UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Mathematics

Time-Changed Self-Similar processes

An Application To High Frequency Financial Data

by

Wahib Arroum

Thesis for the degree of Doctor of Philosophy

July 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF MATHEMATICS

Doctor of Philosophy

TIME-CHANGED SELF-SIMILAR PROCESSES

by Wahib Arroum

We consider processes of the form $X(t) = \tilde{X}(\theta(t))$ where $\tilde{X}$ is a self-similar process with stationary increments and θ is a deterministic subordinator with a periodic activity function $a = \theta' > 0$. Such processes have been proposed as models for high-frequency financial data, such as currency exchange rates, where there are known to be daily and weekly periodic fluctuations in the volatility, captured here by the periodic activity function. We propose three new methods for estimating the activity and review an existing estimator for it. We present some experimental studies of the estimators performances. We finish with an application to high frequency financial data such as foreign exchange rates and FTSE100 futures contract.

Contents

List of Tables

List of Figures

Acknowledgements

First of all, I would like to thank very much my supervisor Dr. Owen Dafydd Jones for the assistance, advice, help and support he has given freely throughout the course of my PhD study whether it was in Melbourne or at Southampton. I am thankful to the comments and corrections he provided; in particular for providing proofs of Lemma 4.5.1 and Lemma 5.1.1.

I am also grateful to Professor Chris Potts for being my formal supervisor at the University of Southampton and to Professor Russell Cheng for advice and help.

I am also grateful to the School of Mathematics for offering funding for my PhD and to the University of Melbourne for financial support, which made my visit to Melbourne possible. I am also thankful to the Department of Finance (Faculty of Economics and Commerce, University of Melbourne) to provide me with high frequency financial foreign exchange data from SIRCA (Securities Industry Research Centre of Asia-Pacific), otherwise all financial application would not have been possible to do.

A special thanks to Dr. Arougg Jbira for his advice and help. I would also like to thank Israel Vieira for help in computer programming and tips to improve speed of a code, during the times I shared the office with him. I would also like to thank Zaeem Burq with whom I have had many interesting conversations during the times I spent in Melbourne and Fréderic Cortat with whom I have shared good times during my visit to Melbourne.

I would also like to thank those who make my time enjoyable whether it was in Melbourne or at Southampton.

Last but not least, I would very much like to thank my parents, my brothers Anis and Ilyes and my fiance Leila, who has shown continued support, and patience throughout the course of my PhD study.

Abbreviations

Symbols

$\#$	Number of elements

Sets

$\mathbb{R}$	The set of real numbers
$\mathbb{R}^+$	The set of positive real numbers
$\mathbb{N}$	The set of natural numbers
$\mathbb{N}^*$	The set of natural numbers without 0
$\mathcal{C}^n(\mathbb{R})$	Space of continuous functions n times differentiable
$\mathcal{D}(\mathbb{R})$	Space of real valued càdlàg functions

Equalities and Convergence Symbols

$x \approx a$	x is approximatively equal to a
$x \overset{def}{=} y$	x and y are equal by definition
$x \to a$	x tends to a
$x_n \underset{n\to\infty}{\sim} x$	x_n is equivalent to x as n tends to infinity
$X \overset{(d)}{=} Y$	X and Y have the same distribution
$X \overset{(\mathcal{P})}{=} Y$	X and Y are equal in probability
$X \overset{(a.s.)}{=} Y$	X and Y are equal almost surely
$X \overset{(fdd)}{=} Y$	X and Y are equal in the sense of finite-dimensional distributions
$X_n \overset{(d)}{\to} X$	X_n converges to X in distribution
$X_n \overset{(\mathcal{P})}{\to} X$	X_n converges to X in probability
$X_n \overset{(a.s.)}{\to} X$	X_n converges to X almost surely
$X_n \overset{(fdd)}{\to} X$	X_n converges to X in the sense of finite-dimensional distributions

Statistical Notation

$\mathbb{V}ar\,(X)$	Variance of X
$Cov\,(X,Y)$	Covariance between X and Y
$\mathbb{E}\,[X]$	Expectation of X
$\mathbb{P}\,(A)$	Probability of the event A

Probability Distributions

$X \sim \mathrm{P}$	X follow a law P
$\mathcal{U}(a,b)$	Uniform distribution with parameters a and b
$\mathcal{N}(\mu,\sigma^2)$	Normal distribution on with with mean μ and variance σ^2
$\mathcal{T}(\nu,\sigma,\mu)$	Symmetric scaled Student t-distribution
$\Gamma(\alpha,\beta)$	Gamma distribution
$\mathrm{R}\Gamma(\alpha,\beta)$	Inverse Gamma distribution
$\chi^2(\nu)$	Chi squared distribution
$S_\alpha(\sigma,\beta,\mu)$	α-stable distribution
$S\alpha S$	symmetric α-stable law
$\mathcal{E}(a)$	exponential distribution

Abbreviations

i.i.d.	independent, identically distributed
H-ss	H-self-similar
H-sssi	H-self-similar with stationary increments
H-sssi,$S\alpha S$	H-sssi with symmetric α-stable distributed increments
fBm	fractional Brownian motion
Lsm	stable Lévy motion
càdlàg	continue à droite, limite à gauche
H-lass	locally asymptotically self-similar with index H
H-lss	locally self-similar with index H
H-lsssi	locally self-similar with index H and stationary increments
MSE	Mean Square Error
MLE	Maximum Likelihood Estimator
MM	Method of moments
MCS	Minimum chi squared
EDF	Empirical density function
PDF	Probability density function
CDF	Cumulative density function
ACF	Autocorrelation function
UHFD	Ultra High Frequency Data
HFD	High Frequency Data
EBP	Embedded Branching Process
R/S	Rescaled Range (Range/Standar deviation)
DFA	Detrend Fluctuation Analysis

Notation

P_t^{bid}	Bid price at time t
P_t^{ask}	Ask price at time t

Preface

The analysis of high-frequency financial data is becoming an issue in understanding market behaviour. Millions of quotes are traded everyday and the advances in technology such as data storage and electronic trading systems implementation have made high-frequency financial data accessible to academicals and commercial users. High-frequency data (HFD) in finance refers to prices that are recorded several times a day. More precisely, HFD dataset contains quotes for each transaction made during the day. High-frequency financial data present an advantage compared to daily data. The additional intraday prices gives more details on how the price reacts to information and so a better examination of the source and the volatility of the return can be made. However, it also presents a great challenge to econometric modeling and statistical analysis for various reasons. Periodic fluctuation exhibited in the volatility is one of those reasons for which our research will be focused on.

Throughout the thesis, we describe how to remove the daily and weekly seasonality exhibited in high frequency financial data. We recall the subordinated model of Dacorogna [3] $X(t) = \tilde{X}(\theta(t))$ for high frequency financial data, where $\tilde{X}$ has some scaling properties and has stationary increments, and θ an increasing deterministic function. Assuming the logarithmic price of a given asset being of the form $\tilde{X}(\theta(t))$, we present three methods to estimate the time change function θ. An estimate of the function θ will describe the daily and weekly periodic fluctuation exhibited in the volatility.

The thesis is divided into three parts. In the first part, we introduce the mathematical background on self-similar stochastic processes. In the second part, we present three methods to estimate the deterministic time-change function θ and in the last part, we apply the time-changed process as a model for high frequency financial data.

Part I

In the first Chapter, an introduction of self-similar processes is given. We study especially the case where such processes has stationary increments. We introduce the Hurst index H, which in a certain way represents the roughness of the process in some case; when the process has finite second moment, it is also used to detect long/short range dependency of a time series. Several methods to estimate this index are presented, such as the R/S Analysis[4], the Variance plot[5], the Wavelet Analysis[6], the Crossing Tree method[7].

We introduce the α-stable distribution. We present some methods to estimate the parameters $(\alpha, \beta, \sigma, \mu)$ of the α-stable distribution, such as McCulloch [8], Koutrouvélis [9], Paulson, Holcomb and Leitch [10] and the maximum likelihood method. We then compare their performances.
Finally, we limit our study to the family of self-similar processes with α-stable distribution [11] and give some examples such as the Brownian motion ($H = 1/\alpha = 1/2$) or more generally the fractional Brownian motion (fBm) ($H \in (0,1), \alpha = 2$ value at which the distribution of the process is Gaussian), also the α-stable Lévy process ($H = 1/\alpha$), called Lévy stable motion (Lsm). We extend to the case $\alpha \neq 1/H$ by providing some examples.
In the second Chapter, we introduce briefly, some mathematical tools describing the roughness of a given function sample path. We present some notion on the Hausdorff dimension [12] and the Hölder exponent. We present as well some notion on the local structure of a random process, from which the class of locally self-similar processes will be introduced. Obviously a self-similar process is locally asymptotically self-similar.

Part II

In the third Chapter, we consider processes of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is a self-similar process with stationary increments and θ is a deterministic subordinator. We study the effect of the time change function on the roughness of the stochastic process sample path. Namely we look at the Hölder exponent and the local self-similarity of the process. Then, after describing how to simulate time-changed self-similar processes, we test the self-similarity index estimator's performances described in the first Chapter on the time-changed process $X(t) = \tilde{X}(\theta(t))$. We end the Chapter by presenting the appropriate estimator for time-changed processes.
In the fourth Chapter, we develop two time change estimators: the $1/H$-variation and log-variation. These two methods are based on the path variation of the process. We show that these two estimators are consistent under some assumption of X. More specifically, we assume $\tilde{X}$ as being locally self-similar and having stationary and ergodic increments. A statistical analysis and a numerical analysis of these estimators will be carried out. In particular, we test the performances of these two estimators on some well known self-similar processes.
In the fifth Chapter, we introduce a new estimator for the time change function. This last will be based on counting the level crossing points of the process. However, the continuity of the process sample path is needed as well as the stationarity and the ergodicity of its increments. We show that counting the number of crossing points of a self-similar process at different level size, allow us to estimate the self-similarity index of the process. In this Chapter, some experimental study of the estimator will be performed on self-similar processes to test their performances.

Part III

In the sixth Chapter, we present an overview on high frequency financial data. The main characteristics of such data are described as well as their inconvenience when we study them. Since such data contain spurious observations, we present a simple algorithm used to clean the data. The main task of the filter is to remove aberrant observations.

In the seventh Chapter, we apply our time change estimator given in part two, to some high frequency financial data. In particular on some Foreign exchange rate indices and the FTSE100 future contract. Assuming that the time change function θ has periodic derivative such that $\theta'(t+\tau) = \theta'(t)$, where τ is the one week periodicity, we introduce Dacorogna's time change estimator, which uses the scaling law properties of the data and compares it to ours. The derivative of the time change function represents the activity of the market. The function θ describes the daily and weekly seasonality of the data.

In the eighth Chapter, we present a model for high frequency financial data, using a time-changed Heyde's FATGBM model [13]. We analyse the AUD/USD rate index and the FTSE100 future contract after removing the intraday seasonality. A statistical analysis will be performed on these data. For both dataset, we study the time duration, which is the time interval separating two ticks. We analyse the autocorrelation function of the log-prices return as well as its absolute and squared value. The distribution on the log-prices return will be fitted to three different distribution functions. After some statistical analysis of the data, we set the characteristic of these data as assumptions for a model. The model, that uses the FATGBM will be fitted and compared respectively to the AUD/USD rate index and the FTSE100 future contract to close the Chapter.

Part I

Mathematical background

Chapter 1

Introduction to self-similar processes

"Self-similar processes are stochastic processes that are invariant under suitable scaling of time and space..." [14]. In this Chapter, we present an introduction to self-similar processes. We start by defining them, then we describe the α-stable distribution and we finish by providing some examples of self-similar processes.

1.1 Self-similar processes

We introduce some definitions. We suppose a stochastic process $X = \{X(t)\}_{t\in \mathrm{T}}$, a collection of random variables $X(t)$, defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $\mathrm{T} \subseteq \mathbb{R}$, called the index set of the process X.

Definition 1.1.1. *The process X ...*

(i) ...has independent increments if for any $t_1, \ldots, t_n \in \mathrm{T}$ such that $t_1 < \cdots < t_n$, $X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent.

(ii) ...is stochastically continuous at t, if for any $\varepsilon > 0$,

$$\lim_{s\to 0} \mathbb{P}\left(|X(t+s) - X(t)| > \varepsilon\right) = 0$$

(iii) ...has stationary increments if

$$\{X(t+h) - X(h)\}_{t\in \mathrm{T}} \overset{(fdd)}{=} \{X(t) - X(0)\}_{t\in \mathrm{T}} \text{ for all } h \in \mathrm{T}$$

(iv) ...is trivial (or degenerate) if $X(t)$ is a constant for every $t \in \mathrm{T}$.

1.1.1 Definition

Definition 1.1.2. *A stochastic process X is said to be self-similar if for any $a > 0$, there exist $b > 0$ such that*

$$\{X(at)\}_{t\in \mathrm{T}} \overset{(fdd)}{=} \{bX(t)\}_{t\in \mathrm{T}} \tag{1.1}$$

In other words, a self-similar process appears to be the same in statistical terms at any temporal scale. In Definition 1.1.2, Lamperti [15] shows under some assumption of X that it exists a relation between a and b. This last relation is given in Theorem 1.1.3, when X is a non-trivial self-similar process. Note that in [15], Lamperti used the term *semi-stable* instead of *self-similar*.

Theorem 1.1.3. *Let X be a non-degenerate and stochastically continuous self-similar process, then there exists a unique $H \geq 0$, such that b in (1.1) can be expressed as $b = a^H$.*

Proof. see proof in [14] page 2. □

Note that H cannot be also strictly negative. Indeed, let X be a process as defined in Theorem 1.1.3; suppose that $H < 0$, then the relation (1.1) becomes

$$\{X(at)\}_{t\in \mathrm{T}} \overset{(fdd)}{=} \left\{\frac{1}{a^{|H|}}X(t)\right\}_{t\in \mathrm{T}}$$

one has for $\varepsilon \geq 0$ and $0 < a < 1$

$$\mathbb{P}\left(|X(t) - X(0)| > \varepsilon\right) = \mathbb{P}\left(a^{|H|}|X(at) - X(0)| > \varepsilon\right) \leq \mathbb{P}\left(|X(at) - X(0)| > \varepsilon\right)$$

since the process X is stochastically continuous, one gets

$$\mathbb{P}\left(|X(at) - X(0)| > \varepsilon\right) \to 0 \text{ as } a \to 0^+$$

hence for each $t \in \mathrm{T}$, $X(t) \overset{(a.s.)}{=} X(0)$. Moreover, $X(0) \overset{(d)}{=} a^{|H|}X(a0) \overset{(a.s.)}{\to} 0$, as $a \to +\infty$. One deduces for $H < 0$, $X(t) \overset{(a.s.)}{=} 0$, for any $t \in \mathrm{T}$.
For $H = 0$, with Theorem 1.1.3, the relation (1.1) becomes $X(at) \overset{(d)}{=} X(t)$. Moreover, one has $X(at) - X(0) \overset{(d)}{=} X(t) - X(0)$, which yields to

$$\mathbb{P}\left(|X(t) - X(0)| > \varepsilon\right) = \mathbb{P}\left(|X(at) - X(0)| > \varepsilon\right) \text{ for some } \varepsilon > 0$$

X is continuous in probability, so as $a \to 0^+$,

$$\mathbb{P}\left(|X(t) - X(0)| > \varepsilon\right) = \lim_{a\to 0^+} \mathbb{P}\left(|X(at) - X(0)| > \varepsilon\right) = 0$$

Hence $\forall t \in \mathrm{T}$, $X(t) \overset{(a.s.)}{=} X(0)$.

One can rewrite the following definition for non-degenerate and stochastically continuous self-similar processes.

Definition 1.1.4. *Let X be a non-degenerate and stochastically continuous process. Then the process is said to be H-self-similar if it exists a unique $H > 0$, such that*

$$\forall a > 0,\ \{X(at)\}_{t\in \mathrm{T}} \overset{(fdd)}{=} \{a^H X(t)\}_{t\in \mathrm{T}}$$

H is called the self-similarity index, or also the Hurst index[1].

A direct consequence of self-similarity is that $X(0) \overset{(a.s.)}{=} 0$ (Proposition 1.1.5). Furthermore a self-similar process is never stationary, if such process $X(t)$ exists, then one has for $a > 0$ (X is non degenerate)

$$X(t) = X(at + (1-a)t) \overset{(d)}{=} X(at) \overset{(d)}{=} a^H X(t)$$

and for t fixed when $a \to +\infty$, $X(t) \to +\infty$ which is impossible.
For conciseness, we write H-ss for a H-self-similar process and H-sssi for H-ss process with stationary increments.

Proposition 1.1.5. *X is H-ss , with $H > 0$, then $X(0) \overset{(a.s.)}{=} 0$*

Proof. By self-similarity we have for $a > 0$

$$X(0) \overset{(d)}{=} X(a0) \overset{(d)}{=} a^H X(0)$$

Tending a to 0, one gets $X(0) \overset{(a.s.)}{=} 0$. □

1.1.2 Self-similar process with stationary increments

For applications, the stationarity of a process is important. As we saw, H-ss processes can not be stationary, however, it can have stationary increments. Several studies on self-similar processes with stationary increments exist in the literature ([16], [17], [18]). A consequence of adding this last property is the following Theorem.

Theorem 1.1.6. *Let X be a non-trivial H-sssi process. We suppose that $\mathbb{E}\left[X(1)^2\right] < \infty$. Then*

$$\forall (t,s) \in \mathbb{R}^2,\ \mathbb{E}[X(t)X(s)] = \frac{\mathbb{E}[X(1)^2]}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$$

Proof.

$$\begin{aligned}\mathbb{E}\left[X(t)X(s)\right] &= \mathbb{E}\left[\frac{1}{2}(X^2(t) + X^2(s) - (X(t) - X(s))^2)\right] \\ &= \frac{1}{2}(\mathbb{E}\left[X^2(t)\right] + \mathbb{E}\left[X^2(s)\right] - \mathbb{E}\left[(X(t) - X(s))^2\right])\end{aligned}$$

[1] In fact, for self-similar processes, the Hurst exponent and the self-similar index coincide.

Applying the stationarity of the increments and then the self-similar property, one gets

$$\begin{aligned}\mathbb{E}[X(t)X(s)] &= \frac{1}{2}(\mathbb{E}\left[X^2(t)\right] + \mathbb{E}\left[X^2(s)\right] - \mathbb{E}\left[X^2(t-s)\right]) \\ &= \frac{1}{2}(|t|^{2H}\mathbb{E}\left[X^2(1)\right] + |s|^{2H}\mathbb{E}\left[X^2(1)\right] - |t-s|^{2H}\mathbb{E}\left[X^2(1)\right]) \\ &= \frac{\mathbb{E}\left[X^2(1)\right]}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})\end{aligned}$$

□

Theorem 1.1.7. *Let X be a non-degenerate H-sssi process. Then*

(i) If $\mathbb{E}[|X(1)|^\gamma] < \infty$ for some $0 < \gamma < 1$, then $H < 1/\gamma$

(ii) If $\mathbb{E}[|X(1)|] < \infty$, then $H \leq 1$

(iii) If $\mathbb{E}[|X(1)|] < \infty$ and $H \in (0,1)$, then $\forall t \in \mathbb{R}$, $\mathbb{E}[X(t)] = 0$

(iv) If $\mathbb{E}[|X(1)|] < \infty$ and $H = 1$, then $\forall t \in \mathbb{R}$, $X(t) \overset{(a.s.)}{=} tX(0)$

Proof. see proof in [14] page 20-21. □

So far, we know that if a non-degenerate and stochastically continuous process X is H-ss and $H \leq 0$, then for any $t \in \mathrm{T}$, $X(t) \overset{(a.s.)}{=} 0$. Moreover, if X has stationary increments, $\mathbb{E}[|X(1)|] < \infty$ and for $H \in (0,1)$ then $\mathbb{E}[X(t)] = 0$. When $H = 1$, then $X(t) \overset{(a.s.)}{=} tX(0)$. The Theorem 1.1.8 answers the case $H > 1$.

