. In order to specify the distribution of a Gaussian process, it is sufficient to indicate its mean and covariance functions, therefore the distribution of B^H is uniquely determined by Definition (2.1).

It is important to notice that this definition does not guarantee the existence of fBm, as to show that the fBm exists one needs to check that the covariance function is non-negative defined. The constraint imposed on H parameter is yet strongly linked with the existence of the covariance function of the process, as showed and proved in Nourdin (2012):

Theorem 2.1. *Let $H > 0$ be a real parameter. Then, there exists a continuous Gaussian process $B^H = (B_t^H)_{t\geq 0}$ with covariance function given by*

$$\Gamma_H(s,t) = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H})$$

if and only if $H \leq 1$.

It can be shown that if $H = \frac{1}{2}$, then the fBm is just a standard Brownian motion. This property justifies the name "fractional Brownian motion": B^H is a generalization of Brownian motion obtained by allowing the Hurst parameter to differ from 1/2.

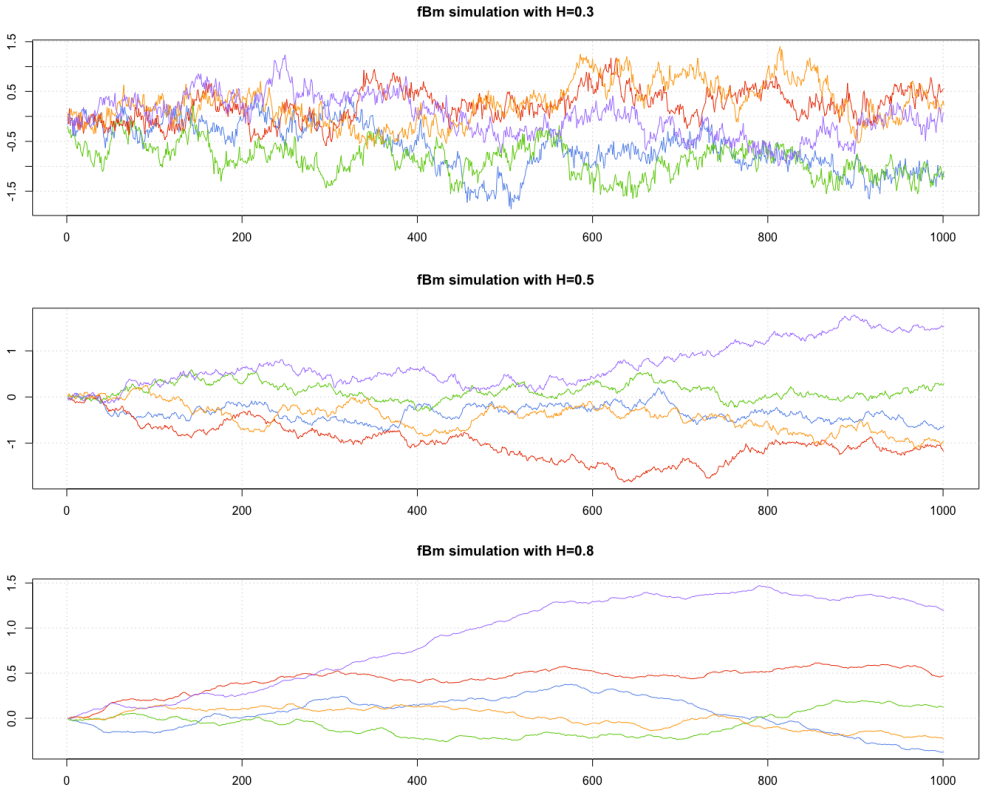

Figure 2.1: Simulated paths for fBm with different values of the Hurst parameter.

Figure 2.1 shows different simulated fBm paths we obtained through the use of specific R packages. In particular, one can see that, when H is close to 0, the process is characterized by a high number of fluctuations, whereas for values of the parameter close to 1, the process presents reduced fluctuations. When $H = 0.5$, the sample paths are trajectories of a Brownian motion.

There are three different integral representations for an fBm. In particular, assuming $H \in (0, 1)$ and $H \neq 1/2$:

- the so-called **time representation** for the process $B^H = (B_t^H)_{t\geq 0}$ is given by

$$B_t^H = \frac{1}{c_H}\left(\int_{-\infty}^{0}\left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\right)dB_u + \int_0^t (t-u)^{H-\frac{1}{2}}dB_u\right)$$

  where

$$c_H = \sqrt{\frac{1}{2H} + \int_0^{+\infty}\left((1+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}}\right)^2 du} < \infty$$

  and B_u is a standard Brownian motion;

- the **spectral representation**, also called harmonized representation, for the process $B^H = (B_t^H)_{t\geq 0}$ is given by

$$B_t^H = \frac{1}{d_H}\left(\int_{-\infty}^{0}\frac{1-\cos(ut)}{|u|^{H+\frac{1}{2}}}dB_u + \int_0^{+\infty}\frac{\sin(ut)}{|u|^{H+\frac{1}{2}}}dB_u\right)$$

  where

$$d_H = \sqrt{2\int_0^{+\infty}\frac{1-\cos(ut)}{u^{2H+1}}du} < \infty$$

  and B_u is a standard Brownian motion;

- the **Volterra process** is based on

$$B_t^H = \int_0^t K_H(t,s)dB_s$$

  where B_s is a standard Brownian motion and K_H is a square integrable kernel.

A detailed description of the representations, which goes beyond the scope of this text, can be found in Nourdin (2012).

Fractional Brownian motion has many really interesting properties, that have been investigated by researchers in many fields because of their uniqueness.
First of all, the process has stationary increments, i.e. the distribution of the increment $B_{t+h}^H - B_t^H$ is independent of t for any h. In particular, the increment process is known

as **fractional Gaussian noise** - FGN. Moreover, fBm is not a Markovian process neither a semi-martingale, so the usual Itô stochastic calculus does not apply. Yet, the two most important properties of fBm are long-range dependence and self-similarity, and they will be discussed in the next sections.

2.2 Long-range dependency

Long-range dependency - LRD - is a measure of decay of statistical dependency. From a financial economic prospective, the measure can be an autocorrelation function of lags of a time series.

Definition 2.2. *Let $(X_t)_{t\geq 0}$ be a stationary process for which the following holds: there exists a real number $\alpha \in (0,1)$ and a finite positive constant c_ρ such that*

$$\lim_{k\to\infty} \rho(k)/\left[c_\rho k^{-\alpha}\right] = 1. \tag{2.1}$$

*Then $(X_t)_{t\geq 0}$ is called a **stationary process with long memory** (or long range dependence, or strong dependence), or a stationary process with slowly decaying or long-range correlations.*

Remark 2.2. Definition (2.2) implies that

$$\rho(k) = Corr(X_t, X_{t+k}) \simeq c_\rho k^{-\alpha}$$

or, in other words, that the correlations $\rho(k)$ are asymptotically equal to a constant c_ρ times $k^{-\alpha}$ for some $\alpha \in (0,1)$.

The interpretation of Definition (2.2) is that the dependence between events that are far apart diminishes very slowly with increasing distance, more specifically slower than an exponential function.

Hurst parameter H can be obtained from α as $H = 1 - \alpha/2$. According to this identity, H itself is a measure of the extent of long-range dependence in a time series. As we know, H takes on values from 0 to 1, and in particular:

- a value of 0.5 indicates the absence of long-range dependence, as $H = 0.5$ implies $\alpha = 1$ and $\rho(k) \simeq 0$ for $k \to \infty$;
- a value close to 1 indicates an high persistence or long-range dependence and, in particular, the higher the value, the stronger the dependence;
- a value of H smaller than 0.5 corresponds to anti-persistency, which indicates strong negative correlation so that the process fluctuates violently.

Hence, in term of parameter H, long-range dependency occurs in fBm when $H > \frac{1}{2}$. In this case, the covariance of the increments decays very slowly, as it can be seen in Figure 2.2.

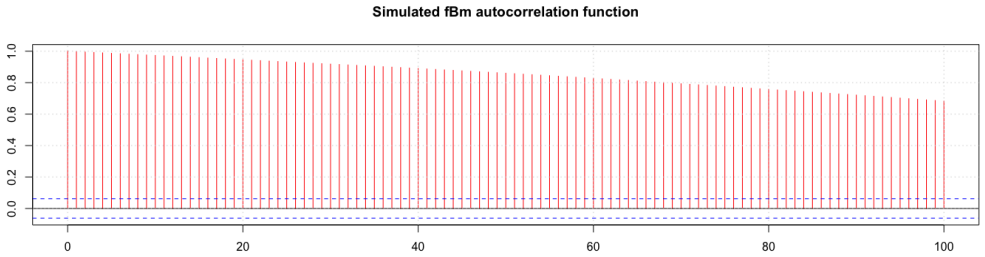

Figure 2.2: ACF for simulated fBm with $H = 0.8$.

It is important to notice that the definition of LRD by (2.1) is an asymptotic definition. It only tells something about the ultimate behaviour of the correlations as the lag tends to infinity. Moreover it only determines the rate of convergence but not the absolute size.