Theorem 1.1.8. *Let X be a non-degenerate H-sssi process with $H > 1$, then*

$$\mathbb{E}[|X(1)|^{1/H}] = \infty$$

Proof. Easily shown using *(i)* of Theorem 1.1.7. □

1.1.3 Long-Range dependance

Long range dependence was studied in particulary by Hurst. He studies the level of the Nile and problems related to the water storage, known for alternating between long periods of dryness followed by long periods of flooding. This effect was called by Mandelbrot [19] "Joseph effect" referring to the seven years of abundance followed by seven years of famine described by a passage of the Bible.

Definition 1.1.9. *A stationary process $\{Y(n)\}_{n\in\mathbb{N}}$, with its autocorrelation function $r(n)$, is said to be long-range dependent, or with long memory, or slowly decaying correlation, if*

$$\sum_{n=0}^{+\infty} |r(n)| = +\infty$$

Proposition 1.1.10. *Let X be a non-trivial H-sssi process with $\mathbb{E}[X(1)^2] < \infty$, and let us define the autocorrelation of its increments*

$$\forall n \in \mathbb{N}, \quad r(n) = \mathbb{E}\left[(X(1) - X(0))(X(n+1) - X(n))\right]$$

Then

- $\forall n \in \mathbb{N}, \quad r(n) \underset{n\to\infty}{\sim} H(2H-1)n^{2H-2}\mathbb{E}\left[X_1^2\right], \; if \; H \neq \frac{1}{2}$

- $\forall n \in \mathbb{N}^*, \quad r(n) = 0, \; if \; H = \frac{1}{2}$

Proof. With $X(0) \overset{(a.s.)}{=} 0, \quad \forall n \in \mathbb{N}$

$$r(n) = \mathbb{E}\left[X(1)X(n+1)\right] - \mathbb{E}\left[X(1)X(n)\right] = \frac{1}{2}\Big((n+1)^{2H} - 2n^{2H} + (n-1)^{2H}\Big)\mathbb{E}\left[X^2(1)\right]$$

For $H = \frac{1}{2}$, one has $r(n) = 0$. Otherwise, $\forall n \in \mathbb{N}$,

$$r(n) = \frac{1}{2}\left(\left(1+\frac{1}{n}\right)^{2H} - 2 + \left(1-\frac{1}{n}\right)^{2H}\right)n^{2H}\mathbb{E}\left[X^2(1)\right]$$

Taylor expansion yields to

$$\begin{cases} \left(1+\frac{1}{n}\right)^{2H} = 1 + 2H\frac{1}{n} + \frac{2H(2H-1)}{2!}\left(\frac{1}{n}\right)^2 + \underset{n\to+\infty}{o}\left(\frac{1}{n}\right)^2 \\ \left(1-\frac{1}{n}\right)^{2H} = 1 - 2H\frac{1}{n} + \frac{2H(2H-1)}{2!}\left(\frac{1}{n}\right)^2 + \underset{n\to+\infty}{o}\left(\frac{1}{n}\right)^2 \end{cases}$$

For $H \neq \frac{1}{2}$, $r(n) \underset{n\to\infty}{\sim} H(2H-1)n^{2H-2}\mathbb{E}\left[X(1)^2\right]$. □

The following conclusion can be set

(i) If $0 < H < \frac{1}{2}$, $\sum_{n=0}^{\infty} |r(n)| < \infty$ $\Rightarrow$ Short-Range dependent increments

(ii) If $H = \frac{1}{2}$, $\forall n \in \mathbb{N}^*, |r(n)| = 0$ $\Rightarrow$ Uncorrelated increments

(iii) If $\frac{1}{2} < H < 1$, $\sum_{n=0}^{\infty} |r(n)| = \infty$ $\Rightarrow$ Long-Range dependent increments

1.1.4 Self-similarity index estimation

Many methods can be found in the literature thatestimate the Hurst parameter. As mentioned in the previous Section, Hurst [20] was the first to analyse the long range dependence of a time series, and therefore he introduced the R/S statistic [4]. Since many authors found other methods to estimate H: Variance of Residuals [5] and [21], Wavelet analysis [6], Crossing Tree [7] and [22], Whittle estimator [23], Higuchi's method [24]. We present the first four cited methods.

Rescaled range analysis

The rescaled range analysis was developed by Hurst. Let X a H-sssi process, we define the range of the deviation of the process as follows

$$R(n) = \left[\max_{0 \le i \le n} \left(X(i) - \frac{i}{n} X(n) \right) - \min_{0 \le i \le n} \left(X(i) - \frac{i}{n} X(n) \right) \right]$$

which will be rescaled by its standard deviation.

$$S(n) = \sqrt{\frac{1}{n} \sum_{1 \le i \le n} (X(i) - X(i-1))^2 - \left(\frac{1}{n}\right)^2 X^2(n)}$$

The rescaled adjusted range is then

$$R/S(n) = \frac{R(n)}{S(n)}$$

We have the following scaling law relation (see [25])

$$\mathbb{E}[R/S(n)] \simeq C_H n^H \text{ as } n \to \infty \tag{1.2}$$

where C_H is a positive constant independent of n.

To estimate H using the R/S statistic for a given time series $\{X(i)\}_{1 \le i \le N}$ with N observations, one subdivides the series into K block k_i of size N/K and then computes the Rescaled range for each block k_i (that we note $R^{k_i}(n)/S^{k_i}(n)$) and for each lag n with the starting points, $k_i = iN/K + 1$, $i = 1, 2, ..,$ such that $k_i + n \le N$. When n is small, ones gets K values of $R^{k_i}(n)/S^{k_i}(n)$ and when n is close to N one gets one value. Let $R/S(n) = \frac{1}{K} \sum_{1 \le i \le K} R^{k_i}(n)/S^{k_i}(n)$, plotting $\log\left(R/S(n)\right)$ versus $\log n$, H can then be estimated by computing the slope of the fitting line of the points plotted (Figure 1.1 using a fractional Brownian motion for $H = 0.6$ see Section 1.3.2).

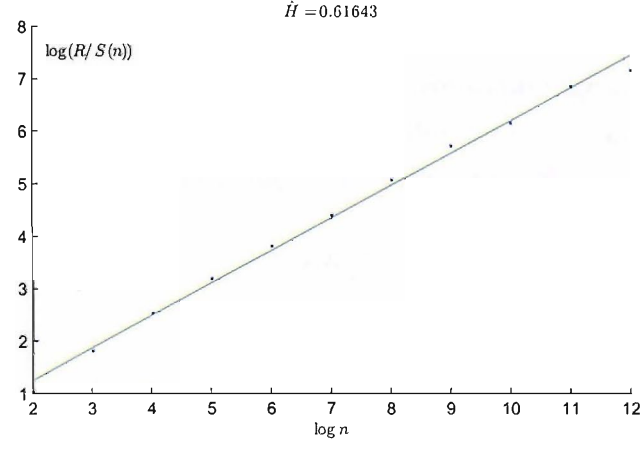

Figure 1.1: $\log(R/S)$ function of $\log(n)$ and its fitted line.

Variance of Residuals

The Detrended Fluctuation Analysis (DFA) was introduced by Peng and all [5]. It is also called Variance of Residuals [21]. To apply this method to a time series $\{X(i)\}_{1\leq i\leq N}$, we first divide the time series into blocks of size $m < N$, then in each block we detrend the sample by subtracting the local trend applied to the i^{th} blocks $y_i(t) = a_i t + b_i$, fitted by a least squares method.

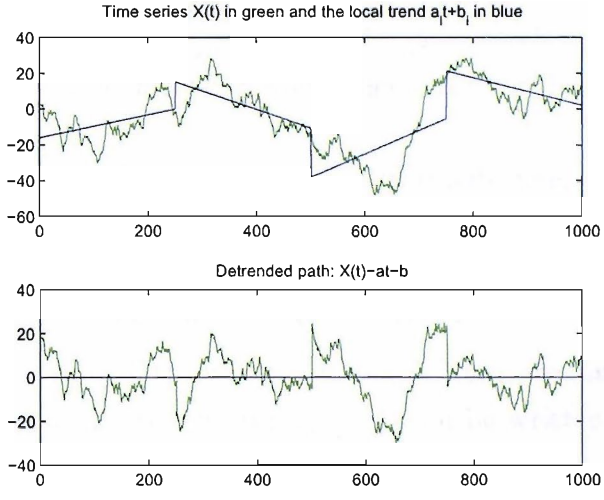

Figure 1.2: Detrend fluctuation analysis.

For a given block i of size m, one has the following characteristic size of fluctuation

$$F_i(m) = \frac{1}{m}\sum_{t=1}^{m}(X(t) - a_i t - b_i)^2$$

and we deduce the average of the characteristic size of fluctuation by

$$F(m) = \frac{m}{N}\sum_{1\leq i\leq N/m} F_i(m)$$

This computation is repeated over different block sizes to provide a relationship between $F(m)$ and m. Taqqu proves the following relationship for a fractional Brownian motion in [21] and gives a sketch proof for some other cases (see [26]).

$$\mathbb{E}[F(m)] \sim C_H m^{2H} \text{ as } m \to \infty$$

where

$$C_H = \left(\frac{2}{2H+1} + \frac{1}{H+2} - \frac{2}{H+1}\right)$$

The slope of the linear regression of $\log F(m)$ against $\log(m)$, determines the scaling exponent H.

Wavelet analysis

The idea behind the wavelet analysis, is to expand a signal $\{X(t)\}_{t\in\mathrm{T}}$ for a given resolution depth J, as Signal=approximation (low pass)+ details (high pass):

$$X(t) = \sum_k a_X(J,k)\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_k d_X(j,k)\psi_{j,k}(t). \tag{1.3}$$

With $\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t-k)$, $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t-k)$, $(j,k)\in\mathbb{Z}^2$ where ϕ_0 is the scaling function and ψ_0 the mother wavelet, derived from the ϕ_0 via The MultiResolution Analysis see [27].

The coefficients $a_X(j,k)$ and $d_X(j,k)$ are defined through inner products respectively with $\phi_{j,k}$ and $\psi_{j,k}$

$$a_X(j,k) =< X(t), \phi_{j,k}(t) >$$
$$d_X(j,k) =< X(t), \psi_{j,k}(t) >$$

To understand the application of the wavelet decomposition, we consider a H-ss process X. The variance of its wavelet coefficients $d_X(j,k)$ can be written as follows (see [6])

$$\mathbb{E}[d_X(j,k)^2] = 2^{j(2H+1)}\mathbb{E}[d_X(0,k)^2] \tag{1.4}$$

Moreover if the process has stationary increments, (1.4) is reduced to

$$\mathbb{E}[d_X(j,k)^2] = 2^{j(2H+1)}C(H,\psi_0)\sigma^2,\ \forall k$$

with

$$C(H,\psi_0) = \int\int |t|^{2H}\psi_0(u)\psi_0(u-t)dudt \text{ and } \sigma^2 = \mathbb{E}[X(1)^2]$$

We deduce an estimator of H by estimating the slope α of the regression line of $\log_2(\mathbb{E}[d_X(j,k)^2])$ against j, with $j\in\{j_1,\ldots,j_2\}$, where the interval $\{j_1,\ldots,j_2\}$ represents the resolutions used. One gets $H = (\alpha-1)/2$.

In practice, one estimates $\mathbb{E}[d_X(j,k)^2]$ by computing the empirical moment of order 2,

$$\mu_j = \mathbb{E}[d_X(j,k)^2] \simeq \frac{1}{Nj}\sum_{k=1}^{Nj} d_X(j,k)^2 = \hat{\mu}_j$$

where N_0 is the data size, and $Nj = 2^{-j}N_0$. The fact that $\log(\mathbb{E}[(.)]) \neq \mathbb{E}[\log(.)]$, Abry and Veitch [6], add a small corrective ξ_j (the estimation error), such that

$$\xi_j = \log(\mu_j) - j\alpha - \log_2(C(H,\psi_0)\sigma^2),\ \forall j\in\{j_1,\ldots,j_2\} \tag{1.5}$$

When the increments of the process X are Gaussian distributed, then the mean and

the variance of ξ_j are given by (see [6])

$$\mathbb{E}[\xi_j] = \frac{\Psi(N_j/2)}{\log(2)} - \log_2(N_j/2) \text{ and } \mathbb{V}ar\,(\xi_j) = \frac{\zeta(2, N_j/2)}{\log^2(2)}$$

where $\Psi(z) = \Gamma'(z)/\Gamma(z)$, $\Gamma(z)$ and $\zeta(z, \nu)$ are respectively the Gamma and the Riemann Zeta function.
The estimator $\hat{\alpha}$ of α is the slope of a linear regression of $\log_2(\mu_j) - \mathbb{E}[\xi_j]$ on j, weighted by $\mathbb{V}ar\,(\xi_j)$, for $j \in \{j_1, \ldots, j_2\}$.

Crossing Tree method

The Crossing Tree method was introduced by O.D. Jones and Y. Shen in [7]. This method consists on studying the process on space scale rather than temporal scale. Let X be a H-ss with continuous sample path, and δ a fixed spacial base scale. Let T_k^n be the k^{th} crossing time of size $\delta 2^n$ with

$$T_0^n = 0 \text{ and } T_{k+1}^n = \inf\{t > T_k^n : X(t) \in \delta 2^n, X(t) \neq X(T_k^n)\}$$

In Figure 1.3, the crossing levels are plotted as well as its crossing tree. An explanation on how the crossing are built can be found in [2]. Also, in appendix F.1 a graph on how the crossing tree is formed is described.

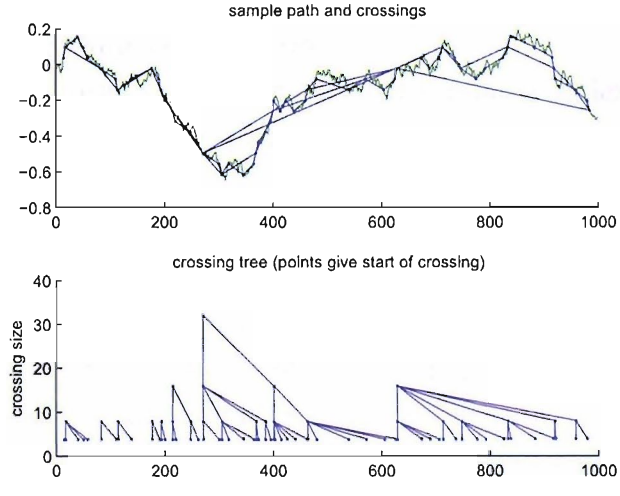

Figure 1.3: Formation of the crossing tree from a sample path. Upper panel shows the path, lower panel shows the structure of the tree which is in fact the crossings for each level linked by lines (Matlab source on [1] to realise this plot).

Let Z_k^n be the number of subcrossings of size $\delta 2^{n-1}$ that make up the k^{th} crossing of size $\delta 2^n$. For different n, $\{Z_k^n\}_k$ are identically distributed, and assumed that for fixed n $Z_1^n, Z_2^n \ldots$ are stationary and ergodic. Let $N(n)$ be the total number of crossings of size $\delta 2^n$. The relation between both quantities must satisfy $N(n) \geq \sum_{k=1}^{N(n+1)} Z_{k+1}^n$. Let $\mu = \mathbb{E}[Z_k^n]$, if we scale the space by 2^k, then we must scale the time by μ^k to get a

crossing of the same expected length, this implies that if X is self-similar, then

$$\{X(t)\}_{t\geq 0} \overset{(fdd)}{=} \{2^{-k}X(\mu^k t)\}_{t\geq 0} = \{(\mu^k)^{-\log(2)/\log(\mu)}X(\mu^k t)\}_{t\geq 0}$$

So from the definition of a self-similar process, we deduce $H = \log(2)/\log(\mu)$.
In practice with the previous assumption, one can use (1.6) as a consistent estimator for μ (see [7]),

$$\hat{\mu}_n = \sum_{k=1}^{N(n)} Z_k^n / N(n) \tag{1.6}$$

the estimator for the Hurst index at scale $\delta 2^n$ is then given by

$$\hat{H}_n = \frac{\log(2)}{\log(\hat{\mu}_n)}$$

If the propriety of self-similarity holds over different scales i.e. $\delta 2^m$ to $\delta 2^n$ [22], then we can combine them to get a more consistent estimator.

$$\hat{H}_{m,n} = \frac{\log(2)}{\log(\hat{\mu}_{m,n})}$$

where $\hat{\mu}_{m,n} = \sum_{m\leq i\leq n} N(i)\hat{\mu}_i / \sum_{m\leq i\leq n} N(i)$

Comments on the crossing tree method

The crossing tree method estimates the global self-similarity index by using crossings level of the process. The crossings points computed on discrete time series are an approximation of the real crossing points of the continuous time process and so the error of this approximation may falsify the estimation of the global index of self-similarity of the process.
Burq and Jones in [28] present a way to simulate a Brownian on a regular space grid. We simulate two standard Brownian motions $\{B_1(t_j)\}_{j=0,\ldots,n}$, where $0 = t_0 < \cdots < t_j < \cdots < t_n = 1$ are the crossing times, and $\{B_2(j/n)\}_{j=0,\ldots,n}$ with $n = 500000$ observations (we omit the first observation). We compare the performance of the crossing tree on both Brownian motion. The index of self similarity estimate is shown in Figure 1.4.
The estimation of the Hurst index of a Brownian, simulated on a regular spacial grid at different levels, are illustrated on Figure 1.4 left panel. We note that values of $\hat{H}$ at different levels are around 0.5. For the regular spaced time Brownian motion, the right panel of Figure 1.4 shows that there is a kind of tail starting from $\hat{H} = 0.65$. A good estimate of H of the regular time spaced Brownian motion can be seen from crossing size of level 5, knowing that the crossing size is the standard deviation of the discrete time process. Note that we simulated 500000 observations for each Brownian motion so we obtained enough levels for a good estimate of H. That means that for time series of small length, the crossing tree is not performing well. For a good performance of the

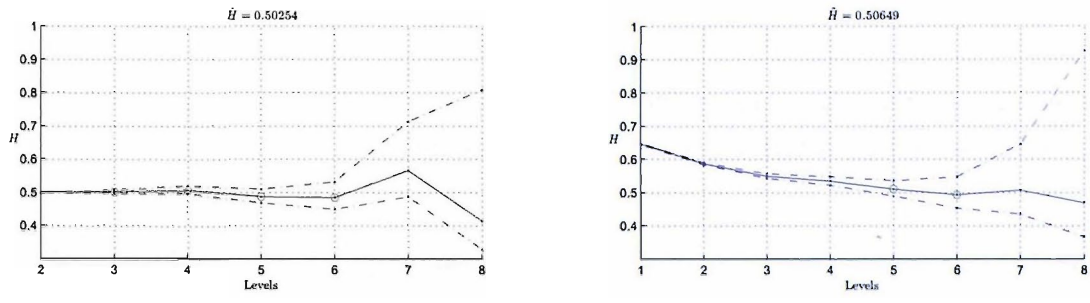

Figure 1.4: Hurst index estimation of two Brownian motions, simulated on regular space grid (left panel) and one regular space time (right panel).

crossing tree estimator, one need to estimate H from time series of a big sample. In finance, the size of a dataset of one year of high frequency financial data for example, may attain 10 millions of quotes. The crossing tree estimator in that case will be an appropriate estimator.