An alternative approach to LRD is the spectral or frequency domain approach. While the time domain approach to time series analysis originates from mathematical statistics, the spectral or frequency domain approach has its root in communication engineering. Whenever a signal fluctuates around a certain stable state we might use periodic functions to describe its behaviour. Spectral analysis aims at splitting the total variability of a stationary stochastic process into contributions related to oscillations with a certain frequency. In this context, the fundamental notion of spectral density needs to be introduced before an analysis can be conducted:

Definition 2.3. *The **spectral density** $f(\lambda)$ of a function with ACF $\rho(k)$ and variance σ^2 can be defined as*

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{ik\lambda}$$

where λ is the frequency, σ^2 is the variance of the observations and $i = \sqrt{-1}$. An alternative definition is given by

$$f(\lambda) = 2c_f(1 - \cos\lambda) \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-2H-1}$$

with $\lambda \in [-\pi, \pi]$ and $c_f = c_f(H, \sigma^2) = \sigma^2(2\pi)^{-1} \sin(\pi H)\Gamma(2H + 1)$.

Loosely speaking, the spectral density within a particular interval of frequencies can be viewed as the amount of the variance explained by those frequencies. In other words, it shows at which frequencies variations are strong and at which ones they are

weak. A deeper analysis is out of the scope of this text, but more information can be found in Priestley (1981).

If Equation (2.1) holds, then a similar definition can be shown to hold in the frequency domain. Hence, according to this Definition (2.3), long-range dependency can be also described as:

Definition 2.4. *Let $(X_t)_{t\geq 0}$ be a stationary process for which the following holds: there exist a real number $\beta \in (0,1)$ and a finite positive constant c_f such that*

$$\lim_{\lambda \to 0} f(\lambda) / \left[c_f |\lambda|^{-\beta} \right] = 1.$$

Then $(X_t)_{t\geq 0}$ is called a stationary process with long memory.

Remark 2.3. Definition (2.4) implies that

$$f(\lambda) \simeq c_f |\lambda|^{-\beta},$$

or, in other words, that the spectral density $f(\lambda)$ has a pole at zero that is equal to a constant c_f times $\lambda^{-\beta}$ for some $\beta \in (0,1)$.

Moreover, it can be shown that the parameter $\beta \in (0,1)$ is related to the Hurst parameter by $H = (1+\beta)/2$.

As the spectral density and the ACF are equivalent descriptions of the linear dynamic properties of a process, knowing the covariances is equivalent to knowing the spectral density. Hence, definitions (2.2) and (2.4) are equivalent in the following sense:

Theorem 2.2. *(1) Suppose (2.2) holds with $0 < \alpha = 2 - 2H < 1$. Then the spectral density f exists and*

$$\lim_{\lambda \to 0} f(\lambda) / \left[c_f(H) |\lambda|^{1-2H} \right] = 1,$$

where $\sigma^2 = Var(X_t)$ and $c_f = \sigma^2 \pi^{-1} c_\rho \Gamma(2H-1) \sin(\pi - \pi H)$.

(2) Suppose (2.4) holds with $0 < \beta = 2H - 1 < 1$. Then

$$\lim_{k \to \infty} \rho(k) / \left[c_\rho k^{2H-2} \right] = 1,$$

where $c_\rho = c_\gamma / \sigma^2$ and $c_\gamma = 2 c_f \Gamma(2-2H) \sin(\pi H - \frac{1}{2} H)$.

For a proof of Theorem (2.2) see Zygmund (1959).

It is finally possible to recall a fundamental theorem from Beran (1994), which allows to consider all the correlations together instead of separately. This property will be used in Chapter 3 to estimate the Hurst parameter.

Theorem 2.3. *Let $(X_t)_{t\in\mathcal{T}}$ be a stationary process with long-range dependence. Then*

$$\lim_{n\to\infty} Var\left(\sum_{i=1}^{n} X_i\right) / \left[c_\gamma n^{2H}\right] = \frac{1}{H(2H-1)}.$$

Remark 2.4. Theorem (2.3) implies that the variance of the sample mean $Var(\bar{X}_n)$ is asymptotically equal to a constant c_γ times $n^{-\alpha}$ for some $\alpha \in (0,1)$.

2.3 Self-Similarity

Since the LRD property is based on autocorrelation function for large lags, quantifying this property is sometimes a difficult task. An alternative is represented by the use of self-similar processes in empirical applications.
Self-similar processes were introduced by Kolmogorov in 1941 in a theoretical context, but the idea of self-similarity can be attributed to Leonardo da Vinci. In the context of stochastic processes, self-similarity is defined in terms of the distribution of the process.

Definition 2.5. *The stochastic process $(X_t)_{t\geq 0}$ is said to be a **self-similar process** if there is a value for H such that for all $c > 0$ one has*

$$X_{ct} \stackrel{d}{=} c^H X_t \tag{2.2}$$

i.e. if for any positive stretching factor c, the rescaled process with time scale ct, $c^{-H}X_{ct}$, is equal in distribution to the original process X_t.

This means that, for any sequence of time points $t_1, ...t_k$, and any positive constant c, $c^{-H}(X_{ct_1}, X_{ct_2}, ..., X_{ct_k})$ has the same distribution as $(X_{t_1}, X_{t_2}, ..., X_{t_k})$.

Remark 2.5. The parameter H is referred to as self-similarity exponent, scaling exponent, and Hurst exponent.

Suppose that $(X_t)_{t\geq 0}$ is a self-similar process with parameter H. By setting $c = 1/t$, Equation (2.2) can be written as follows:

$$t^H X_1 \stackrel{d}{=} X_t \quad t > 0.$$

The following limiting behaviour of X_t as t tends to infinity can be inferred:

- if $H < 0$, then X_t converges in distribution to 0, $X_t \stackrel{d}{\to} 0$;
- if $H = 0$, then $X_t \stackrel{d}{=} X_1$;
- if $H > 0$ and $X_t \neq 0$, then $|X_t| \stackrel{d}{\to} \infty$.

Analogously, for t going to 0, we have:

- if $H < 0$ and $X_t \neq 0$, then $|X_t| \xrightarrow{d} \infty$;
- if $H = 0$, then $X_t \stackrel{d}{=} X_1$;
- if $H > 0$, then $X_t \xrightarrow{d} 0$.

If the case $X_t \equiv 0$ is excluded, then these properties imply that X_t is stationary only if $H = 0$, but this exception is not interesting because it indicates that X_t is equal to X_1 for every t with probability 1. For the purpose of modelling data that look stationary and of explaining the relationship between fBm long-dependency and self-similarity, only self-similar processes with stationary increments will then be considered. Therefore the range of H will be restricted to $H > 0$, as for $H < 0$ the process would not be measurable. Furthermore it is supposed that $X_0 = 0$.
To obtain the covariance function $\gamma_X(t,s) = Cov(X_t, X_s)$ of a self-similar process X_t with stationary increments the following properties have to be considered. Assume $\mathbb{E}(X_t) = 0$ and denote by σ^2 the variance of the increments, given by

$$\begin{aligned}\sigma^2 &= \mathbb{E}\left[(X_t - X_{t-1})^2\right] \\ &= \mathbb{E}\left[(X_1 - X_0)^2\right] \\ &= \mathbb{E}\left[X_1^2\right].\end{aligned}$$

Then, for $t > s$

$$\begin{aligned}\mathbb{E}\left[(X_t - X_s)^2\right] &= \mathbb{E}\left[(X_{t-s} - X_0)^2\right] \\ &= \mathbb{E}\left[(X_{t-s})^2\right] \\ &= \mathbb{E}\left[((t-s)^H X_1)^2\right] = \\ &= \sigma^2 (t-s)^{2H}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left[(X_t - X_s)^2\right] &= \mathbb{E}\left[X_t^2\right] + \mathbb{E}\left[X_s^2\right] - 2\mathbb{E}\left[X_t X_s\right] = \\ &= \sigma^2 t^{2H} + \sigma^2 s^{2H} - 2\gamma_X(t,s).\end{aligned}$$

Hence, matching the two equations, the following expression for the covariance function is obtained

$$\gamma_X(t,s) = \frac{1}{2}\sigma^2\left[t^{2H} - (t-s)^{2H} + s^{2H}\right]$$

Let now $Y_i = X_i - X_{i-1}$ be the increment at time i, for $i \in \mathbb{N}$. The covariances between two elements of the increment sequence can be calculated as

$$\begin{aligned}\gamma_Y(k) &= Cov(Y_i, Y_{i+k}) = Cov(Y_1, Y_{1+k}) \\ &= Cov(X_1 - X_0, X_{k+1} - X_k) = Cov(X_1, X_{k+1} - X_k)\end{aligned}$$

$$\begin{aligned}\gamma_Y(k) &= \mathbb{E}(X_1(X_{k+1} - X_k)) - \mathbb{E}(X_1)\mathbb{E}(X_{k+1} - X_k) \\ &= \mathbb{E}(X_{k+1}X_1) - \mathbb{E}(X_k X_1) \\ &= \gamma_X(k+1, 1) - \gamma_X(k, 1) \\ &= \tfrac{1}{2}\sigma^2 \left[(k+1)^{2H} - k^{2H} + 1^{2H} - k^{2H} + (k-1)^{2H} - 1^{2H}\right]\end{aligned}$$

Therefore the following formula is obtained

$$\gamma_Y(k) = \tfrac{1}{2}\sigma^2 \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \quad \text{for } k \geq 0 \text{ and } \gamma(k) = \gamma(-k) \text{ for } k < 0$$

The correlations are then given by

$$\rho(k) = \tfrac{1}{2} \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \quad \text{for } k \geq 0 \text{ and } \gamma(k) = \gamma(-k) \text{ for } k < 0$$

The equation can be written as

$$\rho(k) = \tfrac{1}{2}k^{2H}g(k^{-1}) \quad \text{where } g(x) = (1+x)^{2H} - 2 + (1-x)^{2H}$$