In Jones and Shen [7], the EBP (Embedded Branching Process) estimator is tested using fractional Brownian motion but not with any discontinuous processes. Accordingly, we tested its performance on simulated α-stable Lévy motion, for a variety of values of $H = 1/\alpha \in [0.5, 1)$. For each value of H we generated 20 series consisting of 10000 evenly spaced observations. In Table 1.1 we give the mean $\pm$ twice the standard error of the EBP estimates for each value of H. In each case, the EBP estimator gave a reasonable estimate of H.

H	0.5	0.6	0.7	0.8	0.9
$\hat{H}$	0.537 ± 0.076	0.629 ± 0.110	0.696 ± 0.100	0.795 ± 0.140	0.897 ± 0.134

Table 1.1: EBP estimates of the Hurst index of a Lévy stable motion (using the crossing tree). Each estimate is given plus or minus twice the standard error (from a Monte-Carlo experiment).

Comments on the efficiency of the estimator

Several studies can be found in the literature about the efficiency of the estimator. We refer to the empirical analysis of Taqqu and Teverovsky in [21] and also in [26], on the efficiency of the estimators. The DFA and the wavelets method are shown to be efficient estimators while the R/S analysis is shown to be less efficient. The crossing tree method is not described in [26].

1.2 α-stable distribution

In this Section, we suppose the random variables X, $X_1, \ldots, X_n$ defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

1.2.1 Introduction

Before we define the α-stable law, we present a more general law: the infinitely divisible law. It is from this law that we define the characteristics of the α-stable (see [29] and [30]).

Definition 1.2.1. *A random variable X is infinitely divisible, if for each $n \in \mathbb{N}$, there is an i.i.d. sequence $X_1, \cdots, X_n$ such that*

$$X_1 + \cdots + X_n \overset{(d)}{=} X$$

Definition 1.2.2. *A random variable X has a stable distribution if and only if, $\forall n \in \mathbb{N}$, and $X_1, \cdots, X_n$ i.i.d random variables with the same law as X, there exist a_n positive and a real b_n such that*

$$X_1 + \cdots + X_n \overset{(d)}{=} a_n X + b_n$$

when $b_n = 0$, X has a strictly stable distribution.

Remark 1.2.3. *It can be shown that it exists $\alpha \in (0, 2]$ such that $a_n = n^{\frac{1}{\alpha}}$. (see [31]). Moreover, we note from these definitions that a stable random variable implies that it is infinitely divisible. The opposite case is false.*

Theorem 1.2.4. *(Levy-Khinchin)[29] If X has a stable distribution, then its characteristic function is*

$$\forall t \in \mathbb{R},\ \varphi_X(t) = \mathbb{E}\left[e^{itX}\right] = e^{\psi(t)}$$

Where

$$\psi(t) = \begin{cases} i\mu t - \sigma^\alpha |t|^\alpha \left(1 - i\beta sign(t) \tan \dfrac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1 \\ i\mu t - \sigma |t| \left(1 + i\beta \dfrac{2}{\pi} sign(t) \log |t|\right) & \text{if } \alpha = 1 \end{cases}$$

$$sign(t) = \begin{cases} -1 & \text{if } t < 0 \\ 0 & \text{if } t = 0 \\ 1 & \text{if } t > 0 \end{cases}$$

and

- $\alpha \in (0, 2]$ *is called the* characteristic exponent. *It measures the thickness of the tails of an α-stable distribution. The smaller the value of α, the higher the probability in the distribution tails (see 1.2.6).*
- $\beta \in [-1, 1]$ *is a* symmetry parameter, *also called* skew parameter.
- $\sigma > 0$ *is a* scale parameter.
- $\mu \in \mathbb{R}$ *is a* location parameter.

We denote an α-stable random variable X: $X \sim S_\alpha(\sigma, \beta, \mu)$

There are two other expressions, which represent the characteristic function of a stable distribution. This are given by Zolotarev [32]:

- The Zolotarev's (M) parameterisation $X \sim S^0_\alpha(\sigma, \beta, \mu_0)$:

$$\forall t \in \mathbb{R},\ \varphi_0(t) = e^{\psi_0(t)}$$

Where

$$\psi_0(t) = \begin{cases} \boldsymbol{i}\mu_0 t - \sigma^\alpha |t|^\alpha \Big(1 + \boldsymbol{i}\beta sign(t) \tan\frac{\pi\alpha}{2}(\sigma^{1-\alpha}|t|^{1-\alpha} - 1)\Big) & \text{if } \alpha \neq 1 \\ \boldsymbol{i}\mu_0 t - \sigma|t|\Big(1 + \boldsymbol{i}\beta\frac{2}{\pi} sign(t)\log\sigma|t|\Big) & \text{if } \alpha = 1 \end{cases}$$

- The Zolotarev's (B) parameterisation $X \sim S^2_\alpha(\sigma_2, \beta_2, \mu_2)$:

$$\forall t \in \mathbb{R},\ \varphi_2(t) = e^{\psi_2(t)}$$

Where

$$\psi_2(t) = \begin{cases} \boldsymbol{i}\mu_2 t - \sigma_2^\alpha |t|^\alpha \exp\Big(-\boldsymbol{i}\frac{\pi}{2}\beta_2 sign(t)(1 - |1-\alpha|)\Big) & \text{if } \alpha \neq 1 \\ \boldsymbol{i}\mu_2 t - \sigma_2|t|\Big(\frac{\pi}{2} + \boldsymbol{i}\beta_2 sign(t)\log|t|\Big) & \text{if } \alpha = 1 \end{cases}$$

These two different parameterisations are linked to the standard parameterisation as follow

- Link between standard and (M) parameterisation

$$\mu = \begin{cases} \mu_0 - \beta\sigma\tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \mu_0 - \beta\sigma\frac{2}{\pi}\log\sigma & \text{if } \alpha = 1 \end{cases} \tag{1.7}$$

- Link between standard and (B) parameterisation

$$\begin{aligned} &\mu_2 = \mu,\ \tan\Big(\beta_2\frac{\pi(1-|1-\alpha|)}{2}\Big) = \beta\tan\frac{\pi\alpha}{2},\ \sigma_2 = \sigma\Big(1 + \beta^2\tan^2\frac{\pi\alpha}{2}\Big)^{\frac{1}{2\alpha}} && \text{if } \alpha \neq 1 \\ &\mu_2 = \mu,\ \beta_2 = \beta,\ \sigma_2 = \frac{2}{\pi}\sigma && \text{if } \alpha = 1 \end{aligned}$$

1.2.2 Properties of the α-Stable Distribution

Proposition 1.2.5. *Let X_1 and X_2 be two independent random variables and so that $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$, $i = 1, 2$. Then $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$ with*

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha},\ \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha},\ \mu = \mu_1 + \mu_2$$

Proof. see [11] page 10-11. □

When $\alpha < 2$, the tails are asymptotically equivalent to a Pareto law.

Proposition 1.2.6. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then*

$$\begin{cases} \lim\limits_{x \to +\infty} x^\alpha \mathbb{P}(X > x) = \sigma^\alpha C_\alpha \dfrac{1+\beta}{2} \\ \lim\limits_{x \to +\infty} x^\alpha \mathbb{P}(X < -x) = \sigma^\alpha C_\alpha \dfrac{1-\beta}{2} \end{cases}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1} = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)} & \alpha \neq 1 \\ \dfrac{2}{\pi} & \alpha = 1 \end{cases}$$

Proof. see [11] page 16-17. □

Let $X \sim S_2(\sigma, \beta, \mu)$, then for $p > 0$, the p^{th} moment is finite. Indeed for $\alpha = 2$, the characteristic function becomes $\phi_X(t) = e^{i\mu t - (\sigma t)^2}$, so $S_2(\sigma, \beta, \mu)$ is the normal distribution (see (1.9)), and all moments of random variables normal distributed are finite. For $\alpha \neq 2$ we have the next Proposition.

Proposition 1.2.7. *Let $X \sim S_\alpha(\sigma, \beta, \mu)$ with $0 < \alpha < 2$. Then*

$$\begin{cases} \mathbb{E}[|X|^p] < +\infty & 0 < p < \alpha \\ \mathbb{E}[|X|^p] = +\infty & p \geq \alpha \end{cases} \tag{1.8}$$

Proof. For a positive integrable random variable Y one has

$$\mathbb{E}[Y] = \int_0^{+\infty} \mathbb{P}(Y > u) du$$

By substituting $Y = |X|^p$ and $u = x^p$ we have

$$\mathbb{E}[|X|^p] = \int_0^{+\infty} p x^{p-1} \mathbb{P}(|X|^p > x^p) dx = p \int_0^{+\infty} x^{p-1} \mathbb{P}(|X| > x) dx$$

we study this integral at 0 and $+\infty$.

$$\lim_{x \to 0} x^{p-1} \frac{\mathbb{P}(|X| > x)}{x^{p-1}} = \lim_{x \to 0} \mathbb{P}(|X| > x) = \mathbb{P}(|X| \neq 0) = 1$$

hence

$$x^{p-1} \mathbb{P}(|X| > x) \underset{x \to 0}{\sim} x^{p-1}$$

and for $M \in (0, +\infty)$ and $p > 0$,

$$\int_0^M x^{p-1} dx < +\infty \Leftrightarrow \int_0^M x^{p-1} \mathbb{P}(|X| > x) dx < +\infty$$

At $+\infty$, Proposition 1.2.6 yields to

$$\lim_{x\to+\infty} x^{p-1}\frac{\mathbb{P}(|X|>x)}{x^{p-\alpha-1}} = \lim_{x\to+\infty} x^{\alpha}\mathbb{P}(|X|>x) = \sigma^{\alpha}C_{\alpha}$$

hence

$$x^{p-1}\mathbb{P}(|X|>x) \underset{x\to+\infty}{\sim} x^{p-\alpha-1}$$

and for $M \in (0,+\infty)$ and $p<\alpha$,

$$\int_M^{+\infty} x^{p-\alpha-1}dx < +\infty \Leftrightarrow \int_M^{+\infty} x^{p-1}\mathbb{P}(|X|>x)dx < +\infty$$

which implies for $p \in (0,\alpha)$, $\mathbb{E}[|X|^p] < +\infty$
for $p \geq \alpha$

$$\int_M^{+\infty} x^{p-\alpha-1}dx = +\infty \Leftrightarrow \int_M^{+\infty} x^{p-1}\mathbb{P}(|X|>x)dx = +\infty$$

which implies for $p \geq \alpha$, $\mathbb{E}[|X|^p] = +\infty$ □

From this property we deduce that for $\alpha < 2$, the variance of a stable random variable is infinite. For $\alpha \leq 1$, the mean becomes infinite. For $\alpha > 1$, the mean of a stable random variable X is μ. Indeed by computing the derivative of its characteristic function for $\alpha \neq 1$,

$$\varphi'_X(t) = \left(\mathbf{i}\mu - \text{sign}(t)\alpha\sigma^{\alpha}|t|^{\alpha-1}\left(1-\mathbf{i}\beta\text{sign}(t)\tan\frac{\pi\alpha}{2}\right)\right)e^{\mathbf{i}\mu t-\sigma^{\alpha}|t|^{\alpha}\left(1-\mathbf{i}\beta\text{sign}(t)\tan\frac{\pi\alpha}{2}\right)}$$

hence $\varphi'_X(0) = i\mu$ and so $\mathbb{E}[X] = \mu$

1.2.3 Density function

Except from the normal, Cauchy and Lévy distribution, there is no known explicit form of the α-stable density function.

- $S_2(\sigma,0,\mu)$ is the normal law

$$f(x) = \frac{1}{2\sigma\sqrt{\pi}}\exp\left(\frac{(x-\mu)^2}{4\sigma^2}\right) \tag{1.9}$$

- $S_1(\sigma,0,\mu)$ is the Cauchy law

$$f(x) = \frac{2\sigma}{\pi((x-\mu)^2+4\sigma^2)} \tag{1.10}$$

- $S_{\frac{1}{2}}(\sigma,1,\mu)$ is the Lévy law

$$f(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2}(x-\mu)^{-3/2}\exp\left(-\frac{\sigma}{2(x-\mu)}\right)\times 1_{(\mu,\infty)}(x) \tag{1.11}$$

However, an integral representation is given by Zolotarev in [32] as an expression of the density function using the inverse Fourier transform

$$f(x;\alpha,\beta,\sigma,\mu) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi_X(t)dt \tag{1.12}$$

One can use either the fast Fourier transform (FFT) to compute this density numerically or the direct integration method, proposed by Nolan [33], which consists of a numerical integration of Zolotarev's (M) parameterisation.
The density function is computed from (1.12) (see [33] for more details on the computation). One gets

$$f(x;\alpha,\beta,1,0) = \frac{1}{\pi}\int_0^{+\infty} \cos(h(x,t;\alpha,\beta))e^{-t^\alpha}dt$$

$$\text{where} \quad h(x,t;\alpha,\beta) = xt - \Im\Big(\log \mathbb{E}\big[e^{itY}\big]\Big) = \begin{cases} xt + \beta\tan\dfrac{\pi\alpha}{2}(t-t^\alpha) & \alpha \neq 1 \\ xt + \beta\dfrac{2}{\pi}\log t & \alpha = 1 \end{cases}$$

we have the density function of $S^0_\alpha(\sigma,\beta,\mu_0)$

$$f(x;\alpha,\beta,\sigma,\mu_0) = \begin{cases} \dfrac{1}{\sigma}f\left(\dfrac{x-\mu_0}{\sigma};\alpha,\beta,1,0\right) & \alpha \neq 1 \\ \dfrac{1}{\sigma}f\left(\dfrac{x-\mu_0}{\sigma};\alpha,\beta,1,0\right) & \alpha = 1 \end{cases}$$

Note that we can also compute the density function of $S_\alpha(\sigma,\beta,\mu)$, using its relation with the Zolotarev's (M) parameterisation given in (1.7).
We can see the consequence of the different parameterisation on the density function (see [34] for more detail), using a numerical computation for both densities (see [35]), with $\alpha \in [0.4, 2]$, $\beta = 1$, $\sigma = 1$ and $\mu = 0$, we obtain Figure 1.5. Note the discontinuity on $\alpha = 1$ for the standard parameterisation.

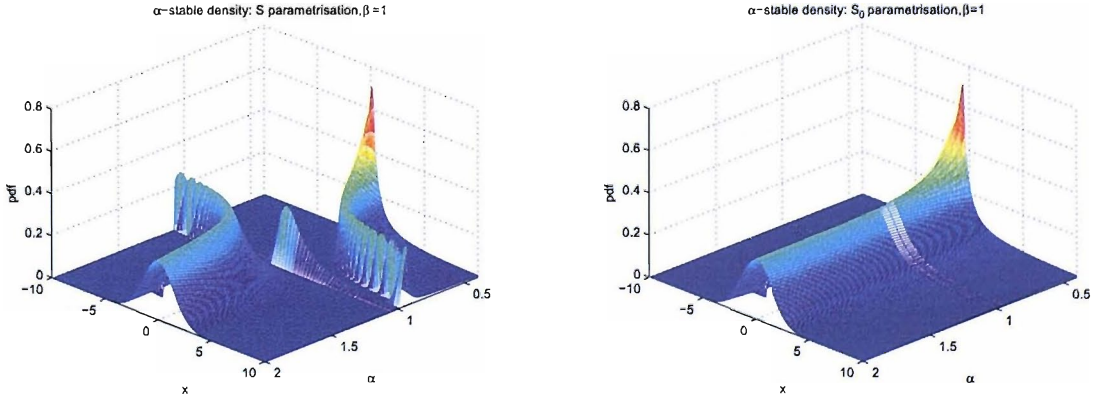

Figure 1.5: Density function of $S_\alpha(1,1,0)$ (left) and $S^0_\alpha(1,1,0)$ (right) with $\alpha \in [0.4, 2]$.

1.2.4 Simulation of α-stable random variable

We present a method to generate an α-stable random variable introduced by Chambers, Mallows, and Stuck [36], and corrected by Rafal Weron [37] and [38].
This method consists on simulating $X \sim S_\alpha(1, \beta, 0)$, with $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$, using the standard characteristic function.
First, we simulate two independent random variables U and W such that

- $U \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$
- $W \sim \mathcal{E}(1)$

and then X is obtained as follow

- If $\alpha \neq 1$:

$$X = \mathcal{S}_{\alpha,\beta} \frac{\sin(\alpha(U + \mathcal{B}_{\alpha,\beta}))}{\cos(U)^{\frac{1}{\alpha}}} \left(\frac{\cos(U - \alpha(U + \mathcal{B}_{\alpha,\beta})}{W} \right)^{\frac{1-\alpha}{\alpha}}$$

- If $\alpha = 1$:

$$X = \frac{2}{\pi} \left(\left(\frac{\pi}{2} + \beta U \right) \tan(U) - \beta \ln \left(\frac{\frac{\pi}{2} W \cos(U)}{\frac{\pi}{2} + \beta U} \right) \right)$$

Where:

- $\mathcal{B}_{\alpha,\beta} = \dfrac{\arctan\left(\beta \tan(\frac{\pi\alpha}{2})\right)}{\alpha}$
- $\mathcal{S}_{\alpha,\beta} = \left(1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{\frac{1}{2\alpha}}$

One can generate a standard α-stable R.V. $X \sim S_\alpha(\sigma, \beta, \mu)$, and also $X^{(0)} \sim S^0_\alpha(\sigma, \beta, \mu_0)$ by using the following equivalences; if $Y \sim S_\alpha(1, \beta, 0)$ then

1. from $Y \sim S_\alpha(1, \beta, 0)$ to $X \sim S_\alpha(\sigma, \beta, \mu)$

$$X = \begin{cases} \sigma Y + \mu & \text{if } \alpha \neq 1; \\ \sigma Y + \frac{2}{\pi}\beta\sigma \ln \sigma + \mu & \text{if } \alpha = 1. \end{cases}$$

2. from $Y \sim S_\alpha(1, \beta, 0)$ to $X^{(0)} \sim S^0_\alpha(\sigma, \beta, \mu_0)$

$$X^{(0)} = \begin{cases} \sigma Y - \sigma\beta \tan\frac{\pi\alpha}{2} + \mu_0 & \text{if } \alpha \neq 1; \\ \sigma Y + \mu_0 & \text{if } \alpha = 1. \end{cases}$$

1.2.5 Estimation of the parameters

Several stable distribution parameters estimators exist. Fama and Roll [39], [40] estimate the parameters using order statistics limited to $\beta = 0$ and $\alpha \in [1, 2]$. McCulloch [8] extended this method to $\beta \in [-1; 1]$ and $\alpha \in [0.6, 2]$. Arad [41] and Koutrouvélis [9] use the characteristic function, whereas Zolotarev [42] estimate α, β and σ with the method of moments for a known μ. DuMouchel (1971) developed an algorithm using the maximum likelihood. Press [43], Paulson, Holcomb and Leitch [10] use the fourier transform of the data. We introduce some of these methods, and compare their performances.