The asymptotic behaviour of $\rho(k)$ follows by the Taylor expansion of $g(x)$ at the origin. More precisely

$$\begin{aligned}g(x) &= g(0) + g'(0)(x-0) + \frac{1}{2}g''(0)(x-0)^2 + \cdots \\ &= 2Hx - 2Hx + \frac{1}{2}\left[2H(2H-1) + 2H(2H-1)\right]x^2 + \cdots \\ &= 2H(2H-1)x^2 + \cdots\end{aligned}$$

So, if $0 < H < 1$ and $H \neq 1/2$, the first non-negative term of the series is $2H(2H-1)x^2$. Hence, as k tends to infinity, $\rho(k)$ is equivalent to $H(2H-1)k^{2H-2}$, which implies that

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \overset{k\to\infty}{\longrightarrow} 1.$$

For $1/2 < H < 1$, the correlations decay to zero so slowly that the process $(Y_t)_{t\geq 0}$ has long memory, according to Definition (2.2). Moreover, the correlations are such that

$$\sum_{k=-\infty}^{+\infty} \rho(k) = \infty.$$

For $H = 1/2$, all correlations at non-zero lags are zero, which implies that the observations Y_i are uncorrelated. Finally, for $0 < H < 1/2$ the process has short-range memory and the correlations sum up to zero. To conclude, what happens when $H \geq 1$ has to be considered. For $H = 1$, all correlations are equal to 1, no matter how far apart in time the observations are. For $H > 1$, the function $g(k^{-1})$ diverges to infinity and contradicts the fact that $\rho \in [-1, 1]$.

It can be then inferred that if the increments of a self-similar process are stationary, then the increments themselves show the property of long-range dependence when $1/2 < H < 1$ and fBm is itself an example of a self-similar process with stationary increments.

Alternatively, in order to better understand the structure of the power law for a self-similar process the following quantities can be considered

$$F_t(x) = F_1(\frac{x}{t^H})$$
$$f_t(x) = \frac{1}{t^H} f_1(\frac{x}{t^H})$$

where F and f are respectively the cumulative distribution function and the probability density function of the process X_t.
By setting $x = 0$, f_t is represented by the scaled probability function at time 1

$$f_t(0) = \frac{1}{t^H} f_1(0). \tag{2.3}$$

Remark 2.6. In general, there are two approaches to check self-similarity. The first one is based on Equation (2.3) for estimating the Hurst parameter, initially evaluating $f_t(0)$ through an empirical histogram and then using regression for an estimation of H. The second approach is called the curve fitting method and is based on the comparison of the aggregation properties of empirical densities.

Chapter 3

Modelling Financial Data

Stochastic modelling is a form of financial modelling that includes one or more random variables. The purpose of such modelling is to estimate how probable outcomes are and to forecast the variations of prices and returns on assets and asset classes, such as bonds and stocks, over time. For example, when used for portfolio evaluation, various simulations of how a portfolio may perform are developed based on probability distributions of individual stock returns.
In particular, Brownian motion models for financial markets are based on the work of Robert C. Merton and Paul A. Samuelson and are concerned with defining the concepts of financial assets and markets, portfolios, gains and wealth in terms of continuous-time stochastic processes. Under this models, assets have continuous prices evolving continuously in time, driven by Brownian motion processes.

3.1 Distribution of Stock Prices and Log Returns

Before presenting various types of stochastic models, it is necessary to introduce the quantities that will be used in the following sections.
If one assumes that the price of a stock P follows a geometric Brownian motion

$$dP_t = \mu P_t dt + \sigma P_t dB_t$$

then the logarithm of the price follows a generalized Brownian motion

$$d\ln(P_t) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t$$

where P_t is the price of the stock at time t and B is a Wiener process. Therefore the change in log price from t to T, $\ln(P_T) - \ln(P_t)$, is normally distributed as

$$\ln(P_T) - \ln(P_t) \sim N\left[\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right]$$

$$\ln\left(\frac{P_T}{P_t}\right) \sim N\left[\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right].$$

Consequently, conditional on the price P_t at time t, the log price at time $T > t$ is normally distributed as

$$\ln(P_T) \sim N\left[\ln(P_t) + \left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right]$$

As shown in Tsay (2015), the conditional mean and variance of P_T are

$$\mathbb{E}(P_T) = P_t e^{\mu(T-t)}$$
$$Var(P_T) = P_t^2 e^{2\mu(T-t)}\left\{\exp\left[\sigma^2(T-t)\right] - 1\right\}$$

Knowing prices and their distributions, one can calculate the log returns, or continuously compound returns, of an asset as

$$r_{T-t} = \ln\left(\frac{P_T}{P_t}\right) = p_T - p_t$$

Their distribution is clearly given by

$$r_{T-t} \sim N\left[\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right].$$

3.2 The Black-Scholes-Mertons Model

The Black-Scholes-Merton, or Black-Scholes, model is a mathematical model, developed by F.Black and M.Scholes in the late 60s, that provides a closed-form formula for European call and put options pricing. This model assumes that the market consists of at least one risky asset, called stock, and one riskless asset, called bond or cash. The following assumptions have to be made on the assets:

- the rate of return of the riskless asset is constant and equal to r;
- the instantaneous log returns of the stock price follow a geometric Brownian motion with constant drift and volatility;
- the stock does not pay dividends;

and on the market:

- there is no arbitrage opportunity, i.e. there is no chance to make riskless profit;
- any quantity of cash can be sold and bought at the riskless rate;
- short selling is allowed for every amount of money;
- transactions don't require any additional cost, i.e. the market is frictionless.

In the hypothesis that the prices of the financial assets at time t, S_t, satisfy the following equation, i.e. follow a geometric Brownian motion

$$dS_t = rS_t dt + \sigma S_t dB_t,$$

where r and σ are the risk-free interest rate and the volatility, the stochastic differential equation of the Black-Scholes model for the pricing of a European option on a non-dividend paying stock is given by

$$S_t = S_0 + \int_0^t rS_u du + \int_0^t \sigma S_u dB_u.$$

Let $V(t, S_t)$ be the price of a derivative in function of time t and stock price S_t. With the application of Itô's formula, the differential of $V(t, S_t)$ is obtained as

$$dV(S_t, t) = \left[\frac{\partial V(t, S_t)}{\partial t} + rS_t \frac{\partial V(t, S_t)}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V(t, S_t)}{\partial S_t^2}\right] dt + \sigma S_t \frac{\partial V(t, S_t)}{\partial S_t} dB_t.$$

Then one can consider a certain portfolio, called the delta-hedge portfolio, consisting of $\frac{\partial V(t,S_t)}{\partial S_t}$ shares of the risky asset S_t and the derivative on that asset. Its value is

$$\phi(t) = \frac{\partial V(t, S_t)}{\partial S_t} S_t - V(t, S_t)$$

Calculating the total profit or loss from changes in the values of the holdings over time and considering the absence of arbitrage, one can derive the partial differential equation for a European option, called **Black-Sholes equation**, which is given by

$$\frac{\partial V(t, S_t)}{\partial t} + rS_t \frac{\partial V(t, S_t)}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V(t, S_t)}{\partial S_t^2} - rV(t, S_t) = 0. \tag{3.1}$$

More details about the derivation of the model can be found in Focardi et al. (2017).

Recalling that the value of an option at the time that the option matures, i.e. T, is

$$\begin{aligned} C(T, S_T) &= \max\{S_T - K, 0\} \quad \text{for European call options} \\ P(T, S_T) &= \max\{K - S_T, 0\} \quad \text{for European put options} \end{aligned}$$

Black, Scholes and Merton obtained from Equation (3.1) the following closed-form formulas for the prices of European call options, $C(T, S_T)$, and European put options, $P(t, S_t)$, at time $t < T$:

$$\begin{aligned} C(t, S_t) &= G(d_1)S_t - G(d_2)Ke^{-r(T-t)} \\ P(t, S_t) &= Ke^{-r(T-t)} - S_t + C(t, S_t) \\ &= G(-d_2)Ke^{-r(T-t)} - G(-d_1)S_t \end{aligned}$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right]$$
$$d_2 = d_1 - \sigma\sqrt{T-t}$$

and where G, S and K are the cumulative distribution of the standard Gaussian distribution, spot price, and strike price, respectively.

The Black-Scholes model is widely employed as a useful approximation to reality, but proper applications require an understanding of its limitations. First of all the Black-Scholes model cannot capture the behaviour of tails and central peaks that have been noticed for empirical returns distribution. The Black-Scholes equation assumes a Gaussian distribution for the price changes of the underlying asset, but it has been observed that asset prices have significant skewness and kurtosis. This means high-risk downward moves often happen more often in the market than a Gaussian distribution predicts. Furthermore, in order to be able to use the theory in a real situation, numerical estimates of all the input parameters are needed. In the Black-Scholes model the input data consists of S, K, r, σ, T and t. Out of these six parameters, S, K, r, T and t can be observed directly, whereas an estimate of the volatility σ is needed. There are two basic approaches to find it: historic volatility and implied volatility. **Historical volatility** is the realized volatility of the underlying asset over a previous time period. It is determined by measuring the standard deviation of the underlying asset from the mean during that time period. This differs from the implied volatility determined by the Black-Scholes method, as it is based on the actual volatility of the underlying asset. **Implied volatility** is a measure of the estimation of the future variability for the asset underlying the option contract. The value of σ is obtained through the valuation of a benchmark option by the market.