McCulloch method

This method is based on the distribution quantiles. However the restriction of this method is that it gives an estimation of $\alpha \in [0.6, 2]$. The author did not provide a precise reason of this restriction (see [8]). However this restriction is not an inconvenience for most applications.

Let U be a α-stable random variable with $\sigma = 1$ and $\mu = 0$ and $S = \sigma U + \mu$, we define Q_x and q_x the quantiles of distribution such that $\mathbb{P}(S < Q_x)$ and $\mathbb{P}(U < q_x)$. Since $\mathbb{P}(U < q_x) = \mathbb{P}(S < \sigma q_x + \mu)$, one has

$$Q_x = \sigma q_x + \mu \tag{1.13}$$

To estimate α and β, we first need to compute the following quantities defined by McCulloch

$$\begin{cases} \nu_\alpha = \dfrac{Q_{0.95} - Q_{0.05}}{Q_{0.75} - Q_{0.25}} \\ \nu_\beta = \dfrac{Q_{0.95} + Q_{0.05} - 2Q_{0.50}}{Q_{0.95} - Q_{0.05}} \end{cases}$$

From (1.13) the previous equalities are reduced to

$$\begin{cases} \nu_\alpha = \dfrac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}} \\ \nu_\beta = \dfrac{q_{0.95} + q_{0.05} - 2q_{0.50}}{q_{0.95} - q_{0.05}} \end{cases}$$

which implies that ν_α and ν_β depend only on α and β and by inversion α and β depend on ν_α and ν_β, one has

$$\begin{cases} \alpha = \Psi_1(\nu_\alpha, \nu_\beta) \\ \beta = \Psi_2(\nu_\alpha, \nu_\beta) \end{cases}$$

One reads Tables F.1 and F.2 given respectively Ψ_1 and Ψ_2 to find $\hat{\alpha}$ and $\hat{\beta}$ respectively

estimators of α and β. We refer to [8] for the tables.

$$\begin{cases} \hat{\alpha} = \Psi_1(\hat{\nu}_\alpha, \hat{\nu}_\beta) \\ \hat{\beta} = \Psi_2(\hat{\nu}_\alpha, \hat{\nu}_\beta) \end{cases}$$

The scale and location parameters σ and μ, are estimate in a similar way.
We define

$$\nu_c = \frac{Q_{0.75} - Q_{0.25}}{\sigma} = q_{0.75} - q_{0.25} = \Phi_3(\alpha, \beta)$$

and

$$\nu_\mu = \frac{\mu - Q_{0.50}}{\sigma} = -q_{0.50} = \Phi_4(\alpha, \beta)$$

we deduce an estimator of σ

$$\hat{\sigma} = \frac{\hat{Q}_{0.75} - \hat{Q}_{0.25}}{\Phi_3(\hat{\alpha}, \hat{\beta})}$$

and μ

$$\hat{\mu} = \hat{\sigma}\Phi_4(\hat{\alpha}, \hat{\beta}) + \hat{Q}_{0.50}$$

see Tables F.3 and F.4 for the function Φ_3 and Φ_4.

Koutrouvélis Method

This method, introduced by Koutrouvélis, uses the characteristic function (1.14) to estimate the α-stable distribution parameters. He uses a regression-type method that starts with an initial estimation of the parameters and proceeds iteratively.

$$\phi(t) = e^{i\mu t - |\sigma t|^\alpha \left(1 - i\beta \frac{|t|}{t} w(t,\alpha)\right)}, \text{ where } w(t,\alpha) = \begin{cases} \tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi}\log|t| & \text{if } \alpha = 1 \end{cases} \tag{1.14}$$

one has

$$\log\Big(-\log|\phi(t)|^2\Big) = \log(2\sigma^\alpha) + \alpha\log t \tag{1.15}$$

Using a regression for (1.15), one gets $\hat{\alpha}$ and $\hat{\sigma}$, estimator of α and σ.

$$\log\Big(-\log|\hat{\phi}(t_k)|^2\Big) = \log(2\hat{\sigma}^{\hat{\alpha}}) + \hat{\alpha}\log t_k \tag{1.16}$$

where for a sample of stable random variables $\{S_i\}_{i=\{1,\dots,n\}}$,

$$\hat{\phi}(t) = \frac{1}{n}\sum_{i=1}^{n} \exp(iS_i t)$$

is an estimator of $\phi(t)$, and for $K \in \mathbb{N}^*$, $\{t_k\}_{\{k=1,\dots,K\}}$ is an appropriate set of real numbers depending on α.

To estimate β and μ, we take the real and imaginary part of $\phi(t)$ for $\alpha \neq 1$

$$\begin{cases} \Re\big(\phi(t)\big) = e^{-|\sigma t|^\alpha} \cos\left(\mu t + |\sigma t|^\alpha \beta \frac{|t|}{t} \tan\frac{\pi\alpha}{2}\right) \\ \Im\big(\phi(t)\big) = e^{-|\sigma t|^\alpha} \sin\left(\mu t + |\sigma t|^\alpha \beta \frac{|t|}{t} \tan\frac{\pi\alpha}{2}\right) \end{cases}$$

hence

$$\arctan\left(\frac{\Im\big(\phi(t)\big)}{\Re\big(\phi(t)\big)}\right) = \mu t + \beta\sigma^\alpha \frac{|t|^{\alpha+1}}{t} \tan\frac{\pi\alpha}{2} \tag{1.17}$$

Using a multiple regression for (1.17), one can then estimate the parameters β and μ as follows

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mu} \end{bmatrix} = [(X^T X)^{-1} X^T Y] \tag{1.18}$$

where

$$X = \begin{bmatrix} \vdots & \vdots \\ \hat{\sigma}^{\hat{\alpha}} \frac{|u_l|^{\hat{\alpha}+1}}{u_l} \tan\frac{\pi\hat{\alpha}}{2} & u_l \\ \vdots & \vdots \end{bmatrix} \text{ and } Y = \begin{bmatrix} \vdots \\ \arctan\left(\frac{\Im\left(\hat{\phi}(u_l)\right)}{\Re\left(\hat{\phi}(u_l)\right)}\right) \\ \vdots \end{bmatrix}$$

where for $L \in \mathbb{N}^*$, $\{u_l\}_{\{l=1,\ldots,L\}}$ is an appropriate set of real numbers depending on α. The appropriate sets of real numbers, fixed by the author for reason of consistency, are

$$\left\{t_k = \frac{\pi k}{25}\right\}_{k=\{1,\ldots,K\}} \text{ and } \left\{u_l = \frac{\pi l}{50}\right\}_{l=\{1,\ldots,L\}}$$

where K and L are the optimum values depending on α and the number of observations. This optimum values are given in [9].

Once having our estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\mu}$, one can improve upon the estimation by iterating this method (see [44]). We assume $\mathrm{S} = \{S_i\}_{i=\{1,\ldots,n\}}$ being the random variables being estimated. We describe the recursive algorithm as follow.

1- We set the initial values $\hat{\alpha}^{(0)} = \hat{\alpha}$, $\hat{\beta}^{(0)} = \hat{\beta}$, $\hat{\sigma}^{(0)} = \hat{\sigma}$ and $\hat{\mu}^{(0)} = \hat{\mu}$, where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\mu}$ are obtained by the McMculloch method. We set the optimum values K and L depending on $\hat{\alpha}$ given in [9]. We also set our stoping conditions: the maximum number of iterations M and the admissible error $\varepsilon > 0$ so that $\left(\alpha^{(i+1)} - \alpha^{(i)}\right)^2 + \left(\mu^{(i+1)} - \mu^{(i)}\right)^2 < \varepsilon$.

2- We standardise S so that $\mathrm{S}^{(0)} = \left(\mathrm{S} - \hat{\mu}^{(0)}\right)/\hat{\sigma}^{(0)}$

3- We estimate $\hat{\alpha}_0$ and $\hat{\sigma}_0$ using equation (1.16) and we set $\mathrm{S}_0 = \mathrm{S}^{(0)}/\hat{\sigma}_0$.

4- We estimate $\hat{\beta}_0$ and $\hat{\mu}_0$ using equation (1.18) and we set $\mathrm{S}^{(1)} = \left(\mathrm{S}_0 - \hat{\mu}_0\right)/\hat{\sigma_0}$. Our

new estimators are

$$\left\{\hat{\alpha}^{(1)}, \hat{\beta}^{(1)}, \hat{\sigma}^{(1)}, \hat{\mu}^{(1)}\right\} = \left\{\hat{\alpha}_0, \hat{\beta}_0, \hat{\sigma}^{(0)}\hat{\sigma}_0, \hat{\mu}^{(0)} + \hat{\sigma}^{(0)}\hat{\mu}_0\right\} \tag{1.19}$$

5- If our stopping conditions are not satisfied, then we go back to point '3-', otherwise we keep our estimator $\left\{\hat{\alpha}^{(1)}, \hat{\beta}^{(1)}, \hat{\sigma}^{(1)}, \hat{\mu}^{(1)}\right\}$.

The maximum likelihood estimator (MLE)

Since we know an approximation of the density function (either using the FFT method or the direct integration method), MLE can be applied to estimate the stable parameters. For n independent observations $S = \{S_1, \ldots, S_n\}$, an estimate of $\theta = (\alpha, \sigma, \beta, \mu)$ is obtained by maximizing the log-likelihood function

$$L_\theta(S) = \sum_{i=1}^{n} \log f(S_i; \theta) \tag{1.20}$$

Nolan [45] uses a direct approximated integration of the density function [35] and maximises the log-likelihood function using the quasi Newton method. The McCulloch method is used to initialise the parameters values before the minimisation procedure.

Comparison

We compare these three methods. We simulate a sample of α-stable random variables (see Section 1.2.4) for different parameters values. We run 50 samples of 10000 α-stable random variables. Table 1.2 shows the mean and in brackets the standard deviation of the estimated parameters for each case.

Method	$\alpha = 1.80$	$\beta = 0.80$	$\sigma = 0.25$	$\mu = -10.00$
McCulloch	1.8008(0.0366)	0.8759(0.1230)	0.2488(0.0032)	−9.9976(0.0063)
Koutrouvélis	1.7992(0.0150)	0.7921(0.0707)	0.2493(0.0019)	−10.0007(0.0050)
MLE	1.7995(0.0140)	0.7938(0.0398)	0.2494(0.0018)	−10.0001(0.0047)
Method	$\alpha = 1.30$	$\beta = 0.30$	$\sigma = 2.00$	$\mu = -0.50$
McCulloch	1.3085(0.0202)	0.3150(0.0281)	1.9907(0.0304)	−0.4895(0.1328)
Koutrouvélis	1.3051(0.0199)	0.3025(0.0278)	1.9971(0.0292)	−0.5118(0.1166)
MLE	1.3049(0.0163)	0.3031(0.0220)	1.9976(0.0254)	−0.5001(0.0277)
Method	$\alpha = 0.85$	$\beta = 0.00$	$\sigma = 5.00$	$\mu = 0.00$
McCulloch	0.8541(0.0129)	0.0006(0.0274)	4.9645(0.0915)	−0.0257(0.6511)
Koutrouvélis	0.8494(0.0104)	0.0017(0.0281)	4.9957(0.0783)	−0.0475(0.6243)
MLE	0.8500(0.0085)	−0.0002(0.0149)	4.9951(0.0763)	−0.0107(0.0791)
Method	$\alpha = 0.70$	$\beta = -0.50$	$\sigma = 20.00$	$\mu = 5.00$
McCulloch	0.7128(0.0132)	−0.5321(0.0334)	19.6372(0.9459)	7.4925(1.9243)
Koutrouvélis	0.6978(0.0117)	−0.5010(0.0260)	19.9893(0.4603)	4.9155(1.4101)
MLE	0.6986(0.0056)	−0.4994(0.0127)	20.0120(0.4202)	4.9846(0.2911)

Table 1.2: Comparison of McCulloch, Koutrouvélis and MLE for some values of α, β, σ and μ.

These three methods provide good estimates of the α-stable parameters. However their

differences are related to the speed of the computation; The slower the computation, the better the estimate: MLE is the most accurate but slowest.

1.3 Self-similar process with α-stable distribution

This section deals with self-similar processes with α-stable distribution. We introduce here some properties, and give examples of such processes.

1.3.1 Properties

From now, we consider self-similar processes with symmetric stable distribution and we note such a process H-sssi,$S\alpha S$.

Proposition 1.3.1. *Let X H-sssi,$S\alpha S$. If such a process exists, then*

$$H \leq \max\left(1, \frac{1}{\alpha}\right)$$

Proof. By self-similarity and stationarity of the increments of X, one has by Theorem 1.1.7 (ii) and 1.1.8,

$$\mathbb{E}[|X(1)|] < \infty \Rightarrow H \leq 1 \Leftrightarrow H > 1 \Rightarrow \mathbb{E}[|X(1)|] = \infty$$

For $1 < \alpha \leq 2$, one has for $1 \leq \gamma < \alpha$

$$\mathbb{E}[|X(1)|] < \mathbb{E}[|X(1)|^\gamma] < \infty$$

whence by Theorem 1.1.7 (ii) we have $H \leq 1$.
For $0 < \alpha \leq 1$, one has for $0 < \gamma < \alpha$

$$\mathbb{E}[|X(1)|^\gamma] < \infty$$

and from Theorem 1.1.7 (i), this implies $H < 1/\gamma$ whence $H \leq 1/\alpha$. □

A well known H-sssi,$S\alpha S$ is the Brownian motion, which is a self-similar process with index $H = 1/\alpha = 1/2$. The Brownian motion was used by Brown in 1827 to describe the particle motion, and later studied by Bachelier in 1900, who gives a theoretical definition in his PhD thesis "Theorie de la spéculation". A Brownian motion $\{B(t)\}_{t\geq 0}$ is defined as follows

Definition 1.3.2. *A stochastic process $\{B(t)\}_{t\geq 0}$ is a Brownian motion if*

(i) $B(0) \overset{(a.s.)}{=} 0$

(ii) it has independent and stationary increments

(iii) $\forall t \geq 0, \forall s \in [0,t]$, $B(t) - B(s) \sim \mathrm{N}(0, t-s)$

(iv) it has continuous sample paths a.s.

Theorem 1.3.3. *Brownian motion on $\mathbb{R}$ exists and it is unique.*

Proof. The existence of the Brownian motion is based on its construction. A complete proof can be found in [46], page 12 Theorem (6). This proof is based on Wiener's work [47]. □

Proposition 1.3.4. *A Brownian motion $\{B(t)\}_{t\geq 0}$ is H-sssi with index $H = 1/2$. That is $\{a^{-1/2}B(at)\}_{t\geq 0} \overset{(fdd)}{=} \{B(t)\}_{t\geq 0}$ for $a > 0$.*

Proof.

$$\begin{aligned} a^{-\frac{1}{2}}B(at) - a^{-\frac{1}{2}}B(as) &\sim a^{-\frac{1}{2}}\mathrm{N}\big(0, a(t-s)\big) \\ &\sim \mathrm{N}\big(0,(t-s)\big) \end{aligned}$$

choosing $s = 0$, we have $B(0) \overset{(a.s.)}{=} 0$. $a^{-1/2}B(at)$ inherits independent increments and continuous sample paths from $B(t)$. The result follows from the uniqueness of the Brownian motion. □

Remark 1.3.5. *The increments Brownian motion $\{B(t)\}_{t\geq 0}$ are α-stable distributed with $\alpha = 2$. Indeed, a stable distribution with $\alpha = 2$ is in fact the normal distribution. See (1.9).*

The fractional Brownian motion and the stable Lévy motion are also both well known self-similar processes studied in the literature. These processes will be described in the next two Sections.

1.3.2 Fractional Brownian motion (fBm)

The advantage of the Gaussian distribution is that a Gaussian process can be characterised up to finite-dimensional distributions only by its mean and covariance function (see [11] 7.2 page 318). It exists then a unique Gaussian process X with mean 0 and with the following covariance function (see [14] Remark 1.3.1).

$$\forall (t,s) \in \mathrm{T}^2,\ r_X(t,s) = \mathbb{E}[X(t)X(s)] = \frac{\mathbb{E}[X(1)^2]}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \qquad (1.21)$$

There is a different way of defining the fractional Brownian motion. One way is to consider Proposition 1.3.9 as a definition of the fBm, see for example [14]. Another way that we chose is the following definition given by Samorodnitsky and Taqqu in [11].

Definition 1.3.6. *A Gaussian H-sssi process $\{B^H(t)\}_{t\geq 0}$ with $H \in (0,1)$ is called fractional Brownian motion.*

Proposition 1.3.7. *The fractional Brownian motion exists and it is unique.*

Proof. This is shown in [11] page 318 lemma 7.2.1. □

Remark 1.3.8. *For $H = \frac{1}{2}$, $\{B^H(t)\}_{t\in \mathrm{T}}$ is a Brownian motion. We refer to Theorem 1.3.2 page 5 of [14].*

Proposition 1.3.9. *A stochastic process X is a fractional Brownian motion if and only if*

(i) $X(0) \overset{(a.s.)}{=} 0$

(ii) $X(t)$ *is Gaussian with mean 0*

(iii) has stationary increments

(vi) $(t,s) \in \mathrm{T}^2$, $\mathbb{V}ar\,(X(t+s) - X(t)) = s^{2H}\mathbb{E}[(X(1))^2]$, *with* $H \in (0,1)$

(v) has almost surely continuous sample paths

Proof. First suppose that X satisfies Definition 1.3.6.

(i) By self-similarity of the fBm and Proposition 1.1.5.

(ii) By definition, the fBm has Gaussian increments and

$$\mathbb{E}[X(1)] = \mathbb{E}[X(2) - X(1)] = (2^H - 1)\mathbb{E}[X(1)]$$

for $H \in (0,1)$, one has $2^H - 1 \neq 1$ and obtain then $\mathbb{E}[X(1)] = 0$. Moreover, $\forall t \in \mathrm{T}$ $\mathbb{E}[X(t)] = |t|^H\mathbb{E}[X(\mathrm{sign}(t))]$ and $\mathbb{E}[X(1)] = \mathbb{E}[X(0) - X(-1)] = -\mathbb{E}[X(-1)]$ hence $\mathbb{E}[X(t)] = 0$.

(iii) By Definition 1.3.6.

(vi) One use 1.21.

(v) One has for $n \in \mathbb{N}$, $\mathbb{E}[(X(t) - X(s))^{2n}] = \mathbb{E}[(X(1))^{2n}](t-s)^{2nH}$, moreover $\mathbb{E}[(X(1))^{2n}]$ is finite. Using the Kolmogorv's criterion[2], one shows that almost surely X has a continuous sample path.

For the converse see [11] or [14]. □

The next theorem gives an integral representation of the fractional Brownian motion.