In both estimation methods, volatility is assumed to be constant, while in reality it is often varying over time. If one plots implied volatility as a function of the exercise price, an horizontal straight line should be obtained. Contrary to this, the graph of the observed implied volatility function often looks like the smile of the Cheshire cat and for this reason the implied volatility curve is termed the **volatility smile**. The main reason for this distortion is the fact that the volatility heavily depends on the calendar time, the time to maturity and the moneyness of the option.

3.3 Stochastic Volatility Models

Stochastic volatility models - SV models - for options were developed because of a need to modify the Black-Scholes model for option pricing, as it failed to effectively take into account the volatility in the price of the underlying asset. In particular,

models based on Black-Scholes assume that the underlying volatility is constant over the life of the derivative, and unaffected by the changes in the price level of the underlying security. Furthermore, these models cannot explain long-observed features of the implied volatility surface such as volatility smile and skew. By assuming that the volatility of the underlying price is a stochastic process rather than a constant, it becomes possible to price derivatives more accurately.

The basic model for SV models is given by

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dB_t^S$$
$$d\nu_t = \alpha_{\nu,t} dt + \beta_{0_{\nu,t}} dB_t^\nu$$

where S, the derivative's underlying asset price, follows a geometric Brownian motion B^S and ν_t, the variance function of S_t, is also driven by a Brownian motion B^ν; B_t^S and B_t^ν are correlated with correlation ρ.
Different specifications of the volatility process are possible. Widely used specifications are the Hull and White model (1993), the Heston model (1993) and the SABR model (2002).

Hull and White model Hull and White have proposed an option pricing model in which the volatility of the underlying asset appears not only time-varying, but also associated with a specific risk, according to the following paradigm

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t^S \tag{3.2}$$
$$d(\ln \sigma_t) = k(\theta - \ln \sigma_t)dt + \gamma dB_t^\nu \tag{3.3}$$

where S_t is the price of the stock, σ_t is its instantaneous volatility and (B_t^S, B_t^ν) is a standard bivariate Brownian process.

Heston Model The basic Heston model assumes that the randomness of the variance process varies as the square root of variance:

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dB_t^S$$
$$d\nu_t = \kappa(\theta - \nu_t)dt + \xi\sqrt{\nu_t} dB_t^\nu$$

where:

- μ is the rate of return of the asset;
- θ is the long variance and the limit, as t tends to infinity, for the expected value of ν_t;
- κ is the rate at which ν_t reverts to θ;
- ξ is the volatility of the volatility and determines the variance of ν_t.

In other words, the Heston SV model assumes that the variance is a random process that exhibits a tendency to revert towards a long-term mean θ at a rate κ, exhibits a volatility proportional to the square root of its level and whose source of randomness is correlated with the randomness of the underlying's price processes.

SABR Model The SABR model, standing for Stochastic Alpha, Beta, Rho, is specified by the system of SDEs

$$dS_t = \sigma_t S_t^{\beta} dB_t^S$$
$$d\sigma_t = \alpha \sigma_t dB_t^{\nu}$$

where $\sigma_0 = \alpha > 0, 0 \leq \beta \leq 1, \nu \geq 0$, and B^S and B^{ν} are Brownian motion with correlation ρ. The main feature of the SABR model is being able to reproduce the smile effect of the volatility smile.

For more details about the models see Hull and White (1993), Heston (1993) and Hagan et al. (2002).

3.4 Fractional SV models

Since empirical studies have shown that the decay of the autocorrelation of volatility follows a power law, an extension to SV models has been presented. In particular, in order to respect the evidence of volatility persistence, it was proposed to model volatility via fractional processes.
The first **continuous-time fractional stochastic volatility model** - FSV model - was introduced by Comte and Renault in 1998. With their studies, they were able to extend the Hull and White SV model to a continuous-time long-memory model, by replacing the Brownian motion B^{ν} in (4.3) with a fBm B^H, considering H restricted to $H \in (1/2, 1)$. The FSV model is then described as

$$dS_t = \mu S_t dt + \sigma_t S_t dB_t^S$$
$$d(\ln \sigma_t) = k(\theta - \ln \sigma_t)dt + \gamma dB_t^H.$$

Under this model, the volatility process is described by a fractional Ornstein-Uhlenbeck process, that is the standard Ornstein-Uhlenbeck process where the Brownian motion is replaced by a fractional Brownian motion.
The **Ornstein-Uhlenbeck process** is a stochastic process that is both a Gaussian and Markovian process and that allows linear transformations of the space and time variables. Furthermore, it is a mean-reverting process, which tends to drift towards its long-term mean. The process can be hence considered to be a modification of the random walk in continuous time in which the properties of the process have been changed

so that there is a tendency of the walk to move back towards a central location, with a greater attraction when the process is further away from the centre.
In particular, a fractional Ornstein-Uhlenbeck process $(X_t)_{t\geq 0}$ is defined as the solution to the stochastic differential equation

$$dX_t = k(\theta - X_t)dt + \gamma dB_t^H$$

where $\theta \in \mathbb{R}$ and γ,k are positive parameters. An explicit solution can be found in Focardi at al. (2017) and is given by

$$X_t = \gamma \int_{-\infty}^{t} e^{-k(t-s)} dB_t^H + \theta. \tag{3.4}$$

One of the main problems in using fBm is that this kind of model is not arbitrage-free. In particular, it is know that arbitrage possibilities can be eliminated if and only if the underlying asset price is a semimartingale, as the semimartingale property provides the general framework for the theoretical development of arbitrage pricing. The following theorem holds, proving the possibility of the existence of arbitrage for fractional Brownian motion.

Theorem 3.1. *Let $(B_t)_{t\geq 0}$ a stochastic process with Hurst parameter $H \in (0, 1/2) \cup (1/2, 1)$, then the process is not a semimartingale.*

For a proof of Theorem (3.1) see Rogers (1997).

3.5 Rough FSV model

An alternative to traditional models was presented in 2014 by Jim Gatheral, Thibault Jaisson and Mathieu Rosenbaum. Their work gave life to a new generation of stochastic volatility models, as they proved that, in order to be realistic and to better forecast future behaviours, the driving processes of the model should be considered of short-memory nature.

Empirically, it has been proven that the distribution of increments of the log-volatility is close to a Gaussian and that its smoothness is constant in time. This analysis suggest the following model

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu(B_{t+\Delta}^H - B_t^H) \tag{3.5}$$

where ν is a positive constant and B^H is a fractional Brownian motion with Hurst parameter equal to the measured smoothness of the volatility. The previous equation

can be written as

$$\sigma_t = \sigma \exp\left\{\nu B_t^H\right\}$$

where σ is another positive constant. The main issue with the proposed model would be the fact that it is not stationary, whereas stationarity is required to ensure mathematical tractability. Hence, the final specification of the **Rough Fractional Stochastic Volatility model** equation for volatility in $[0, T]$ is

$$\sigma_t = exp\left\{X_t\right\}$$

where X_t satisfies Equation (3.4) for some $\gamma > 0$, $k > 0$, $\theta \in \mathbb{R}$ and $H < 1/2$.

Remark 3.1. Taking $k \ll 1/T$ implies that the dynamics of our process is close to that of a fBm. Furthermore one can notice that the choice of $H < 1/2$ permits to reproduce the observed smoothness of the volatility process. Moreover this constraint is also coherent with mean-reversion.

It can be demonstrated that the following property holds:

Proposition 3.1. *Let $q > 0$, $t > 0$, $\Delta > 0$ and let X^k be defined by (3.4) for some $k > 0$. As k tends to zero,*

$$Cov(X_t^k, X_{t+\Delta}^k) = Var(X_t^k) - \frac{1}{2}\nu\Delta^{2H} + o(1).$$

As a consequence, for fixed t, the covariance between X_t and $X_{t+\Delta}$ is linear with respect to Δ^{2H}. The covariance function of the volatility in the RFSV model is then obtained as

$$\mathbb{E}(\sigma_{t+\Delta}\sigma_t) = e^{2\mathbb{E}(X_t^k)+2Var(X_t^k)}e^{-\nu^2\frac{\Delta^{2H}}{2}}.$$

Hence, in the RFSV model, $\log(\mathbb{E}(\sigma_{t+\Delta}\sigma_t))$ is linear in Δ^{2H}.

Finally, regarding long-memory existence, this model proves that, although the volatility may exhibit some form of persistence, it does not present any long-memory in the classical power law sense.
More details about the concept of this section and the all the proves can be found in Gatheral et al. (2014).

Chapter 4

The Hurst Parameter

Since ancient times, the Nile River has been known for its characteristic long-term behaviour, as long periods of dryness were followed by long periods of yearly returning floods. In 1951 the celebrated British hydrologist H.E. Hurst published a paper entitled, "The Long-Term Storage Capacity of Reservoirs", in which he investigated the question of how to regularize the flow of the Nile River so that architects could construct an appropriately sized reservoir system. The paper dealt specifically with the modelling of reservoirs, but as it turned out, the results also held for a number of other natural systems.
In fact, many years later, while investigating the fractal nature of financial markets, Benoit Mandelbrot came across Hurst's work and, recognizing the potential therein, introduced to fractal geometry the term Generalized Hurst Exponent, referring to the measure of the long-term memory of a time series.

The phenomenon of long memory was observed in applications long before appropriate stochastic models were known. Several methods have been now introduced to estimate the long-memory parameter. Despite their different applications and characteristics, it is extremely important to introduce both heuristic and more theoretical procedure based on Maximum Likelihood.

4.1 Heuristic methods for H estimation

Several heuristic methods to estimate the long-memory parameter H were suggested, such as the R/S statistic, the log-log correlogram, the log-log plot of $Var(\bar{X}_n)$ versus n, least squares regression in the spectral domain and periodogram.
These non-parametric methods are mainly useful as diagnostic tools, and, in particular, they are adopted to identify an initial estimation of H. They are yet less suitable for statistical inference, as, for most of these methods, it is not easy to obtain confidence intervals.

4.1.1 R/S Statistic

The **rescaled range statistic**, or R/S statistic, is a statistical measure of the variability of a time series introduced by the British hydrologist Harold Edwin Hurst in 1951. Its purpose is to provide an assessment on how the apparent variability of a series changes with the length of the time-period being considered.

The R/S statistic is given by the ratio between the range of the data aggregated over blocks of length k and the sample standard deviation of the data aggregated at the same scale. More precisely, let $(y_t)_{t=1}^N$ be a time series. Its range is defined according to the following relation

$$R(t_i, k) = \left[\max_{0 \le u \le k} (W(t_i, k, u)) - \min_{0 \le u \le k} (W(t_i, k, u)) \right]$$

where $W(t_i, k, u)$ is considered equal to

$$W(t_i, k, u) = y_{t_i+u} - y_{t_i} - u\mathbb{E}(t_i, k)$$

and $\mathbb{E}(t_i, k)$ is the sample mean with index t_i in the interval $(t_i, t_i + k)$, that is

$$\mathbb{E}(t_i, k) = \frac{1}{k} \sum_{j=t_i}^{t_i+k} y_j.$$

Moreover, the sample standard deviation in the interval $(t_i, t_i + k)$ is

$$S(t_i, k) = \sqrt{\frac{1}{k} \sum_{j=t_i}^{t_i+k} (y_j - \mathbb{E}(t_i, k))^2}.$$

Hence, the rescaled range statistic is given by

$$\frac{R}{S}(t_i, k) = \frac{R(t_i, k)}{S(t_i, k)} \quad \forall t_i \forall k.$$

The ratio R/S generally behaves like a power law, i.e. $\mathbb{E}\left[\frac{R}{S}(t_i, k)\right] = Ck^H$ where C is a positive, finite constant independent of k. To calculate the estimation of parameter H, one can consider that

$$\log \mathbb{E}\left[\frac{R}{S}(t_i, k)\right] \approx c + H \log k.$$

and plot the logarithm of the size of each series, i.e. k, versus the logarithm of the rescaled range, in what is called a **pox plot**. A straight line can then be fitted, that represents the ultimate behaviour of the data, corresponding to their expected value. The coefficients can be estimated by least square: the slope of the line gives $\widehat{H}$.

Remark 4.1. For any stationary process with short-range dependence, R/S should behave asymptotically like a constant times $k^{1/2}$, so $\log(R/S)$ should be scattered around a straight line with slope $\frac{1}{2}$. On the other hand, for long-memory processes, R/S behaves like a constant times k^H for some $H > \frac{1}{2}$.

One of the problems with this technique, which is common to many Hurst parameter estimators, is knowing which value of k to consider. For small k short term correlations dominate and the analyses are not valid. For large k then there are few samples and the value of $(R/S)_k$ will not be accurate. Furthermore, the exact distribution of $(R/S)_k$ seems to be difficult to derive and depends on the actual distribution of the data generating the process.

Two algorithms have been proposed for computing the Hurst parameter: dynamic blocks R/S Statistic algorithm and static blocks R/S Statistic algorithm. The first one is usually applied to financial, economic and hydrological time series, whereas the second one is used in computer engineering. The **dynamic blocks R/S Statistic algorithm** partitions the original length N of the time series in blocks of size k, and computes the R/S Statistic for each block, i.e., obtains $R(t_i, k)$, $t_i = ik$ and $i = 0, 1, 2, ..., N/k$. This procedure is repeated for several blocks size values k. A log-log plot of R/S Statistic values versus the k values results in an estimate of the Hurst parameter H. When the algorithm has been implemented, it is really important to select the cut-offs points, to guarantee the highest accuracy possible. Their values depend on the data that are considered and on the required accuracy.

4.1.2 Log-log correlograms analysis

Correlograms, i.e. plots of empirical correlations $\hat{\rho}(k)$ against the lag k, are useful diagnostic plots for short-memory processes, but they can be used even for detecting long-memory.
As we know, long-memory is characterized by a slow decay of correlations proportional to k^{2H-2} for some $\frac{1}{2} < H < 1$. Despite standard correlograms show this behaviour, it is difficult to tell whether ultimately correlations follow an hyperbolic curve, proving long-range dependency, or an exponential curve. Furthermore, distinguishing different values of H results to be complicated, as well as effectively estimating correlations at big lags.
A more suitable plot to identify long-memory can be obtained by representing $\log|\rho(k)|$ against $\log k$, in what is called a **log-log correlogram**. In this case, if the asymptotic decay follows k^{2H-2}, then the points in the graph should be scattered around a straight line with negative slope approximately equal to $2H - 2$. It results that

$$\widehat{H} = 1 + \frac{1}{2}(slope)$$

On the other hand, for short memory processes, the log-log correlogram should show divergence to $-\infty$ at an exponential rate.
Despite being a useful diagnostic method for very long time series, for relatively short ones and if H is close to 1/2, it is very difficult to detect long-memory in the data by looking at the correlogram only.

4.1.3 Aggregate variance technique

An important property of long-memory processes is that the variance of the sample mean converges slower to zero than n^{-1} according to the Theorem (2.3), where N is dimension of the sample. Hence, it can be inferred that $Var(\bar{X}_n) \approx cn^{2H-2}$, where $c > 0$ and $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_i$.
This consideration allows to introduce the following procedure to estimate H:

1. Let k be an integer. For different values of k in $2 \leq k \leq \lceil N/2 \rceil$ and a sufficient number, m_k, of subset of length k, calculate the sample means $\bar{X}_1(k), .., \bar{X}_{m_k}(k)$ and the overall mean

$$\widehat{X}(k) = \frac{1}{m_k}\sum_{j=1}^{m_k} \widehat{X}_j(k)$$

2. For each k, calculate the sample variance of the sample means $\bar{X}_1(k), .., \bar{X}_{m_k}(k)$

$$s^2(k) = \frac{1}{m_k - 1}\sum_{k=1}^{m_k} \left(\widehat{X}_j(k) - \widehat{X}(k)\right)^2$$

3. Plot $\log s^2(k)$ against $\log(k)$, in what is called a **variance plot**, and fit a simple least square line through the resulting points in the plane.

As well as in log-log correlograms, for large values of k, the points in the plot are expected to be scattered around a straight line with a negative slope 2H-2. It results that

$$\widehat{H} = 1 + \frac{1}{2}(slope).$$

For short-range dependence or independence among the observations, the slope is equal to -1.
The problems with this method are clearly the same as for R/S statistic and log-log correlograms.

4.1.4 Least square regression in spectral domain

In the frequency domain, analysis of time series is merely the analysis of a stationary process by means of its spectral representation at the origin. So the least square

regression in the spectral domain exploits Remark (2.3):

$$f(\lambda) \simeq c_f |\lambda|^{1-2H},$$
$$\text{or equivalently } \log f(\lambda) \simeq \log c_f + (1-2H)\log|\lambda|.$$

An estimator for the spectral density function $f(\lambda)$ is the **periodogram**, given by

$$I(\lambda_j) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}(X_t - \hat{X}_n)e^{it\lambda_j}\right|^2 = \frac{1}{2\pi}\sum_{k=-(n-1)}^{n-1}\hat{\gamma}(k)e^{ik\lambda_j}$$

where $\lambda_j = 2\pi j/n$ are the Fourier frequencies. In practice, the periodogram is the sample equivalent of the spectral density in which the covariances $\gamma(k)$ are replaced by $\widehat{\gamma}(k)$. As the estimator is asymptotically unbiased the following holds

$$\lim_{n\to\infty}\mathbb{E}\left(I(\lambda)\right) = f(\lambda).$$

Usually, $I(\lambda)$ is calculated at the Fourier frequencies $\lambda_{k,n} = 2\pi k/n$, for $k = 1, .., n^*$ where n^* is the integer part of $(n-1)/2$. For long-memory processes, the following result can be demonstrated:

$$\log I(\lambda_{k,n}) \simeq \log c_f + (1-2H)\log\lambda_{k,n} + \log\xi_k \tag{4.1}$$

where ξ_k are independent standard exponential random variables. If one defines

$$y_k = \log I(\lambda_{k,n})$$
$$x_k = \log\lambda_{k,n}$$
$$\beta_0 = \log c_f - C \quad \beta_1 = 1-2H$$

and the error terms $e_k = \log\xi_k + C$, then (4.1) can be written as

$$y_k = \beta_0 + \beta_1 x_k + e_k.$$

This is a regression equation with i.i.d. errors e_k with zero mean, so the coefficients β_0 and β_1 can be estimated by least square regression. The estimation of H is then

$$\widehat{H} = \frac{1-\widehat{\beta_1}}{2}$$

One of the main problems with this method is that LRD, as known, has an asymptotic definition. So the spectral density is proportional to λ^{1-2H} only in a small neighbourhood of zero, but not in the whole interval $[-\pi, \pi]$, where the estimation could be highly biased. A deep analysis on spectral density estimations goes beyond the scope of the text, but more details can be found in Beran (1994).