Theorem 1.3.10. *Let $H \in (0,1)$ and $\{B(t)\}_{t\geq 0}$ a Brownian motion, the process*

$$B^H(t) = \frac{1}{C_H}\int_{-\infty}^{\infty} (t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}} dB(s), \quad t \geq 0$$

where

$$C_H = \mathbb{E}[B^H(1)^2]^{-\frac{1}{2}}\left\{\int_0^{\infty} (x+1)^{H-\frac{1}{2}} - (x)^{H-\frac{1}{2}} dx + \frac{1}{2H}\right\}^{\frac{1}{2}}$$

[2]The Kolmogorv's criterion will be introduced in Chapter 2

is a fractional Brownian motion.

We can show that this integral representation of a fractional Brownian motion is self-similar (see [14]), and this satisfies Definition 1.3.6.
In the following graphs, we compare fractional Brownian motion for different values of H. We use the algorithm developed by Wood and Chan [48] to simulate our fractional Brownian motion.

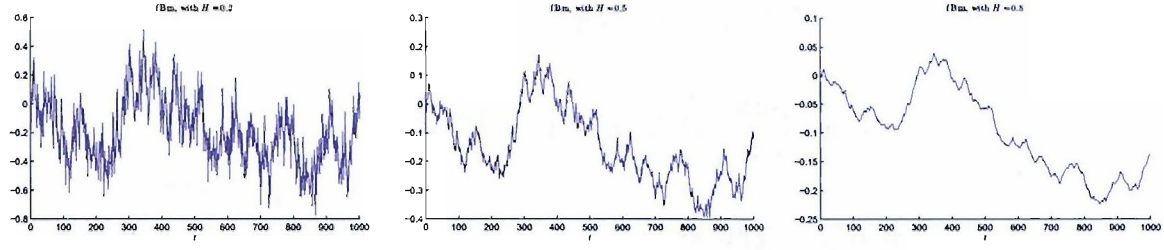

Figure 1.6: fBm sample path for H=0.2, 0.5 and 0.8.

1.3.3 α-stable Lévy process

We describe briefly the class of Lévy processes. For more details on Lévy processes, we refer to [30] and [29].

Definition 1.3.11 (Lévy process). *A stochastic process $\{L(t)\}_{t\geq 0}$ is called a Lévy process if*

(i) $L(0) \overset{(a.s.)}{=} 0$

(ii) it has independent and stationary increments

(iii) it is càdlàg (continue à droite, avec limite à gauche)

(iv) it is stochastically continuous

Theorem 1.3.12 (Characteristic function of a Lévy process). *Let $\{L(t)\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}$, then there exists a continuous function $\psi : \mathbb{R} \to \mathbb{C}$ called the characteristic exponent of L such that:*

$$\forall \theta \in \mathbb{R}, \quad \mathbb{E}[e^{i\theta L(t)}] = e^{t\psi(\theta)}$$

where:

$$\psi(\theta) = i\mu\theta - \frac{\sigma^2}{2}\theta^2 - \int_{\mathbb{R}} (1 - e^{i\theta x} - i\theta x \mathbf{1}_{|x|<1})\nu(dx)$$

and ν is a Lévy measure on $\mathbb{R}$ such that

$$\nu(\{0\}) = 0 \text{ and } \int_{\mathbb{R}} \min(1, x^2)\nu(dx) < +\infty$$

Proof. See Theorem 1 page 13 of [29]. □

There are different subclasses of Lévy processes (see [49]). However, our interest will be limited to Lévy processes that have stable distribution. We define the stable Lévy process and then show the links between self-similarity and stable Lévy motion.

Definition 1.3.13 (Stable Lévy motion). *A stochastic process $\{L_\alpha(t)\}_{t\geq 0}$ is a stable Lévy motion if it is a Lévy process with $L_\alpha(1)$ an α-stable random variable. Its characteristic function is given by*

$$\forall\theta\in\mathbb{R},\ \varphi_{L_\alpha(t)}(\theta)=\mathbb{E}[e^{i\theta L_\alpha(t)}]=e^{t\psi(\theta)}$$

Where

$$\psi(\theta)=\begin{cases} i\mu\theta-\sigma^\alpha|\theta|^\alpha\Big(1-i\beta sign(\theta)\tan\dfrac{\pi\alpha}{2}\Big) & \text{if } \alpha\neq 1;\\[2ex] i\mu\theta-\sigma|\theta|\Big(1+i\beta\dfrac{2}{\pi}sign(\theta)\log|\theta|\Big) & \text{if } \alpha=1. \end{cases}$$

Figure 1.7 shows some simulated paths of stable Lévy motion L_α, for $\alpha=\{0.5,1,1.8\}$, $\beta=0$, $\sigma=1$ and $\mu=0$. The simulations are achieved by taking the cumulative sum of simulated α-stable random variables $\{S_i\}_{i\in\mathbb{N}}$ described in Section 1.2.4, so that

$$\{L_\alpha(n)\}_{n\in\mathbb{N}}=\left\{\sum_{i=0}^{n}S_i\right\}_{n\in\mathbb{N}}$$

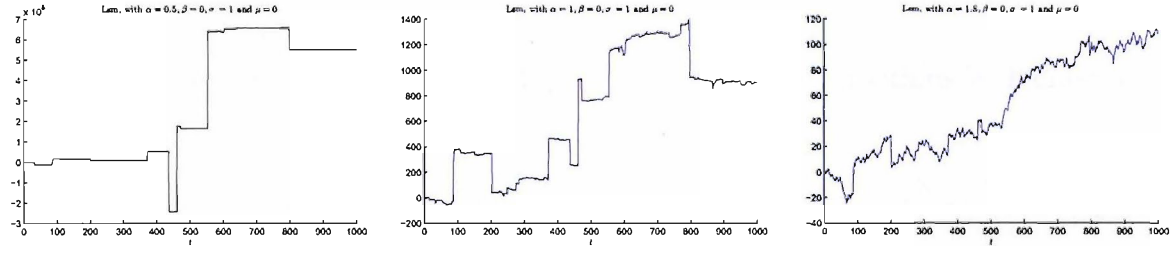

Figure 1.7: Stable Lévy motion sample path for $\alpha=0.5$, 1 and 1.8

To understand when a stable Lévy process is self-similar, we give the following proposition.

Proposition 1.3.14. *An α-stable Lévy process X is self-similar with index $H=1/\alpha$ if and only if, it has a strictly symmetric stable law ($\mu=0$ and $\beta=0$).*

Proof. If

If $\mu=0$ and $\beta=0$, then from the definition of the exponent function ψ of the characteristic function, we have for $t=(t_0,t_1,\ldots,t_n)\in \mathrm{T}^n$ and $i\in\{0,\ldots,n\}$

$$\forall\theta\in\mathbb{R},\ a>0,\ a\psi(\theta)=\psi(a^{1/\alpha}\theta)\quad\Leftrightarrow\quad\Big(\varphi_{X(t_i)}(\theta)\Big)^a=\varphi_{X(t_i)}(a^{1/\alpha}\theta)$$

however

$$\big(\varphi_{X(t_i)}(\theta)\big)^a=\big(e^{t_i\psi(\theta)}\big)^a=e^{at_i\psi(\theta)}=\mathbb{E}\big[e^{i\theta X(at_i)}\big]=\varphi_{X(at_i)}(\theta)$$

and

$$\varphi_{X(t_i)}(a^H\theta) = \mathbb{E}\big[e^{\mathrm{i}\theta a^H X(t_i)}\big] = \varphi_{a^H X(t_i)}(\theta)$$

The stochastic process $X = \{X(t_i)\}_{i\in\{0,\dots,n\}}$ has independent increments, for $t = (t_0, t_1, \dots, t_n) \in \mathrm{T}^n$. One has

$$\varphi_{X(at)}(\theta) = \prod_{i=0}^{n} \varphi_{X(at_i)}(\theta) = \prod_{i=0}^{n} \varphi_{a^H X(t_i)}(\theta) = \varphi_{a^H X(t)}(\theta)$$

Hence $\varphi_{a^H X(t)}(\theta) = \varphi_{X(at)}(\theta)$. Whence $\{X(at_i)\}_{i\in\{0,\dots,n\}} \overset{(fdd)}{=} \{a^H X(t_i)\}_{i\in\{0,\dots,n\}}$.
Only if
Similarly, if $\mu \neq 0$ and $\beta \neq 0$, then $\varphi_{a^H X(t)}(\theta) \neq \varphi_{X(at)}(\theta)$. Whence

$$\{X(at_i)\}_{i\in\{0,\dots,n\}} \overset{(fdd)}{\neq} \{a^H X(t_i)\}_{i\in\{0,\dots,n\}}$$

□

1.3.4 Other examples of H-sssi,$S\alpha S$ processes

There are other H-sssi,$S\alpha S$ processes (see [11] chapter 7). We introduce some of them without giving any details.

Linear fractional stable motion

For $0 < H < 1$, $0 < \alpha < 2$ with $H \neq \frac{1}{\alpha}$, Linear fractional stable motions is defined as follows

$$Z(t) = \int_{-\infty}^{0} \Big(a\big((t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha}\big) + b\big((t-s)_-^{H-1/\alpha} - (-s)_-^{H-1/\alpha}\big) \Big) dL_\alpha(s) \quad (1.22)$$

where $(a, b) \in \mathbb{R}^2$ with $ab \neq 0$, $x_+ = \max(x, 0)$, $x_- = \min(x, 0)$ and L_α is a stable Lévy process.

Log fractional stable motion

For $H = \frac{1}{\alpha}$, $0 < \alpha < 2$, Log fractional stable motion is defined as follows

$$Z(t) = \int_{-\infty}^{+\infty} \log\left|\frac{t-s}{s}\right| dL_\alpha(s) \quad (1.23)$$

where L_α is a stable Lévy process.
The advantage of self-similar processes is that it allows us to know the statistical properties at any other time scale from one time scale. But this advantage generally restricts its application as a model to some real phenomena, where the index of self-similarity changes over different time scales.

1.4 Fundamental limit Theorem

We present another type of self-similar process called Rosenblatt process, see [50]. We first present the following Theorem from which the existence of the process follow.

Theorem 1.4.1. *Assume X is stochastically continuous and non-trivial. If there exists a process $\{Y(t)\}_{t\in \mathrm{T}}$ and real numbers $\{A(\lambda), \lambda \geq 0\}$ with $A(\lambda) > 0$, $\lim_{\lambda\to+\infty} A(\lambda) = +\infty$ such that*

$$\frac{1}{A(\lambda)} Y(\lambda t) \overset{(fdd)}{\to} X(t), \text{ as } \lambda \to +\infty$$

Then for some $H > 0$, X is H-ss . $A(\lambda)$ is of the form $A(\lambda) = \lambda^H L(\lambda)$, where L is a slowly varying function $\left(\text{for all } x > 0, \lim_{t\to+\infty} \frac{L(tx)}{L(t)} = 1\right)$.

Proof. see [14] page 13. □

Proposition 1.4.2. *Let $\{X_i\}_{i\geq 0}$ be a stationary normal distributed sequence with mean 0 and variance 1, Such that $\mathrm{Cov}\,(X_i, X_{i+k}) \sim k^{-\gamma} L(k)$ as $k \to +\infty$, with $\gamma \in (1/2, 1)$ and L slowly varying. Then*

$$Z_n(t) = \frac{1}{n^{1-\gamma} L(n)} \sum_{i=1}^{\lfloor nt \rfloor} (X_i^2 - 1) \overset{(fdd)}{\to} R(t)$$

Where $R(t)$ is called the Rosenblatt process, which is H-sssi with $H = 1 - \gamma$.

Chapter 2

Introduction to locally asymptotically self-similar processes

2.1 Local regularity measure of a function

In this Section, we present the notion of Hausdorff dimension and the Hölder exponent as a measure of the roughness of a function sample path. The roughness of a process has been studied by many authors. Falconer in [51]; Peltier and Lévy Véhel in [52], studied the local structure of a process by looking at the local self-similarity index. Seuret and Lévy Véhel in [53] and Daoudi in [54], studied the same local structure in terms of the Hölder exponent. The local self-similarity index and the Hölder exponent both allow us to bound the fractal dimension of the process sample path.

2.1.1 Hausdorff measure and the Hausdorff dimension

Let $\mathcal{U}$ a non-empty subset of n-dimensional Euclidean space $\mathbb{R}^n$, then the diameter of $\mathcal{U}$ defined as $\text{diam}(\mathcal{U}) = \sup\{|x-y| : x, y \in \mathcal{U}\}$. If $\{\mathcal{U}_i\}_{i\in \text{I}}$ is a countable collection of sets of diameter at most δ that cover E; that is $E \subset \bigcup_{i\in \text{I}} \mathcal{U}_i$, then $\{\mathcal{U}_i\}_{i\in \text{I}}$ is called a δ-cover of E. Let $E \subseteq \mathbb{R}$, and let $\{\mathcal{U}_i\}_{i\in \text{I}}$ be a δ-cover of E, we define for $s \in \mathbb{R}^+$

$$\mathcal{H}^s_\delta(E) = \inf\left\{\sum_{i\in \text{I}} \text{diam}(\mathcal{U}_i)^s\right\}$$

where $\mathcal{H}^s_\delta(E)$ is minimised over all countable δ-cover $\{\mathcal{U}_i\}_{i\in \text{I}}$. We define the s-dimensional Hausdorff measure of E by

$$\mathcal{H}^s(E) = \lim_{\delta\to 0} \mathcal{H}^s_\delta(E)$$

This limit exists because as δ decreases, $\mathcal{H}^s_\delta(E)$ increases, and so approaches a limit as $\delta \to 0$. This is illustrate in Figure 2.1.
The Hausdorff dimension $\dim_{\mathcal{H}}(E)$ of E is defined by (we refer to [12] for more detail)

$$\dim_{\mathcal{H}}(E) = \inf\{s \geq 0,\ \mathcal{H}^s(E) = 0\}$$

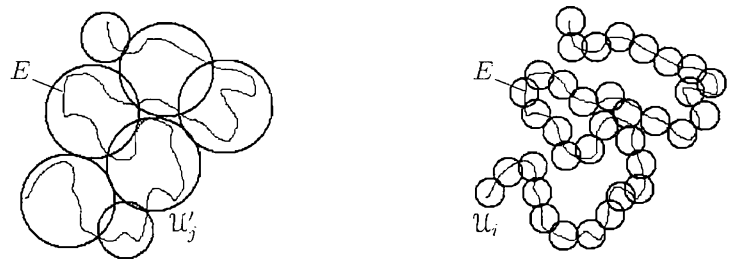

Figure 2.1: Two δ-cover of E with $\sup_{j\in J}(\text{diam}(\mathcal{U}'_j)) > \sup_{i\in I}(\text{diam}(\mathcal{U}_i))$

Note that for $0 \leq s < \dim_{\mathcal{H}}(E)$ we have $\mathcal{H}^s(E) = +\infty$, and for $s > \dim_H(E)$, we have $\mathcal{H}^s(E) = 0$. This is illustrated in Figure 2.2. For $s = \dim_H(E)$, $\mathcal{H}^s(E) \in [0, +\infty]$

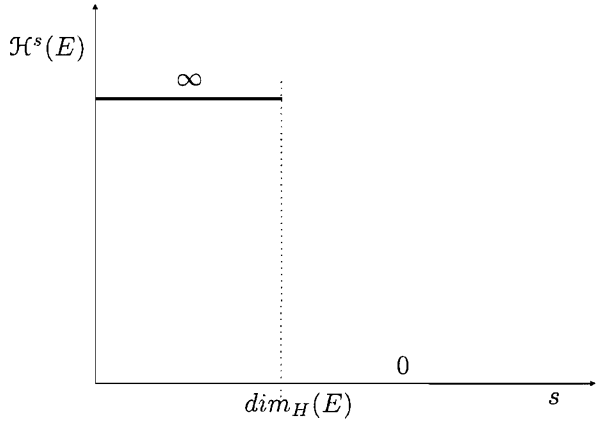

Figure 2.2: Graph of $\mathcal{H}^s(E)$

Unfortunately, in practice, it is difficult to calculate the Hausdorff dimension. However, the box-counting dimension (see Definition 2.1.1) can be used instead (for more details on box-counting see [12]).

Definition 2.1.1 (box-counting dimension [12]). *Let F be any non-empty bounded subset of $\mathbb{R}^n$ and let $N_\delta(F)$ be the smallest number of sets of diameter at most δ that can cover F. The lower and upper box-counting dimension of F are given by*

$$\underline{\dim}_B(F) = \lim_{\delta\to 0} \frac{\log N_\delta(F)}{-\log \delta} \text{ and } \overline{\dim}_B(F) = \overline{\lim_{\delta\to 0}} \frac{\log N_\delta(F)}{-\log \delta}$$

When these two limits coincide, the box-counting dimension of F is then defined by

$$\dim_B(F) = \lim_{\delta\to 0} \frac{\log N_\delta(F)}{-\log \delta}$$

whenever this limit exists.

2.1.2 Hölder exponent

To characterise the roughness of a stochastic process X, one can measure its regularity at point t by the pointwise Hölder exponent (2.1) or by the local Hölder exponent (2.2).

$$\alpha_X^P(t_0,\omega) = \sup\left\{\alpha > 0, \quad \limsup_{\delta\to 0} \frac{X(t_0+\delta,\omega) - X(t_0,\omega)}{|\delta|^\alpha} = 0\right\} \tag{2.1}$$

$$\alpha_X^L(t_0,\omega) = \sup\left\{\alpha > 0, \exists\epsilon > 0, \sup_{(t,t')\in]t_0-\epsilon,t_0+\epsilon[} \frac{|X(t,\omega) - X(t',\omega)|}{|t-t'|^\alpha} < \infty\right\} \tag{2.2}$$

Note that we have $\alpha_X^L \leq \alpha_X^P$ (see [53]).

Definition 2.1.2. *Let X be a stochastic process. X is Hölder continuous of order $\alpha \in (0,1)$ on T, where $\mathrm{T} \subseteq \mathbb{R}$ if*

$$\mathbb{P}\left(\omega \in \Omega : \sup_{\substack{0<|t-s|<h(\omega)\\(s,t)\in T}} \frac{|X(t,\omega) - X(s,\omega)|}{|t-s|^\alpha} \leq \delta\right) = 1 \tag{2.3}$$

where h is an almost surely positive random variable and $\delta > 0$ is an appropriate constant.

For simplicity, we use $X(t)$ instead of $X(t,\omega)$ and $\alpha_X(t_0)$ instead of $\alpha_X(t_0,\omega)$.

Theorem 2.1.3. Kolmogorov criterion (see [55])
Let X be a random process. If there exists strictly positive constants p, β, C such that $\forall t \in \mathrm{T}$ and $h > 0$, $\mathbb{E}|X(t+h) - X(t)|^p \leq Ch^{1+\beta}$ then

$$\lim_{h\to 0} \sup_{\substack{(t,s)\in \mathrm{T}\\|t-s|<h}} |X(t) - X(s)| \overset{(a.s.)}{=} 0$$

and X has a modification[1] $\left\{\tilde{X}(t)\right\}_{t\in\mathrm{T}}$ *whose sample paths are locally Hölder continuous of index $\alpha \in (0, \beta/p)$.*

Corollary 2.1.4. *Let X be a random process such that there exists strictly positive constants p, H, C_p so that $\forall t \in \mathrm{T}$ and $h > 0$*

$$\mathbb{E}|X(t+h) - X(t)|^p \leq C_p h^{pH}$$

Then using the Kolmogorov criterion, X has a modification whose sample paths are locally Hölder continuous of index $\alpha \in (0, H - 1/p)$.