4.1.5 Examples on real data

In the following section a real data example will be presented and the estimation methods introduced in the chapter will be tested and compared through the use of R. Figure 4.1 displays the series of daily stock closing prices for Tesla Motors Inc (TSLA), from 29/06/2010 to 23/06/2017. The data are available on Yahoo Finance.
We will focus on the adjusted close prices because they have been corrected to include any distributions and corporate actions that occurred at any time prior to the next day's open and, for this reason, they are more useful when examining daily returns.

```
> library(timeSeries)

> TSLA <- read.csv("TSLA.csv", header = T)
> head(TSLA)
        Date  Open  High   Low Close Adj.Close   Volume
1 2010-06-29 19.00 25.00 17.54 23.89     23.89 18766300
2 2010-06-30 25.79 30.42 23.30 23.83     23.83 17187100
3 2010-07-01 25.00 25.92 20.27 21.96     21.96  8218800
4 2010-07-02 23.00 23.10 18.71 19.20     19.20  5139800
5 2010-07-06 20.00 20.00 15.83 16.11     16.11  6866900
6 2010-07-07 16.40 16.63 14.98 15.80     15.80  6921700

> date <- as.Date(TSLA[, "Date"])
> adj.close <- TSLA[, "Adj.Close"]
> prices <- timeSeries(adj.close, date)
> head(prices)
            TSLA
2010-06-29 23.89
2010-06-30 23.83
2010-07-01 21.96
2010-07-02 19.20
2010-07-06 16.11
2010-07-07 15.80

> seriesPlot(prices)
```

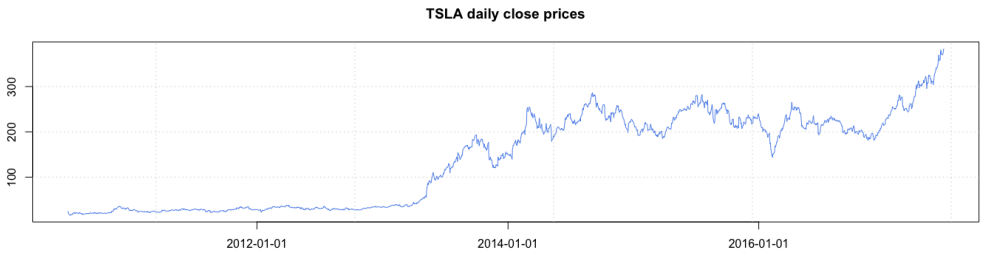

Figure 4.1: TSLA daily close prices.

When studying the characteristics of financial time series, it is preferable to consider returns instead on prices, because they are a complete and scale-free summary of the investment opportunities and because they have more attractive statistical properties. In particular, we will consider log returns or continuously compound returns, given by

$$r_t = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1}$$

where P_t is the asset price at time t and $p_t = \ln P_t$. In R they can be obtained as

```
> logprices <- log(prices)
> returns <- diff(logprices, 1)
> returns <- removeNA(returns)
> head(returns)
                    TSLA
2010-06-30 -0.002514628
2010-07-01 -0.081722720
2010-07-02 -0.134312240
2010-07-06 -0.175470072
2010-07-07 -0.019430319
2010-07-08  0.099902553
```

Alternatively, the function `returns` can be used

```
> returns<-returns(prices, methods="continuous")
> head(returns)
                    TSLA
2010-06-30 -0.002514628
2010-07-01 -0.081722720
2010-07-02 -0.134312240
2010-07-06 -0.175470072
2010-07-07 -0.019430319
2010-07-08  0.099902553
```

To estimate the value of H one can use different R packages. One of the most complete is the `fArma` package, which includes many different functions to estimate the parameter. In particular the R/S Rescaled Range Statistic method estimation can be obtained with the function `rsFit()`, defined as follows

```
rsFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),
  doplot = FALSE, trace = FALSE, title = NULL)
```

The arguments of the function are:

- `x` is the numeric vector of data, an object of class timeSeries, or any other object which can be transformed into a numeric vector;
- `levels` is the number of aggregation levels or number of blocks for which the variances or moments are computed;

- `minnpts` is the minimum number of points or blocksize to be used to estimate the variance or moments at any aggregation level, i.e. k;
- `cut.off` is a numeric vector with the lower and upper cut-off points for the estimation, with default values c(0.7, 2.5).

The function can be applied to the time series `returns`, considering 100 blocks for the iterations and at least 2 points per each block. Applying the algorithm explained in Section 4.1.1 one obtains:

```
> library(fArma)

> RS <- rsFit(returns, levels = 100, minnpts = 2,  doplot = T,
    trace = F)
> RS

Title:
Hurst Exponent from R/S Method

Call:
rsFit(x = returns, levels = 100, minnpts = 2, doplot = T)

Method:
R/S Method

Hurst Exponent:
        H      beta
0.5049195 0.5049195

Hurst Exponent Diagnostic:
   Estimate    Std.Err  t-value     Pr(>|t|)
X 0.5049195 0.02809711 17.97051 3.392375e-31

Parameter Settings:
   n   levels  minnpts cut.off1 cut.off2
1759      100        2        5      316
```

The returned estimation of H is 0.5049 and it indicates that the returns of the TSLA stocks are very close to be truly independent and uncorrelated and that there is no trend effect in the stock prices.

In Figure 4.2 different values of $\log(k)$ versus the corresponding values of $\log(R/S)$ have been plotted. In particular, green points represent those values that resulted in being out of the cut-off points and that have not been considered for the estimation of the regression line.

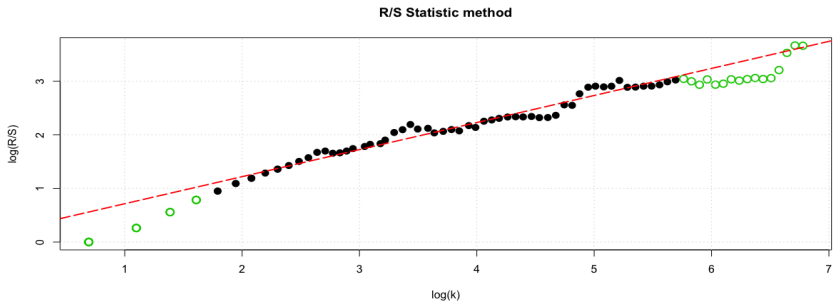

Figure 4.2: R/S statistic method. H=0.5049.

As explained in the previous sections, an alternative to R/S statistic method is the aggregate variance method. To apply it to a financial time series, the function `aggvarFit()` can be used, which is given by

```
aggvarFit(x, levels = 50, minnpts = 3, cut.off = 10^c(0.7, 2.5),
  doplot = FALSE, trace = FALSE, title = NULL)
```

The parameters have the same meanings as for `rsFit()` function.
In particular, considering 100 blocks and at least 2 elements per block, an estimation of H for the TSLA daily returns is obtained as

```
> aV <- aggvarFit(returns, levels = 100, minnpts = 2, doplot = T)
> aV

Title:
Hurst Exponent from Aggregated Variances

Call:
aggvarFit(x = returns, levels = 100, minnpts = 2, doplot = T)

Method:
Aggregated Variance Method

Hurst Exponent:
        H       beta
0.5416380 -0.9167239

Hurst Exponent Diagnostic:
  Estimate    Std.Err  t-value     Pr(>|t|)
X 0.541638 0.02265854 23.90437 1.121317e-42

Parameter Settings:
   n   levels  minnpts cut.off1 cut.off2
1759      100        2        5      316
```

The estimation of H obtained through this method is slightly different from the one obtained through the R/S statistic, but in both cases the value of H is close to 0.5, to indicate the absence of persistence in the data. Figure 4.3 shows the values of the logarithm of the aggregated variance for every value of the logarithm of k. The green points resulted in being out of the cut-off points and have not been considered for the estimation of the regression line.

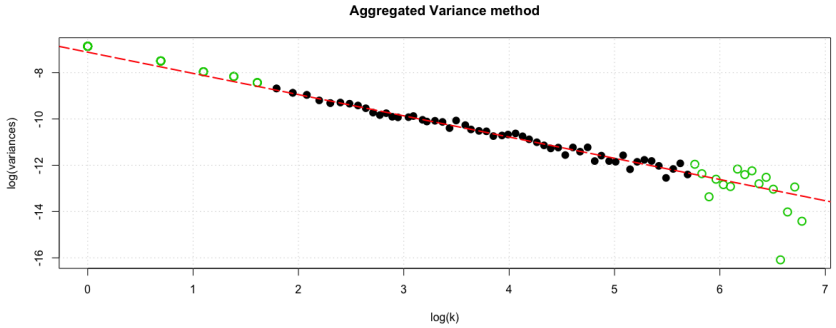

Figure 4.3: Aggregate variance method. H=0.5416.

4.2 Maximum Likelihood Estimators

In the previous section, an introduction to the main heuristic methods for estimating H has been presented. These methods are useful as first informal diagnostic tool for checking the existence of long-memory, but they are not sufficient if an analysis on short-term properties wants to be conducted and, moreover, they are not suitable for statistical inference. A possible alternative approach is to use parametric models and estimate the parameters by maximizing the likelihood. One of the most used methods is Whittle's estimator, which assumes a functional form for the spectral density $f(\lambda)$ and seeks to minimize parameters based upon specific assumptions on the model.