Proof. Using the Theorem 2.1.3, with $\beta = pH - 1$, the result is straightforward. □

[1] $\tilde{X}(t,\omega)$ is a modification of $X(t,\omega)$, if $\mathbb{P}\left(\{\omega \in \Omega : \tilde{X}(t,\omega) = X(t,\omega)\}\right) = 1$.

This corollary gives us a lower bound of the local Hölder exponent of X at time t

$$H - 1/p \leq \alpha_X^L(t) \tag{2.4}$$

Remark 2.1.5. *If we have a stochastic process X satisfying conditions of Corollary 2.1.4 and which has moments of all order finite (e.g: fractional Brownian motion B_H), then we have $H \leq \alpha_X^L(t)$.*

2.2 Local structure of a random process

2.2.1 Tangent processes

We define $\mathcal{F}$ the class of random process $X = \{X(t)\}_{t\geq 0}$ of $\mathcal{D}(\mathbb{R})$, where $\mathcal{D}(\mathbb{R})$ is the space of real valued càdlàg functions. Let $\mathcal{F}_0 = \{Y \in \mathcal{F},\ Y(0) \overset{(a.s.)}{=} 0\}$, then the *scaling operator* $\mathcal{T}_{t,\epsilon} : \mathcal{F} \to \mathcal{F}_0$ is defined by

$$(\mathcal{T}_{t,\epsilon}X)(u) = X(t+\epsilon u) - X(t) \qquad u \geq 0$$

Definition 2.2.1. *[51] For $X \in \mathcal{F}$, we say $T_{X,t} \in \mathcal{F}_0$ is a tangent process of X at point $t \geq 0$, if there exist sequences $r_n \to 0$ and $c_n \to 0$ such that*

$$\lim_{n\to+\infty} c_n^{-1}\mathcal{T}_{t,r_n}X \overset{(fdd)}{=} T_{X,t}$$

The tangent processes reflect the local structure of a random process. We define the tangent space $\mathrm{Tan}(X,t)$ of X at point $t \geq 0$ by

$$\mathrm{Tan}(X,t) = \{Y \in \mathcal{F}_0 :\ \text{there exists } r_n \to 0 \text{ and } c_n \to 0 \text{ such that } c_n^{-1}\mathcal{T}_{t,r_n}X \overset{(fdd)}{\to} Y\}$$

Tangent processes are shift invariant and self-similar. These properties are described in [51]. To deal with the regularity of the sample path of a process X, we study the class of tangent processes where $c_n = r_n^\alpha$ with $\alpha > 0$ so that $r_n^{-\alpha}\mathcal{T}_{t,r_n}X \overset{(fdd)}{\to} T_{X,t}$.

Lemma 2.2.2. *Let X be a random process of $\mathcal{D}(\mathbb{R})$, such that*

$$r_n^{-\alpha}\mathcal{T}_{t,r_n}X \overset{(fdd)}{\to} T_{X,t}, \quad t \geq 0$$

Then the tangent process $T_{X,t}$ is self-similar with index α.

Proof. Setting $r'_n = ar_n \to 0$ for some $a > 0$, one gets

$$\begin{aligned}
\{T_{X,t}(au)\}_{u\geq 0} &\overset{(fdd)}{=} \lim_{r_n\to 0}\{r_n^{-\alpha}\mathcal{T}_{t,r_n}X(au)\}_{u\geq 0} \\
&\overset{(fdd)}{=} a^\alpha \lim_{r'_n\to 0}\{(r'_n)^{-\alpha}\mathcal{T}_{t,r'_n}X(u)\}_{u\geq 0} \\
&\overset{(fdd)}{=} a^\alpha\{T_{X,t}(u)\}_{u\geq 0}
\end{aligned}$$

□

Lemma 2.2.3. *Let X be a random process of $\mathcal{D}(\mathbb{R})$ with stationary increments, such that*

$$r_n^{-\alpha}\mathcal{T}_{t,r_n}X \overset{(fdd)}{\to} T_{X,t}, \quad t \geq 0$$

Then

$$\forall t \geq 0,\ T_{X,t} \overset{(fdd)}{=} T_{X,0}$$

Proof. For $t \geq 0$ we have:

$$\begin{aligned}\{T_{X,t}(u)\}_{u\geq 0} &\overset{(fdd)}{=} \lim_{r_n\to 0}\{r_n^{-\alpha}\big(X(t+r_nu)-X(t)\big)\}_{u\geq 0}\\ &\overset{(fdd)}{=} \lim_{r_n\to 0}\{r_n^{-\alpha}X(r_nu)\}_{u\geq 0}\\ &\overset{(fdd)}{=} \{T_{X,0}(u)\}_{u\geq 0}\end{aligned}$$

□

2.2.2 Locally asymptotically self-similar

Definition 2.2.4. *[51] A process $X = \{X(t)\}_{t\geq 0}$ is locally asymptotically self-similar (lass for short) at point $t \geq 0$ with index $h(t)$ if,*

$$\lim_{\epsilon\to 0^+}\left\{\frac{X(t+\epsilon u)-X(t)}{\epsilon^{h(t)}}\right\}_{u\geq 0} \overset{(fdd)}{=} \{T_{X,t}(u)\}_{u\geq 0} \tag{2.5}$$

where the non-degenerate process $\{T_{X,t}(u)\}_{u\geq 0}$ is the tangent process at point t of X

Definition 2.2.5. *We say that the process X is lass with multifractional function $h : [0,+\infty) \to (0,1)$, if for each $t \geq 0$, process is lass with index $h(t)$*

Note that the tangent process $T_{X,t}$ is self-similar (see [51]). The process X is lass with multifractional function $h : [0,+\infty) \to (0,1)$, if for each $t \in [0,+\infty)$, the process is lass with index $h(t)$. A definition of local self-similarity (*lss* for short) is given by Benassi, Cohen and Istas [56]. A process X is *lss* at point t with index $h(t)$, if there exist a non-degenerate random variable $T_{X,t}$ such that

$$\lim_{\epsilon\to 0}\frac{X(t+\epsilon)-X(t)}{\epsilon^{h(t)}} \overset{(d)}{=} T_{X,t}$$

Clearly *lass* implies *lss* with $T_{X,t} = T_{X,t}(1)$.

2.3 Self-similar and locally asymptotically self-similar process

2.3.1 From self-similar to lass process

Let us see how self-similar processes are linked to lass processes. Considering a H-sssi processes X, then this process is lass. Indeed, one has for some $\epsilon > 0$,

$$\begin{aligned} X \text{ is } H\text{-sssi } &\Rightarrow \left\{ \frac{X(t+\epsilon u) - X(t)}{\epsilon^H} \right\}_{u\geq 0} \overset{(fdd)}{=} \{X(u)\}_{u\geq 0} \\ &\Rightarrow \lim_{\epsilon \to 0^+} \left\{ \frac{X(t+\epsilon u) - X(t)}{\epsilon^H} \right\}_{u\geq 0} \overset{(fdd)}{=} \{X(u)\}_{u\geq 0} \end{aligned}$$

2.3.2 From lass to self-similar process

Now if we consider a process that is *lass* almost everywhere on $\mathbb{R}$, with index $h_X(t) = H$, then this process does not need to be self-similar i.e.: The *Filtered White Noise* [57], which is defined by the following representation

$$X(t) = \int_{\mathbb{R}} \frac{a(t,s)(e^{its} - 1)}{|s|^{1/2+H}} dW(s)$$

Where $dW(s)$ is a Brownian random measure and $a \in \mathcal{C}^2$. For the *lass* properties of this process see [57]. For self-similarity, we show that the relation $X(ct) \overset{(d)}{=} c^H X(t)$, $c > 0$ is not satisfied, due to the function $t \mapsto a(t,s)$. Indeed one has

$$X(ct) = \int_{\mathbb{R}} \frac{a(ct,s)(e^{icts} - 1)}{|s|^{1/2+H}} dW(s) \overset{(d)}{=} c^H \int_{\mathbb{R}} \frac{a(ct,u/c)(e^{itu} - 1)}{|u|^{1/2+H}} dW(u)$$

Assuming for $t \in \mathbb{R}$, $a(ct, u/c) \neq a(t,u)$ $\forall u \in \mathbb{R}$, then the Filtered White Noise process is not self-similar. A lass process with stationary increments and $h_X(t) = H$ (we will see later that processes with stationary increments have a constant index of self-similarity over $\mathbb{R}$) does not imply that the process is self-similar. For example, the Real Harmonizable Fractional Lévy Motion (RHFLM), represented in (2.6), is lass with stationary increments. We refer to the paper of S. Cohen et al. [58] for more details on the properties of the RHFLM.

$$X_H(t) = \int_{\mathbb{R}} \frac{e^{-ixt} - 1}{x^{H+1/2}} L(dx) \tag{2.6}$$

where for $(a,b) \in \mathbb{R}^2$, $L(dx) = aM(dx) + bW(dx)$ is a Lévy random measure with $W(dx)$ a Wiener measure independent of $M(dx)$, and $M(dx)$ a Lévy random measure without Brownian component. We present some example of lass processes.

Multifractional Brownian motion

Let $H : \mathbb{R} \to (0,1)$ be a function, then the Multifractional Brownian motion, is repre-

sented as follows

$$X(t) = \int_{-\infty}^{0} \Big(a\big((t-s)_+^{H(t)-1/2}(-s)_+^{H(t)-1/2}\big) + b\big((t-s)_-^{H(t)-1/2}(-s)_-^{H(t)-1/2}\big) \Big) dB(s) \tag{2.7}$$

Where $dB(s)$ is a Brownian random measure. The lass property is shown in [59].

Filtered White Noises

For $0 < H < 1$, the Filtered White Noises [57] is represented by the following expression.

$$X(t) = \int_{\mathbb{R}} \frac{a(t,s)(e^{its}-1)}{|s|^{1/2+H}} dW(s) \tag{2.8}$$

Where $dB(s)$ is a Brownian random measure and $t \longrightarrow a(t,s) \in \mathcal{C}^2(\mathbb{R})$.
Refer to [57] for the lass property of this process.

2.3.3 Relationship between the local self-similar index and the Hausdorff dimension

We state in this section the result given by Benassi, Cohen and Istas [56] on the link between the Hausdorff dimension and the local self-similar index H. Assuming that the process X is locally self-similar with index H at every point t, then the dimension of the graph of X, $\dim_{\mathcal{H}}(X) \geq 2 - H$. The proof of this inequality can be found in [56]. Moreover, they showed that if the sample path of X is $(H-\epsilon)$-Hölder continuous, then $\dim_{\mathcal{H}}(X) \geq 2 - H$.

2.3.4 Link between the Hölder exponent and the local self-similarity index

Let $X = \{X(t)\}_{t\geq 0}$ be a *lass* process with index $h_X(t)$ at time t. By the definition of *lass* process, the pointwise Hölder exponent $\alpha_X^P(t)$ of $X(t)$ can not be strictly greater than $h_X(t)$, but it does not imply that they are equal. We introduce the link between the Hölder exponent and the local self-similar index in the following Proposition.

Proposition 2.3.1. *Let X be a lass process with index $h_X(t)$ at $t \in \mathbb{R}$, then its pointwise Hölder exponent $\alpha_X^P(t)$ at t is such that*

$$\alpha_X^P(t) \leq h_X(t) \text{ almost surely} \tag{2.9}$$

Proof. We set $h = h_X(t)$ for simplicity. Let $\eta > 0$, from the definition of the local self-similarity provided that $\mathbb{P}(T_{X,t}(1) = 0) = 0$, one has

$$\lim_{\epsilon \to 0^+} \frac{\epsilon^{h+\eta}}{|X(t+\epsilon) - X(t)|} \overset{(d)}{=} 0$$

this implies for $\rho > 0$

$$\lim_{\epsilon \to +0} \mathbb{P}\left(\frac{\epsilon^{h+\eta}}{|X(t+\epsilon) - X(t)|} > \rho \right) = 0$$

whence

$$\lim_{\epsilon \to +0} \mathbb{P}\left(\frac{|X(t+\epsilon) - X(t)|}{\epsilon^{h+\eta}} < 1/\rho \right) = 0$$

and thus, there exists a sequence ϵ_n (Theorem 2.4.4 in [60]) such that $\lim_{n \to +\infty} \epsilon_n = 0$

$$\lim_{n \to +\infty} \frac{|X(t+\epsilon_n) - X(t)|}{\epsilon_n^{h+\epsilon}} \overset{(a.s.)}{=} +\infty \tag{2.10}$$

Hence $\alpha_X(t)^P < h + \epsilon$ almost surely or more generally since this inequality is true for all $\epsilon > 0$, $\alpha_X^P(t) \leq h$ a.s. □

For gaussian processes, one has $\alpha_X(t)^P = h_X(t)$ almost surely [61].

Remark 2.3.2. *If the process X satisfies the condition of Corollary 2.1.4, then we have the inequality (2.4). Moreover if the assumption made in Proposition 2.3.1 is also satisfied, then we have the following relation between $\alpha_X^P(t_0)$, $\alpha_X^L(t_0)$ and $H_X(t_0)$*

$$h_X(t_0) - 1/p \leq \alpha_X^L(t_0) \leq \alpha_X^P(t_0) \leq h_X(t_0)$$

For a fractional Brownian motion that has a moment of all order ($p \to +\infty$) then we have

$$\alpha_X^L(t_0) = \alpha_X^P(t_0) = h_X(t_0)$$

Part II

Time-changed H-sssi processes and estimation

Chapter 3

Time-change effects on the roughness of the sample path process

In this Chapter, we study the time change effects on a random process sample path, which will be assumed to be locally self-similar. Let $\mathrm{T} \subseteq \mathbb{R}^+$, we define the set of time change function $\Theta(\mathrm{T})$ by

$$\Theta(\mathrm{T}) = \left\{\theta : \mathrm{T} \to \mathrm{T} \in \mathcal{C}^1, \quad \text{s.t.} \quad \forall t \in \mathrm{T},\, \theta'(t) > 0, \theta(0) = 0\right\} \tag{3.1}$$

We set $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is locally self-similar and $\theta \in \Theta([0, +\infty))$.

3.1 Effect of the time change on the local structure of a random process

We study the effects of the function $\theta \in \Theta(\mathbb{R}^+)$ on the Hölder exponent and on the local index of self-similarity. These indexes may vary over time $t \geq 0$. We then study these indexes in the case where the process $\tilde{X}$ has stationary increments. We will see in this case the local self-similarity index of the process $X(t) = \tilde{X}(\theta(t))$ must be constant over the time.

3.1.1 Effect of the time change on the Hölder exponent

We define the pointwise Hölder exponent $\alpha_X(t)$ of a process X at t by (see [62])

$$\alpha_X(t) = \sup\left\{\alpha; \limsup_{h \to 0} \frac{|X(t+h) - X(t)|}{|h|^\alpha} = 0\right\}$$

The local Hölder exponent $\tilde{\alpha}_X(t)$ of the process X at t is given by

$$\tilde{\alpha}_X(t) = \sup\left\{\alpha; \exists \eta > 0 \sup_{u,v\in B(t,\eta)} \frac{|X(u)-X(v)|}{|u-v|^\alpha} < \infty\right\}$$

Where $B(t,\eta)$ is the open ball centered in t and of radius η. This last definition introduced in [62], is convenient in this section compare to the definition givenat the beginning of Section 2.1.2.

Assuming our process $X(t) = \tilde{X}(\theta(t))$, we study the effect of the time-change on the Hölder exponent. One has the following Proposition

Proposition 3.1.1. *Let $\theta \in \Theta(\mathbb{R}^+)$, the Hölder exponent of X at $t_0 \in \mathbb{R}$ is equal to the Hölder exponent of $\tilde{X}$ at $t_0' = \theta(t_0) \in \mathbb{R}$, one gets*

$$\alpha_X(t_0) = \alpha_{\tilde{X}}(\theta(t_0)) \quad \text{and} \quad \tilde{\alpha}_X(t_0) = \tilde{\alpha}_{\tilde{X}}(\theta(t_0)) \tag{3.2}$$

Proof. Setting $u' = \theta(u)$, $v' = \theta(v)$, $t' = \theta(t)$ and $\eta' = \theta(t+\eta) - \theta(t)$, one has

$$\frac{|X(u)-X(v)|}{|u-v|^\alpha} = \frac{|\tilde{X}(\theta(u)) - \tilde{X}(\theta(v))|}{|u-v|^\alpha} = \left|\frac{u'-v'}{u-v}\right|^\alpha \frac{|\tilde{X}(u')-\tilde{X}(v')|}{|u'-v'|^\alpha}$$

Moreover

$$u, v \in B(t,\eta) \Leftrightarrow u', v' \in B(t',\eta')$$

The function θ is $\mathcal{C}^1$, and $\theta'(u) > 0$, so it exists $\eta > 0$ such that

$$\sup_{u,v\in B(t,\eta)} \left|\frac{u'-v'}{u-v}\right|^\alpha \in (0, +\infty)$$

Hence

$$\sup_{u,v\in B(t,\eta)} \frac{|X(u)-X(v)|}{|u-v|^\alpha} < \infty \Leftrightarrow \sup_{u',v'\in B(t',\eta')} \frac{|\tilde{X}(u')-\tilde{X}(v')|}{|u'-v'|^\alpha} < \infty$$

and thus

$$\tilde{\alpha}_X(t) = \tilde{\alpha}_{\tilde{X}}(\theta(t))$$

Similarly, we set $h' = \theta(t+h) - \theta(t)$. The continuity of the function θ implies $h' \to 0$ as $h \to 0$. For $\alpha < \alpha_X(t)$, one has

$$\limsup_{h\to 0} \frac{|X(t+h)-X(t)|}{|h|^\alpha} = 0 \Leftrightarrow \limsup_{h'\to 0} \frac{|\tilde{X}(t'+h')-\tilde{X}(t')|}{|h'|^\alpha} = 0$$

and for $\alpha > \alpha_X(t)$, one has

$$\limsup_{h\to 0} \frac{|X(t+h)-X(t)|}{|h|^\alpha} = +\infty \Leftrightarrow \limsup_{h'\to 0} \frac{|\tilde{X}(t'+h')-\tilde{X}(t')|}{|h'|^\alpha} = +\infty$$

Whence

$$\alpha_X(t) = \alpha_{\tilde{X}}(\theta(t))$$

□

The last Proposition shows that the Hölder exponent is affected in the sense that it undergoes a time change; one has $\alpha_X = \alpha_{\tilde{X}} \circ \theta$ and $\tilde{\alpha}_X = \tilde{\alpha}_{\tilde{X}} \circ \theta$.

3.1.2 Effect of the time-change on the local self-similarity index

We consider our process $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is lass, with a local index $h_{\tilde{X}}(t)$ at $t \in \mathbb{R}^+$ and $\theta \in \Theta(\mathbb{R}^+)$. We link in Proposition 3.1.3, the local self-similarity index of X to the one of $\tilde{X}$. We first present the following Lemma

Lemma 3.1.2. *Let Y be a lass process of $\mathcal{D}(\mathbb{R})$. Then for any $t \geq 0$*

$$\lim_{\epsilon \to 0^+} \left\{ \frac{Y(\theta(t+\epsilon u)) - Y(\theta(t))}{(\theta(t+\epsilon u) - \theta(t))^{h_Y(\theta(t))}} \right\}_{u>0} \overset{(fdd)}{=} \left\{ u^{-h_Y(\theta(t))} T_{Y,\theta(t)}(u) \right\}_{u>0} \tag{3.3}$$

Proof.
One has $\forall t \geq 0$,

$$\lim_{\epsilon \to 0^+} \left\{ \frac{Y(\theta(t+\epsilon u)) - Y(\theta(t))}{(\theta(t+\epsilon u) - \theta(t))^{h_Y(\theta(t))}} \right\}_{u>0} = \lim_{\epsilon \to 0^+} \left\{ u^{-h_Y(\theta(t))} \frac{Y(\theta(t) + u\eta_\epsilon) - Y(\theta(t))}{\eta_\epsilon^{h_Y(\theta(t))}} \right\}_{u>0}$$

where $\eta_\epsilon = \dfrac{\theta(t+\epsilon u) - \theta(t)}{u}$.
From the equality above, one gets for $t \geq 0$ and $u > 0$

$$\begin{aligned} \lim_{\epsilon \to 0^+} \eta_\epsilon &= \lim_{\epsilon \to 0^+} \frac{\theta(t+\epsilon u) - \theta(t)}{u} \\ &= \lim_{\epsilon \to 0^+} \epsilon \theta'(t) \\ &= 0^+ (\text{ because } \theta'(t) > 0) \end{aligned}$$

So

$$\begin{aligned} \lim_{\epsilon \to 0^+} \left\{ \frac{Y(\theta(t+\epsilon u)) - Y(\theta(t))}{(\theta(t+\epsilon u) - \theta(t))^{h_Y(\theta(t))}} \right\}_{u>0} &= \lim_{\eta_\epsilon \to 0^+} \left\{ u^{-h_Y(\theta(t))} \frac{Y(\theta(t) + u\eta_\epsilon) - Y(\theta(t))}{(\eta_\epsilon)^{h_Y(\theta(t))}} \right\}_{u>0} \\ &\overset{(fdd)}{=} \left\{ u^{-h_Y(\theta(t))} T_{Y,\theta(t)}(u) \right\}_{u>0} \end{aligned}$$

□

The local self-similarity index of processes X and $\tilde{X}$ are linked in the next Proposition

Proposition 3.1.3. *Let $\tilde{X}$ be a lass process with index $h_{\tilde{X}}(t)$ at $t \geq 0$ and $\theta \in \Theta(\mathbb{R}^+)$. Then the process $X(t) = \tilde{X}(\theta(t))$ is lass with index of self-similarity $h_X(t) = h_{\tilde{X}}(\theta(t))$ for any $t \geq 0$.*

Proof. For $u > 0$

$$\lim_{\epsilon\to 0^+}\left\{\frac{X(t+\epsilon u)-X(t)}{\epsilon^{h_X(t)}}\right\}_{u>0}$$
$$=\lim_{\epsilon\to 0^+}\left\{\frac{\tilde{X}(\theta(t+\epsilon u))-\tilde{X}(\theta(t))}{\epsilon^{h_X(t)}}\right\}_{u>0}$$
$$=\lim_{\epsilon\to 0^+}\left\{u^{h_{\tilde{X}}(\theta(t))}\epsilon^{h_{\tilde{X}}(\theta(t))-h_X(t)}\frac{\tilde{X}\left(\theta(t+\epsilon u)\right)-\tilde{X}\left(\theta(t)\right)}{(\theta(t+\epsilon u)-\theta(t))^{h_{\tilde{X}}(\theta(t))}}\left(\frac{\theta(t+\epsilon u)-\theta(t)}{\epsilon u}\right)^{h_{\tilde{X}}(\theta(t))}\right\}_{u>0}$$

From Lemma 3.1.2

$$\lim_{\epsilon\to 0^+}\left\{\frac{\left(\tilde{X}(\theta(t+\epsilon u))-\tilde{X}(\theta(t))\right)}{(\theta(t+\epsilon u)-\theta(t))^{h_{\tilde{X}}(t)}}\right\}_{u>0}\overset{(fdd)}{=}\left\{u^{-h_{\tilde{X}}(\theta(t))}T_{\tilde{X},\theta(t)}(u)\right\}_{u>0}$$

$\theta\in\Theta(\mathbb{R}^+)$ leads to

$$\lim_{\epsilon\to 0^+}\left(\frac{\theta(t+\epsilon u)-\theta(t)}{\epsilon u}\right)^{h_{\tilde{X}}(\theta(t))}=\theta'(t)^{h_{\tilde{X}}(\theta(t))}$$

Finally to obtain a non-trivial tangent process as a limit, one must set $h_{\tilde{X}}(\theta(t))-h_X(t)=0$, which implies

$$\lim_{\epsilon\to 0^+}\left\{\frac{X(t+\epsilon u)-X(t)}{\epsilon^{h_X(t)}}\right\}_{u>0}\overset{(fdd)}{=}\left\{\theta'(t)^{h_X(t)}T_{Y,\theta(t)}(u)\right\}_{u>0}$$

For $u=0$

$$\lim_{\epsilon\to 0^+}\frac{X(t+0\epsilon)-X(t)}{\epsilon^{h_X(t)}}\overset{(d)}{=}0$$

□

Like the Hölder exponent, the local self-similar index is affected by the time change function θ in the sense that the local structure of X at time t is the same as $\tilde{X}$ at time $\theta(t)$.

3.1.3 Index of a process with stationary increments

Since the process $\tilde{X}$ has stationary increments, its pointwise Hölder exponent is almost surely constant on $\mathbb{R}$, see Theorem 2 in [62]. In other words, there exists $H_{\alpha_{\tilde{X}}}\in(0,1)$, so that for all $t\in\mathbb{R}$

$$\mathbb{P}(\alpha_{\tilde{X}}(t)=H_{\alpha_{\tilde{X}}})=1$$

Using the results given by Proposition 3.1.1, we deduce that the pointwise Hölder exponent $\alpha_X(t)=H_{\alpha_{\tilde{X}}}$ for all $t\in\mathbb{R}$. The local index of self-similarity of the process

$\tilde{X}$ is also constant over the time t. Indeed one has for all $t \geq 0$

$$\lim_{\epsilon \to 0^+} \frac{\tilde{X}(t+\epsilon) - \tilde{X}(t)}{\epsilon^{h(t)}} \overset{(d)}{=} \lim_{\epsilon \to 0^+} \epsilon^{h(0)-h(t)} \frac{\tilde{X}(\epsilon) - \tilde{X}(0)}{\epsilon^{h(0)}}$$
$$\overset{(d)}{=} T_{\tilde{X},0}(1) \lim_{\epsilon \to 0^+} \epsilon^{h(0)-h(t)}$$

Since the fractional term converges to a non trivial random variable, we must have $\lim_{\epsilon \to 0^+} \epsilon^{h(0)-h(t)} \in (0, +\infty)$, that is, $h(0) = h(t)$. Hence, the function $h : \mathbb{R} \to (0,1)$ is constant. There exists a constant $H_{h_{\tilde{X}}} \in (0,1)$ so that for all $t \geq 0$, $h(t) = H_{h_{\tilde{X}}}$. We deduce from Proposition 2.3.1

$$H_{\alpha_{\tilde{X}}} \leq H_{h_{\tilde{X}}}$$

Assuming that for all $t \geq 0$, there is $p > 0$ so that

$$\mathbb{E}|\tilde{X}(t+\delta) - \tilde{X}(t)|^p \leq C\delta^{pH_{h_{\tilde{X}}}}$$

Then by the Kolmogorov criterion Theorem (see [55]), the process X has a modification whose sample paths are locally Hölder continuous of index $\alpha \in (0, H_{h_{\tilde{X}}} - 1/p)$. One has $H_{h_{\tilde{X}}} - 1/p \leq H_{\alpha_{\tilde{X}}}$. For processes that have finite moment of every order: e.g. Gaussian processes, one deduces in that case that $H_{\alpha_{\tilde{X}}} = H_{h_{\tilde{X}}}$. Note by Proposition 3.1.1 and 3.1.3, the Hölder exponent and the local self-similarity index of the process $X(t) = \tilde{X}(\theta(t))$ must be constant. Note in practice, a necessary condition for X to be a time change of a lass process with stationary increments, is that its Hölder exponent and its local self-similarity index are constants over the time $t \geq 0$.

3.2 Simulation of time-changed H-sssi processes

3.2.1 α-stable stochastic integrals of a deterministic function

We want to define a stochastic integral with respect to an α-stable Lévy motion, of a deterministic function $f : E \to \mathbb{R}$, where $E \subset \mathbb{R}$, represented by the following stochastic integral in (3.4).

$$I(f) = \int_E f(s) d\tilde{X}(s) \tag{3.4}$$

where

$$\int_E |f(s)|^\alpha ds < \infty \tag{3.5}$$

For stochastic functions f, we refer to works of Rosiński and Woyczyński in [63] and Janiki and Weron in [64]. In this last reference ([64] page 73-74), the authors also define integral (3.4), where f is deterministic. They show for f satisfying (3.5), the stochastic integral (3.4) is well defined.

Setting the simple function $f(t) = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}(t)$, with disjoint sets A_k, Janiki and Weron

showed that $I(f)$ is α-stable. Then, they extended to a more general continuous function f satisfying (3.5) by showing there exists a sequence of simple functions $\{f^{(n)}\}_{n=1}^{\infty}$ such that for almost all $t \in E$

$$f^{(n)}(t) \to f(t)$$

and for any n, $|f^{(n)}(t)| < g(t)$, for some continuous function g satisfying (3.5). The sequence of integrals $\{I(f^{(n)})\}_{n=1,2,...}$ is a Cauchy sequence in the complete space of α-stable random variables with the metric space induced by convergence in probability. So, there exists a random variable $I(f)$ which is the limit $\{I(f^{(n)})\}_{n=1,2,...}$ in this space. Whence the definition

Definition 3.2.1. *[64] An α-stable stochastic integral of any continuous function satisfying (3.5), is by definition*

$$I(f) = \lim_{n\to\infty} I(f^{(n)}) \text{ in probability} \tag{3.6}$$

3.2.2 Time-changed α-stable H-sssi processes

A stochastic integral representation of a random time-changed α-stable Lévy motion is given by Rosiński and Woyczyński in [63]. This stochastic integral defined for V non-negative random process and $\int_0^t V^\alpha(s)ds < \infty$, and $\tilde{X}$ an α-stable motion. More precisely, there exists a process $\tilde{X}' \overset{(fdd)}{=} \tilde{X}$, such that

$$\tilde{X}' \circ \int_0^t V^\alpha(s)ds \overset{(a.s.)}{=} \int_0^t V(s)d\tilde{X}(s) \tag{3.7}$$

Using (3.4) and (3.7), we introduce the integral representation of our time-changed α-stable Lévy motion. This representation is also satisfied for time-changed Brownian motion, since the 2-stable Lévy motion is a Browanian. Substituting in (3.4) the random process V by the deterministic function $\theta'^{1/\alpha}$, there exists a process $\tilde{X}' \overset{(fdd)}{=} \tilde{X}$ so that

$$\tilde{X}'(\theta(t)) \overset{(a.s.)}{=} \int_0^t (\theta'(s))^{1/\alpha} d\tilde{X}(s) \tag{3.8}$$

Integral (3.8) is well defined since the function θ' is positive and $\theta(t) = \int_0^t \theta'(s)ds < \infty$. In order to simulate a time-changed H-sssi stable process, an approximation of integral (3.8) can be defined using Definition 3.2.1. We set the simple function $\theta'^{(n)}$ as

$$\theta'^{(n)}(t) = \sum_{j=0}^{n-1} \theta'(t_j^*)\mathbb{I}_{[t_j,t_{j+1})}(t)$$

where $t_0, \ldots, t_n$ partition of $[0,t]$, and $t_j^* \in [t_j, t_{j+1})$, such that $\theta'(t_j^*) = \dfrac{\theta(t_{j+1}) - \theta(t_j)}{t_{+1} - t_j}$. The mean value Theorem ensures the existence of such t_j^*. Setting $t_0 = 0$ and $t_n = t$,

one has

$$\begin{aligned} I\left((\theta'^{(n)})^{1/\alpha}\right) &= \int_0^t (\theta'^{(n)}(s))^{1/\alpha} d\tilde{X}(s) \\ &= \int_0^t \left(\sum_{j=0}^{n-1} \theta'(t_j^*) \mathbb{I}_{[t_j, t_{j+1})}(s)\right)^{1/\alpha} d\tilde{X}(s) \\ &= \int_0^t \sum_{j=0}^{n-1} \left(\theta'(t_j^*)\right)^{1/\alpha} \mathbb{I}_{[t_j, t_{j+1})}(s) d\tilde{X}(s) \\ &= \sum_{j=0}^{n-1} \left(\theta'(t_j^*)\right)^{1/\alpha} \int_{t_j}^{t_{j+1}} d\tilde{X}(s) \\ &= \sum_{j=0}^{n-1} \left(\theta'(t_j^*)\right)^{1/\alpha} (\tilde{X}(t_{j+1}) - \tilde{X}(t_j)) \end{aligned}$$

By Definition 3.2.1, one has the following limit

$$\int_0^t (\theta'(s))^{1/\alpha} d\tilde{X}(s) \overset{def}{=} \lim_{n\to\infty} \sum_{j=0}^{n-1} \left(\theta'(t_j^*)\right)^{1/\alpha} (\tilde{X}(t_{j+1}) - \tilde{X}(t_j)) \text{ in probability} \quad (3.9)$$

3.2.3 Simulation of time-changed H-sssi process with independent increments

As described in the previous section, the continuous time-changed H-sssi α-stable Lévy motion can be approximated by a discrete time process using (3.9). We define an approximation of the continuous time process $\{X(t)\}_{t\geq 0}$, by the discrete time process $\{X_j^{(n)}\}_{j=0,1,\ldots}$, such that for the increasing sequence of positive real $\{t_j\}_{j=0,1,\ldots}$, where $t_0 = 0$, $X_j^{(n)} = X(t_j)$. The discrete time process is obtained by the following expression

$$X_{j+1}^{(n)} = X_j^{(n)} + \left(\theta'(t_j^*)\right)^{(1/\alpha)} d\tilde{X}_j^{(n)} \quad (3.10)$$

where $d\tilde{X}_j^{(n)} = \tilde{X}(t_{j+1}) - \tilde{X}(t_j)$ and $X_0^{(n)} = X(0) = 0$ almost surely.
More generally, continuous time H-sssi processes with independent increments are approximated by a discrete time process satisfying (3.10). As the process X has independent increments, one has

$$\{\tilde{X}(\theta(t_j)) - \tilde{X}(\theta(t_{j-1}))\}_{j\in\{0,\ldots,n\}} \overset{(fdd)}{=} \left\{ \left(\frac{\theta(t_j) - \theta(t_{j-1})}{t_j - t_{j-1}}\right)^H \left(\tilde{X}(t_j) - \tilde{X}(t_{j-1})\right) \right\}_{j\in\{0,\ldots,n\}}$$

Indeed, by self-similarity, and stationarity and independency of the increments of X,

one has

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}(\theta(t_1)) - \tilde{X}(\theta(t_0)), \tilde{X}(\theta(t_2)) - \tilde{X}(\theta(t_1)), \dots\right) \\
&= \mathbb{P}\left(\tilde{X}(\theta(t_1)) - \tilde{X}(\theta(t_0))\right) \mathbb{P}\left(\tilde{X}(\theta(t_2)) - \tilde{X}(\theta(t_1))\right) \dots \\
&= \mathbb{P}\left(\left(\frac{\theta(t_1) - \theta(t_0)}{t_1 - t_0}\right)^H \left(\tilde{X}(t_1) - \tilde{X}(t_0)\right)\right) \mathbb{P}\left(\left(\frac{\theta(t_2) - \theta(t_1)}{t_2 - t_1}\right)^H \left(\tilde{X}(t_2) - \tilde{X}(t_1)\right)\right) \dots \\
&= \mathbb{P}\left(\left(\frac{\theta(t_1) - \theta(t_0)}{t_1 - t_0}\right)^H \left(\tilde{X}(t_1) - \tilde{X}(t_0)\right), \left(\frac{\theta(t_2) - \theta(t_1)}{t_2 - t_1}\right)^H \left(\tilde{X}(t_2) - \tilde{X}(t_1)\right), \dots\right)
\end{aligned}$$

Setting $t_0 = 0$, $t_n = t$ and $\theta'(t_j^*) = \dfrac{\theta(t_{j+1}) - \theta(t_j)}{t_{+1} - t_j}$, one has

$$X(t_j) = \sum_{i=1}^{j} X(t_i) - X(t_{i-1}) = \sum_{i=1}^{j} \tilde{X}(\theta(t_i)) - \tilde{X}(\theta(t_{i-1})) \overset{(d)}{=} \sum_{i=1}^{j} (\theta'(t_i^*))^H \left(\tilde{X}(t_i) - \tilde{X}(t_{i-1})\right)$$

We define by construction the discrete time H-sssi process $\{X_j^{(n)}\}_{j=0,1,\dots}$ with independent increments, such that

$$X_{j+1}^{(n)} = X_j^{(n)} + \left(\theta'(t_j^*)\right)^H \left(\tilde{X}(t_{j+1}) - \tilde{X}(t_j)\right)$$

However, this last formula is only true for H-sssi processes with independent increments. The next section will introduce an approximation of simulating time-changed H-sssi with dependent increments.