Suppose that $(X_t)_{t\geq 0}$ is a Gaussian stationary process with mean μ and variance σ^2, and, more specifically that $X_1, .., X_t, ..$ are the Gaussian independent increments of a fBm, i.e. fractional Gaussian noises. Moreover assume that Definition (2.2) and Definition (2.4) hold for $1/2 < H < 1$. Let the spectral density f be characterized by the parameter vector $\theta = (\sigma_\varepsilon^2, H)$ or, in other words, suppose that the spectral density comes from a parametric family of densities $f(\lambda) = f(\lambda; \theta)$ where $\theta \in \Theta \subset \mathbb{R}^2$. The following notation will be used:

$$x = (x_1, .., x_n)^t$$
$$\Sigma_n(\theta) = [\gamma(j-l)]_{j,l=1,..,n} \text{ covariance matrix of x}$$
$$|\Sigma_n| \text{ determinant of } \Sigma_n$$

Furthermore, μ will be considered equal to 0 and $(X_t)_{t\geq 0}$ will be assumed to be an invertible linear process, i.e. it can be described as a linear combination of its past values. This last condition implies that X_t can be written as

$$X_t = \sum_{s=0}^{\infty} b(s) X_{t-s} + \varepsilon_t \tag{4.2}$$

$$X_t = \sum_{s=0}^{\infty} a(s) \varepsilon_t \tag{4.3}$$

where ε_t are uncorrelated random variables with zero mean and variance σ_ε^2. Moreover, the asymptotic behaviours of $b(s)$ and $a(s)$ as $k \to \infty$ are given by

$$\begin{aligned} b(s) &\sim c_b k^{-H-\frac{1}{2}} \quad \text{with } c_b \in \mathbb{R}^+ \\ a(s) &\sim c_a k^{H-\frac{1}{2}} \quad \text{with } c_a \in \mathbb{R}^+. \end{aligned}$$

Equations (4.2) and (4.3) imply that the new observations X_t depend only on the past, in a linear way. For more details about equations (4.2) and (4.3) see Beran, p.103 (1994).

Remark 4.2. The presented method is based on the Gaussian likelihood function, as it can be characterized by mean and variance only. Through the use of the central limit theorem, this hypothesis can be carefully extended to other distributions.

Given the observations $x_1, .., x_n$ one wants now to estimate the unknown parameters in $\theta = (\sigma_\varepsilon^2, H)$. The joint probability density function, i.e. the density function of a n-dimensional Gaussian distribution, is associated with the likelihood function and is given by

$$L(x;\theta) = (2\pi)^{-\frac{n}{2}} \, |\Sigma_n(\theta)|^{-\frac{1}{2}} \, e^{-\frac{1}{2} x^t \Sigma_n^{-1}(\theta) x}.$$

The log-likelihood function is then obtained as

$$\ell(x;\theta) = \log L(x,\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_n(\theta)| - \frac{1}{2} x^t \Sigma_n^{-1}(\theta) x.$$

The maximum likelihood estimator - MLE - for θ is obtained by maximizing $\ell(x;\theta)$ with respect to the vector θ. If one defines the bi-dimensional vector

$$\ell'(x;\theta) = \left(\frac{\partial}{\partial \sigma_\varepsilon^2} \ell(x;\theta) \quad \frac{\partial}{\partial H} \ell(x;\theta) \right)^t$$

then the estimator $\hat{\theta}$ is the solution to the system of two equations

$$\ell'(x;\widehat{\theta}) = 0.$$

In particular, the MLE for H is obtained by maximizing the H-dependent part of the

log-likelihood function. In general, $\widehat{\theta}$ is unbiased as $n \to \infty$ and its asymptotic distribution can be computed by looking at the Taylor expansion of ℓ'.

The MLE estimators are clearly more efficient that those studied in the previous section, but the calculation of the exact MLE poses computational problems, as the evaluation of the inverse and the determinant of the variance matrix may be numerically unstable and the number of operations needed grows as the square of the dimension of the dataset. It is therefore necessary to approximate the MLE, to obtain a more computationally efficient procedure. The Whittle's approximation MLE is based on Whittle's approximation for Gaussian processes and can be derived as follows.
The terms in the log-likelihood function that depend on θ are the determinant of the covariance matrix $\log|\Sigma_n(\theta)|$ and the quadratic form $x^t\Sigma_n^{-1}(\theta)x$. More specifically:

- $\log|\Sigma_n(\theta)|$ can be replaced by

$$n(2\pi)^{-1}\int_{-\pi}^{\pi}\log f(\lambda;\theta)d\lambda, \tag{4.4}$$

  according to Grenander et al. (1958);

- Σ_n^{-1} can be replaced by a matrix whose elements are easier to calculate:

$$A(\theta) = [\alpha(j-l)]_{j,l=1,..,n} \tag{4.5}$$

$$\text{where } \alpha(j-l) = (2\pi)^{-2}\int_{-\pi}^{\pi}\frac{1}{f(\lambda;\theta)}e^{i(j-l)\lambda}d\lambda, \tag{4.6}$$

  according to Beran, p.109 (1994).

Combining (4.4) and (4.5), the log-likelihood function becomes

$$\ell^*(x;\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\theta)d\lambda - \frac{1}{2}x^tA(\theta)x$$

As only the last two terms depend on θ, the **Whittle's approximated MLE** is obtained by minimizing the function

$$\ell_W(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\theta)d\lambda + \frac{x^tA(\theta)x}{n} \tag{4.7}$$

with respect to θ. This operation is equivalent to solving the system of nonlinear equations given by

$$\frac{\partial}{\partial\theta_j}\ell_W(\theta)|_{\theta=\hat{\theta}} = 0 \quad j = 1, 2.$$

It can be demonstrated that Whittle's approximated MLE has the same asymptotic distribution as the exact one, thus it is asymptotically efficient for Gaussian processes.

Despite being an efficient alternative to exact MLE, Whittle's approximated MLE requires the computation of n integrals for each value of θ. To solve this issue, one can observe that (4.6) can be approximated reasonably well by the Riemann sum

$$\tilde{\alpha} = 2\frac{1}{(2\pi)^2}\sum_{j=1}^{m}\frac{1}{f(\lambda_{j,m})}e^{ik\lambda_{j,m}}\frac{2\pi}{m},$$

having

$$\lambda_{j,m} = \frac{2\pi j}{m} \text{ for } j = 1, .., m^*$$

where m^* is given by the integer part of $(m-1)/2$. Furthermore, (4.7) can be written in terms of the periodogram as

$$\ell_W(\theta) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi}\log f(\lambda;\theta)d\lambda + \int_{-\pi}^{\pi}\frac{I(\lambda)}{f(\lambda)}d\lambda\right],$$

hence Whittle's approximated MLE can alternatively be obtained by minimizing the function

$$\tilde{\ell}_W(\theta) = 2\frac{1}{2\pi}\left[\sum_{j=1}^{m^*}\log(\lambda_{j,m};\theta)\frac{2\pi}{m} + \sum_{j=1}^{m^*}\frac{I(\lambda_{j,m})}{f(\lambda_{j,m};\theta)}\frac{2\pi}{m}\right].$$

Because the periodogram can be calculated by the fast Fourier transformation, $\tilde{\ell}_W(\theta)$ can be easily obtained.

4.2.1 Examples on real data

In the following section, as well as for heuristic methods, a real data example will be presented and the estimation methods introduced in the chapter will be tested and compared through the use of R.
Through the use of package `FGN`, one can calculate the exact MLE and the Whittle's MLE of H, considering model fitting for fractional Gaussian noises. In particular, the function of interest, `GetFitFGN`, has the following specification

```
GetFitFGN(z, MeanZeroQ = FALSE, algorithm = c("emle", "wmle"),
    ciQ = FALSE)
```

where

- `z` is a time series data vector;
- `MeanZero` is an optional argument, whose value is `TRUE` if the mean is known to be zero;
- `algorithm` indicates which estimation has to calculated, where `emle` stands for exact MLE and `wmle` for Whittle's MLE;

- `ciQ` takes values `TRUE` or `FALSE` and returns, or not, the 95% confidence interval.

Considering the daily log returns for Tesla Motors Inc, derived in Section 4.1.5, the exact MLE of H is

```
> library(FGN)

> emle <- GetFitFGN(as.vector(returns), algorithm = "emle", ciQ =
    TRUE)
> emle

$H
[1] 0.5029346

$loglikelihood
[1] 6034.394

$alpha
[1] 0.9941308

$algorithm
[1] "emle"

$ci
[1] 0.4738316 0.5335103
```

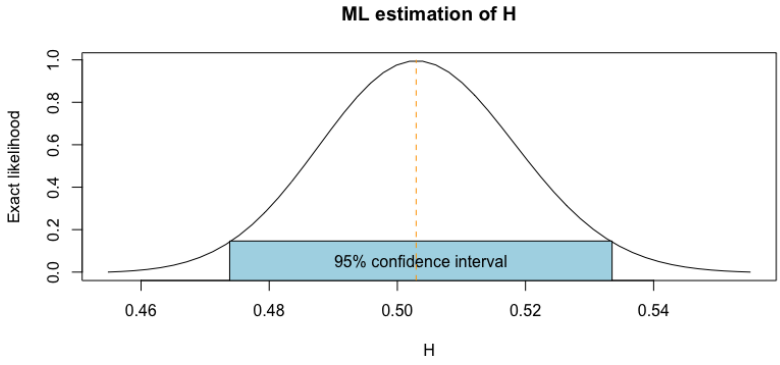

Figure 4.4: Exact MLE. H=0.5029.

The exact MLE of H is equal to 0.5029 and the 95% confidence interval is given by $[0.4738, 0.5335]$.
On the other hand, the Whittle's MLE of H is obtained as

```
> wmle <- GetFitFGN(as.vector(returns), algorithm = "wmle", ciQ =
    TRUE)
> wmle
```

```
$H
[1] 0.5025828

$loglikelihood
[1] 15291.9

$alpha
[1] 0.9948344

$algorithm
[1] "wmle"

$ci
[1] 0.4836450 0.5245051
```

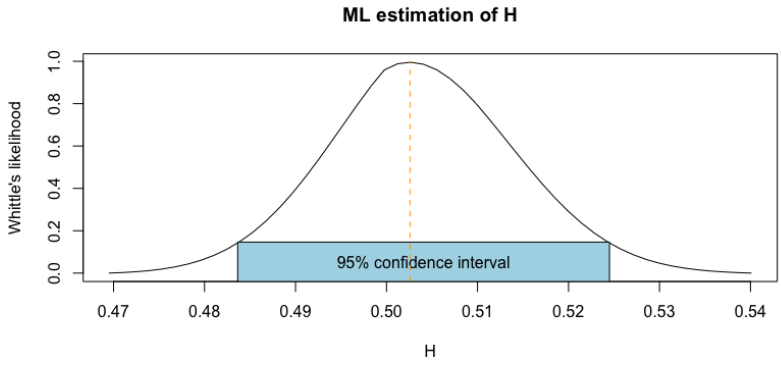

Figure 4.5: Whittle's MLE. H=0.5025.

The Whittle's MLE of H is equal to 0.5026, which is very close to the estimation obtained through the exact method. The 95% confidence interval is $[0.4836, 0.5245]$ and is more narrow than the first one, to indicate higher accuracy and precision.

Conclusions

The characteristics of stochastic processes described in this work make them suitable tools for modelling financial time series. Because of their capability to describe every kind of financial instruments in different ways, all the models presented are nowadays widely used and daily improved.

Concerning Hurst parameter estimations, the examples on real data presented in Sections 4.1.5 and 4.2.1 have allowed us to study and compare the different estimation procedures. More precisely, the estimations obtained through MLE are more accurate, but still very close to the heuristic values. However, as confidence intervals are more explanatory than punctual estimations, the first ones are more suitable for further analysis on H parameter. Depending on his needs, one can decide whether to use heuristic or parametric methods (or both), finding a compromise between computation time and precision.

"Essentially, all models are wrong, but some are useful."

– George E. P. Box

Appendix A
R Code

The following chapter explains the basic code and the packages that have been used to produce the graphs in this work.

Chapter 1

Simulation of a White Noise process.
The `arima.sim()` function can be used to simulate data from a variety of time series models. In particular, an ARIMA with all parameters equal to 0 is a WN process.

```
> WN <- arima.sim(model = list(order = c(0, 0, 0)), n = 500)
> head(WN)
[1]  0.06471903  0.31413574  0.81477864  1.97200008  0.46459139

> plot(WN, col = 3, main = 'WN simulated paths')

> acf(WN)
```

Simulation of a Random Walk process.
As well as for WN, one can consider the `arima.sim()` function to simulate a model with order of integration 1, which consists in a Random Walk.

```
> RW <- arima.sim(model = list(order = c(0,1,0)), n = 500)
> head(RW)
[1]  0.0000000 -0.5672244 -0.7136835 -0.5740660 -3.1056595

> plot(RW,col = 2, main = 'Simulated Random Walk')

> acf(RW)
```

Chapter 2

Simulation of Fractional Brownian Motions.
The package *somebm* includes some functions to generate the time series of Brownian motions. In particular, the function `fbm()` generates a time series of one dimension fractional Brownian motion.

```
> install.packages("somebm")
> library(somebm)

> m1 <- NULL
> for(i in 1:5){
  f <- fbm(hurst = 0.3, n = 1000)
  m1 <- rbind(m1, f)}

> plot.new()
> plot(m1[1, ], type = "l", ylim = range(min(m1), max(m1)))
> for(i in 2:5){
  lines(m1[i, ], col = i)}

> m2 <- NULL
> for(i in 1:5){
  f <- fbm(hurst = 0.5, n = 1000)
  m2 <- rbind(m2, f)}

> plot.new()
> plot(m2[1, ], type = "l", ylim = range(min(m2), max(m2)))
> for(i in 2:5){
  lines(m2[i, ], col = i)}

> m3 <- NULL
> for(i in 1:5){
  f <- fbm(hurst = 0.7, n = 1000)
  m3 <- rbind(m3, f)}

> plot.new()
> plot(m3[1, ], type = "l", ylim = range(min(m3), max(m3)))
> for(i in 2:5){
  lines(m3[i, ], col = i)}
```

Appendix B

Definition 4.1. *A **filtration** $(\mathfrak{F}_t)_{t\geq 0}$ on a probability space $(\Omega, \mathfrak{F}, P)$ is an indexed non-decreasing family of sub-σ-algebras on Ω such that*

$$\mathfrak{F}_t \subseteq \mathfrak{F}, \forall t \geq 0$$
$$s \leq t \Rightarrow \mathfrak{F}_s \subseteq \mathfrak{F}_t.$$

Intuitively, $\mathfrak{F}_t$ represents the available information up to the instance t: more precisely, $\mathfrak{F}_t$ includes all the events that have occurred (or not) by time t.
If $\mathfrak{F}_t^X = \sigma(X_u, 0 \leq u \leq t)$ is the *natural filtration* of the process $(X_t)_{t\geq 0}$, $\mathfrak{F}_t^X$ contains the history of the process until t.

Definition 4.2. *A stochastic process $(X_t)_{t\geq 0}$ is said to be **adapted to the filtration** $(\mathfrak{F}_t)_{t\geq 0}$ if*

$$\sigma(X_t) \subseteq \mathfrak{F}_t \quad \forall t \geq 0$$

which means that X_t is an $\mathfrak{F}_t$-measurable random variable for each time $t \geq 0$.

Bibliography

[1] J. Beran. *Statistics for Long-Memory Processes*, 1994.

[2] T. Björk. *Arbitrage Theory in Continuous Time*, 2009.

[3] J.C. Cano, P. Manzoni. *On the Use and Calculation of the Hurst Parameter with MPEG Videos Data Traffic.* 26th Euromicro Conference, 2000.

[4] V. Capasso, D.Bakstein. *An Introduction to Continuous-Time Stochastic Processes*, 2015.

[5] F. Caravenna. *Moto Browniano e analisi stocastica*, 2011.

[6] R.G. Clegg. *A Practical Guide to Measuring the Hurst Parameter.* International Journal of Simulation: Systems, Science & Technology, Vol. 7(2), p. 3-14, 2006.

[7] F. Comte, E. Renault. *Long memory in Continuous-time Stochastic Volatility Models.* Mathematical Finance, Vol. 8, p. 291-323, 1998.

[8] T. Di Fonzo, F. Lisi. *Serie storiche economiche. Analisi statistiche e applicazioni*, 2005.

[9] H.A. Fallahgoul, S.M. Focardi, F.J. Fabozzi. *Fractional Calculus and Fractional Processes with Application to Financial Economics. Theory and Application,* 2016.

[10] J. Gatheral, T. Jaisson, M. Rosenbaum. *Volatility is rough*, 2014.

[11] U. Grenander, G. Szegö. *Toeplitz Forms and Their Applications*, 1958.

[12] P.S. Hagan, D. Kumar, A.S. Lesniewski, D.E. Woodward. *Managing smile risk.* Wilmott Magazine, p. 84108, 2002.

[13] S.L. Heston *A closed-form solution for options with stochastic volatility with applications to bond and currency options.* The Review of Financial Studies, Vol. 6, p. 327343, 1993.

[14] J. Hull, A. White. *One-factor interest-rate models and the valuation of interest-rate derivative securities.* Journal of Financial and Quantitative Analysis, Vol. 28(02), p. 235254, 1993.

[15] I. Karatzas, S.E. Shreve. *Brownian Motion and Stochastic Calculus*, 1998.

[16] J.F. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*, 2016.

[17] B.B. Mandelbrot, J.W. Van Ness. *Fractional Brownian motions, fractional noises and applications*. SIAM review, Vol. 10, p. 422-437, 1968.

[18] I. Nourdin. *Selected aspects of fractional Brownian motion*, 2012.

[19] J.C. Pacheco, D. Torres Román, L. Estrada Vargas. *R/S Statistic: Accuracy and Implementations*. 18th International Conference on Electronics, Communications and Computers, 2008.

[20] W. Paul, J. Baschnagel. *Stochastic Processes: From Physics to Finance*, 1999.

[21] M.B. Priestley. *Spectral Analysis and Time Series*, 1981.

[22] P.M. Robinson. *Gaussian Semiparametric Estimation of the Long Range Dependence*. The Annals of Statistics, Vol. 23, p. 1630-1661, 1995.

[23] L.C.G. Rogers. *Arbitrage with fractional Brownian motion*. Mathematical Finance, Vol. 7, p. 95-105, 1997.

[24] O.Rose. *Estimation of the Hurst Parameter of Long Range Dependent Time Series*. University of Wurzburg, Research Report Series, 1996.

[25] M.M.A. Sarker. *Estimation of the Self-Similarity parameter in Long Memory Processes*. Journal of Mechanical Engineering, Vol. 38, p. 32-37, 2007.

[26] R.S. Tsay. *Analysis of Financial Time Series*, 2015.

[27] W.W.S. Wei. *Time Series Analysis. Univariate and Multivariate Methods*, 2006.

[28] A. Zygmund. *Trigonometric Series 2*, 1959.