3.2.4 Simulation of time-changed H-sssi process with dependent increments

We assume the process $X(t) = \tilde{X}(\theta(t))$, where the process $\tilde{X}$ is H-sssi and $\theta \in \Theta(\mathbb{R}^+)$. We want to simulate the discrete time process $\{X_j^{(n)}\}_{j=0,1,\dots}$ on a regular time space, such that for the increasing sequence of positive real $\{t_j\}_{j=0,1,\dots}$, where $t_0 = 0$, one has $X_j^{(n)} = X(t_j)$. Let $\{\tilde{X}_j^{(n)}\}_{j=0,1,\dots}$ be the discrete time process of the process $\{\tilde{X}\}_{t\geq 0}$. We simulate $N+1$ observations of $\tilde{X}$ of time length T, we get $\{\tilde{X}_j^{(n)} = \tilde{X}\left(\frac{j}{N}T\right)\}_{j=0,1,\dots,N}$. Taking the inverse of our function θ, one has for $j = 0, 1, \dots, N$

$$X\left(\theta^{-1}\left(\frac{j}{N}T\right)\right) = \tilde{X}\left(\frac{j}{N}T\right)$$

This means in practice, we simulate observation of $\tilde{X}$ regularly spaced in time; say we have $N+1$ observations of $\tilde{X}$ at times $t_j = \frac{j}{N}T$, for $j = 0, 1, \dots, N$. The process X is obtained by keeping the observation of $\tilde{X}$ at each time t_j and changing the time axis $\{t_j\}$ by $\{\theta^{-1}(t_j)\}$. An illustration is shown in Figure 3.1.
The observation of X are irregularly spaced in time. Regular time-spaced observations

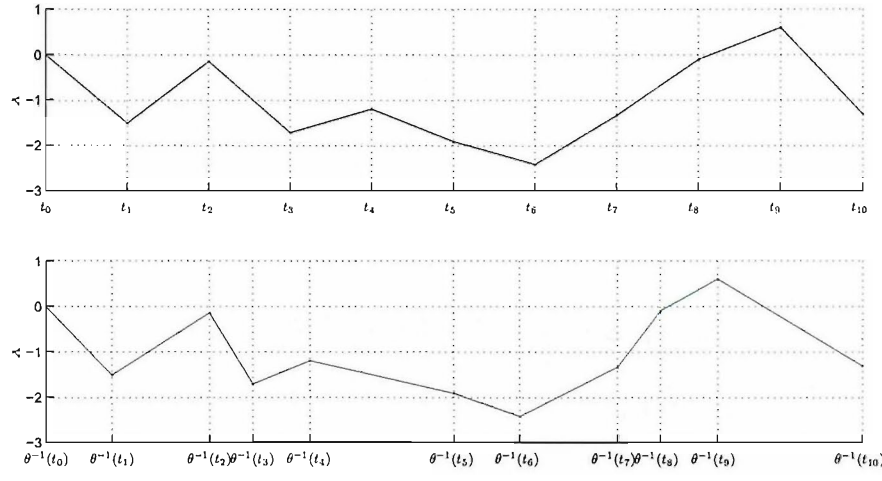

Figure 3.1: Process $\tilde{X}$ with regular observation in time (top panel) and X with observations irregularly spaced in time (bottom panel).

of the process X are obtained by linear interpolation of the observed X. For a positive integer M, let $\{\check{X}(\frac{i}{M}T)\}_{i=0,1,\ldots,M}$ be the discrete time process obtained by this interpolation. Setting $\delta = T/M$, one has when $i \in \left[\dfrac{\theta^{-1}(t_{j_i})}{\delta}, \dfrac{\theta^{-1}(t_{j_i+1})}{\delta}\right)$

$$\check{X}(i\delta) = \frac{i\delta - \theta^{-1}(t_{j_i})}{\theta^{-1}(t_{j_i+1}) - \theta^{-1}(t_{j_i})} \left(X(\theta^{-1}(t_{j_i+1})) - X(\theta^{-1}(t_{j_i}))\right) + X(\theta^{-1}(t_{j_i}))$$

or also

$$\check{X}(i\delta) = \frac{i\delta - \theta^{-1}(t_{j_i})}{\theta^{-1}(t_{j_i+1}) - \theta^{-1}(t_{j_i})} \left(\tilde{X}(t_{j_i+1}) - \tilde{X}(t_{j_i})\right) + \tilde{X}(t_{j_i}) \tag{3.11}$$

In the next Proposition, we compute the interpolation error. This means that we look at the distribution of $|\check{X}(i\delta) - X(i\delta)|$.

Proposition 3.2.2. *Let $\tilde{X}$ be a H-sssi process such that $\mathbb{E}|\tilde{X}(1)|^2 < +\infty$. We suppose having the following observation $\{\tilde{X}(t_i)\}_{t_i \in \mathbb{R}}$, where for $(i,j) \in \mathbb{N}^2$ such that $i < j$, $t_i < t_j$. Let $\{\check{\tilde{X}}(n\delta)\}_{n\in\mathbb{N}}$ a process constructed from $\{\tilde{X}(t_i)\}_{t_i\in\mathbb{R}}$ by linear interpolation such that $t_{i_n} \leq n\delta < t_{i_n+1}$, $(t_{i_n}, t_{i_n+1}) \in \mathbb{R}^2$ and $n \in \mathbb{N}$. Then*

$$\mathbb{E}|\check{\tilde{X}}(n\delta) - \tilde{X}(n\delta)|^2 = \Lambda^H\left(\frac{n\delta - t_{i_n}}{t_{i_n+1} - t_{i_n}}\right)(t_{i_n+1} - t_{i_n})^{2H}\mathbb{E}|\tilde{X}(1)|^2 \tag{3.12}$$

where

$$\forall \lambda \in [0,1], \quad \Lambda^H(\lambda) = \begin{cases} 0 & \text{if } \lambda \in \{0,1\} \\ \lambda(1-\lambda)\left(\lambda^{2H-1} + (1-\lambda)^{2H-1} - 1\right) & \text{if } \lambda \in (0,1) \end{cases} \tag{3.13}$$

Proof. We consider the two successive observations $\tilde{X}(t_i)$ and $\tilde{X}(t_{i+1})$. Let

$$\check{\tilde{X}}(t) = (1-\lambda)\tilde{X}(t_i) + \lambda\tilde{X}(t_{i+1})$$

where $\lambda \in [0,1]$ and $t = (1-\lambda)t_i + \lambda t_{i+1}$.

$$\begin{aligned}\left(\check{\tilde{X}}(t) - \tilde{X}(t)\right)^2 &= \left((1-\lambda)\tilde{X}(t_i) + \lambda\tilde{X}(t_{i+1}) - \tilde{X}(t)\right)^2 \\ &= (1-\lambda)^2\left(\tilde{X}(t_i) - \tilde{X}(t)\right)^2 + \lambda^2\left(\tilde{X}(t_{i+1}) - \tilde{X}(t)\right)^2 \\ &\quad + 2(1-\lambda)\lambda\left(\tilde{X}(t_i) - \tilde{X}(t)\right)\left(\tilde{X}(t_{i+1}) - \tilde{X}(t)\right)\end{aligned}$$

By taking the expectation, one get

$$\begin{aligned}\mathbb{E}|\check{\tilde{X}}(t) - \tilde{X}(t)|^2 &= (1-\lambda)^2\mathbb{E}|\tilde{X}(t_i) - \tilde{X}(t)|^2 + \lambda^2\mathbb{E}|\tilde{X}(t_{i+1}) - \tilde{X}(t)|^2 \\ &\quad + 2(1-\lambda)\lambda\mathbb{E}\left[\left(\tilde{X}(t_i) - \tilde{X}(t)\right)\left(\tilde{X}(t_{i+1}) - \tilde{X}(t)\right)\right]\end{aligned}$$

By self-similarity and stationary increments, using $t = (1-\lambda)t_i + \lambda t_{i+1}$, one gets

$$\begin{aligned}&\mathbb{E}|\tilde{X}(t_i) - \tilde{X}(t)|^2 = \lambda^{2H}(t_{i+1} - t_i)^{2H}\mathbb{E}|\tilde{X}(1)|^2 \\ &\mathbb{E}|\tilde{X}(t_{i+1}) - \tilde{X}(t)|^2 = (1-\lambda)^{2H}(t_{i+1} - t_i)^{2H}\mathbb{E}|\tilde{X}(1)|^2 \\ &\mathbb{E}\left[\left(\tilde{X}(t_i) - \tilde{X}(t)\right)\left(\tilde{X}(t_{i+1}) - \tilde{X}(t)\right)\right] = \frac{1}{2}(\lambda^{2H} + (1-\lambda)^{2H} - 1)(t_{i+1} - t_i)^{2H}\mathbb{E}|\tilde{X}(1)|^2\end{aligned}$$

Whence

$$\mathbb{E}|\check{\tilde{X}}(t) - \tilde{X}(t)|^2 = \Lambda^H(\lambda)(t_{i+1} - t_i)^{2H}\mathbb{E}|\tilde{X}(1)|^2 \tag{3.14}$$

where the function Λ^H is defined in (3.13). Clearly for any $H \in (0,1)$ and $\lambda \in [0,1]$, $\Lambda^H(\lambda) \in [0,1)$. Setting $t = n\delta$ and $i = i_n$, one gets (3.12). □

Corollary 3.2.3. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is a self-similar process such that $\mathbb{E}|\tilde{X}(1)|^2 < +\infty$ and θ a continuous and increasing function. We suppose having the following observation $\{X(t_i)\}_{\{t_i \in \mathbb{R}\}}$, where for $(i,j) \in \mathbb{N}^2$ such that $i < j$, $t_i < t_j$. Let $\{\check{X}(n\delta)\}_{\{n\in\mathbb{N}\}}$ a process constructed from the observation $\{X(t_i)\}_{\{t_i\in\mathbb{R}\}}$ by linear interpolation such that $t_{i_n} \leq n\delta < t_{i_n+1}$, $(t_{i_n}, t_{i_n+1}) \in \mathbb{R}^2$ and $n \in \mathbb{N}$. Assuming θ piecewise linear*[1]*, then*

$$\mathbb{E}|\check{X}(n\delta) - X(n\delta)|^2 = (\theta(t_{i_n+1}) - \theta(t_{i_n}))^{2H}\Lambda^H\left(\frac{n\delta - t_{i_n}}{t_{i_n+1} - t_{i_n}}\right)\mathbb{E}|\tilde{X}(1)|^2 \tag{3.15}$$

Proof. Straightforward from Proposition 3.2.2 □

Figure 3.2 represents the function Λ^H defined in (3.13).
Using Corollary 3.2.3, the interpolation error on the interpolated discrete time process

[1] As we deal with discrete time process. we just consider the function θ being piecewise linear.

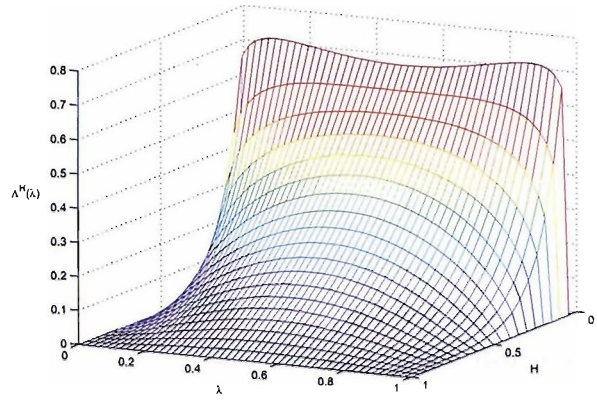

Figure 3.2: $\Lambda^H(\lambda)$ for $H \in (0,1)$ and $\lambda \in [0,1]$.

in (3.11), is

$$\mathbb{E}|\check{X}(i\delta) - X(i\delta)|^2 = (t_{j_i+1} - t_{j_i})^{2H} \Lambda^H \left(\frac{i\delta - \theta^{-1}(t_{j_i})}{\theta^{-1}(t_{j_i+1}) - \theta^{-1}(t_{j_i})} \right) \mathbb{E}|\tilde{X}(1)|^2 \qquad (3.16)$$

Note from (3.16), that to reduce the error, one needs to reduce the interval between t_{j_i} and t_{j_i+1}. If the observations of the process $\tilde{X}$ are regularly spaced in time, one sets $t_{j_i+1} - t_{j_i} = 1/N$. Assuming $U_i = \dfrac{i\delta - \theta^{-1}(t_{j_i})}{\theta^{-1}(t_{j_i+1}) - \theta^{-1}(t_{j_i})}$ is uniformly distributed between 0 and 1, one has

$$\mathbb{E}\left[(\check{X}(i\delta) - X(i\delta))^2 | U_i\right] = \Lambda^H(U_i) \frac{\mathbb{E}|\tilde{X}(1)|^2}{N^{2H}}$$

Whence the expected error in the case where U_i is uniformly distributed

$$\begin{aligned}
\mathbb{E}|\check{X}(i\delta) - X(i\delta)|^2 &= \int_0^1 \mathbb{E}\left[\left(\check{X}(i\delta) - X(i\delta)\right)^2 | U_i = u\right] du \\
&= \frac{\mathbb{E}|\tilde{X}(1)|^2}{N^{2H}} \int_0^1 \Lambda^H(u)\, du \\
&= \left(\frac{1}{(H+1)(2H+1)} - \frac{1}{6} \right) \mathbb{E}\left| \tilde{X}\left(\frac{1}{N}\right) \right|^2
\end{aligned}$$

This shows that the interpolation error becomes more important as H decreases.

3.3 Hurst index estimation of the time-changed H-sssi

Generally, most of the existing methods[2] to estimate the Hurst index assume the process to have stationary increments. However, a time-changed self-similar such that $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-sssi and the function θ is a nonlinear time change

[2]Some of these methods are presented in the first Chapter. A good survey also on the Hurst index estimation given by M.S. Taqqu and V. Teverovsky can be found in [26]

strictly increasing, has no stationary increments. Therefore, the application of the existing methods, for example the detrend fluctuation analysis, the wavelets methods or the R/S analysis, might fail on the self-similarity index estimation. As seen in Section 3.1, the self-similarity index of X is the same as the process $\tilde{X}$ and so the self-similarity index can be estimated from $\tilde{X}$, if this last is known. Unfortunately this is not always the case.

In this section, we present a method for estimating the global index of self-similarity of a time-changed self-similar processes. this is known as the crossing tree method [7]. We show also that there exists a universal estimator introduced by Cohen and Istas in [65]. This last estimator is also adapted to lass processes and time-changed lass processes. We conduct a comparison of the performance of the Hurst index estimators presented in Chapter 1 on time-changed Brownian motion. This test is a fair test since there exists an exact method for simulating time-changed Brownian motion on a regular time grid.

3.3.1 Estimation methods for time-changed self-similar processes

For this application, the crossing tree method for estimating H introduced by O.D. Jones and Y. Shen [7] is well adapted. Indeed, this estimator does not depend on the time-change θ and so one can estimate the index H of $\tilde{X}$ through the process X. We have illustrated this in Figure 3.3. Clearly, the sequence of crossing points is unaffected by a time-change to the process and thus the branching structure of the crossing tree is unaffected by a time-change. The EBP estimator is calculated from the average family size of the crossing tree, so it too is unaffected by a time-change.

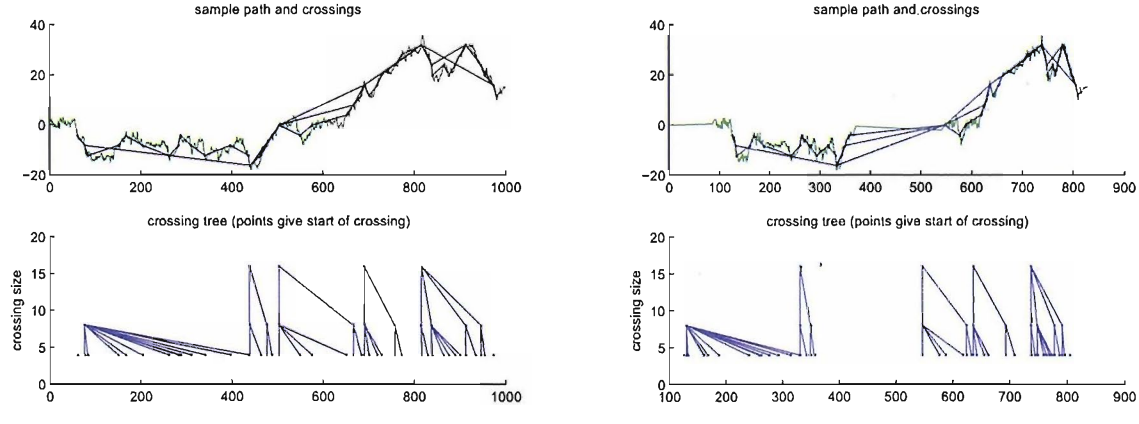

Figure 3.3: The crossing tree of a stochastic process before and after a time change. In each panel the top diagram shows the sample path and three approximations made from crossings of size 4, 8 and 16. The bottom diagram in each panel shows the crossing tree: nodes correspond to crossings and crossings are linked if one is a subcrossing of the other. We see that while the crossing times change, the branching structure of the tree is unaffected by the time change.

For lass processes, the estimator presented in Theorem 1.1 by S. Cohen and J. Istas in

[65] is well adapted to time-changed processes. This estimator is local and so does not depend on the time-change θ. We recall this Theorem and then we show that time-changed lass processes satisfy conditions of Theorem 3.3.1 through Corollary 3.3.2.

Theorem 3.3.1. *[65] Let Z be a process so that there exists a random variables Y such that*

$$\lim_{t,t'\to t_0} \frac{Z(t')-Z(t)}{|t'-t|^H} \stackrel{(d)}{=} Y \tag{3.17}$$

and that the family of $\log^2\left(\frac{|Z(t')-Z(t)|}{|t'-t|^H}\right)$ *are uniformly integrable in neighborhood of t_0. We set* $W_n = -\frac{1}{n v_n}\sum_{k\in\mathcal{V}_n} \log_2|\Delta_{k,n}Z|$ *where for* $t_0\in\mathbb{R}$,

$$\mathcal{V}_n = \left\{k\in\mathbb{Z}\ s.t.\ \left|\frac{k}{2^n}-t_0\right|\le \epsilon_n\right\}$$

for $\epsilon_n > 2^{-n}$, $v_n = \#\mathcal{V}_n$ and $\Delta_{k,n}Z = Z\left(\frac{k+1}{2^n}\right) - Z\left(\frac{k}{2^n}\right)$. Assume $\epsilon_n \to 0$ as $n\to+\infty$. Then one has

$$\lim_{n\to+\infty} W_n \stackrel{(a.s.)}{=} H$$

Corollary 3.3.2. *Assume that we have a process of the form $X(t) = \tilde{X}(\theta(t))$ where $\tilde{X}$ is a lass process so that (3.17) is satisfied and* $\log^2\left(\frac{|\tilde{X}(t')-\tilde{X}(t)|}{|t'-t|^H}\right)$ *are uniformly integrable in neighborhood of t_0 and $\theta\in\Theta(\mathbb{R}^+)$ so that in neighborhood of t_0*

$$\sup_{t,t'}\left|\log\left|\frac{\theta(t')-\theta(t)}{t'-t}\right|^H\right| < +\infty$$

Then X satisfies Theorem 3.3.1.

Proof. The process X satisfies (3.17) since

$$\lim_{t,t'\to t_0}\frac{X(t')-X(t)}{|t'-t|^H} = \lim_{t,t'\to t_0}\frac{\tilde{X}(\theta(t'))-\tilde{X}(\theta(t))}{|\theta(t')-\theta(t)|^H}\left|\frac{\theta(t')-\theta(t)}{t'-t}\right|^H$$
$$\stackrel{(d)}{=} \theta'(t_0)^H T_{\tilde{X},\theta(t_0)}$$

and also

$$\log\left(\frac{|X(t')-X(t)|}{|t'-t|^H}\right) = \log\left(\frac{|\tilde{X}(\theta(t'))-\tilde{X}(\theta(t))|}{|\theta(t')-\theta(t)|^H}\right) + \log\left|\frac{\theta(t')-\theta(t)}{t'-t}\right|^H$$

and $\log\left|\frac{\theta(t')-\theta(t)}{t'-t}\right|^H$ is bounded in neighborhood of t_0 and $\log^2\left(\frac{|X(t')-X(t)|}{|t'-t|^H}\right)$ is uniformly integrable in neighborhood of t_0. Whence X satisfies Theorem 3.3.1. □

We showed that the universal estimator of local self-similarity of S.Cohen and J.Istas

in [65] are adapted to time-changed lass processes, which is also adapted to self-similar processes. In practice we will use the EBP estimator.

3.3.2 Estimation of the self-similarity index of a time changed H-sssi

The aim is to estimate the index of self-similarity on time changed processes. We simulate in this section some time-changed H-sssi processes, using the procedure described in Section 3.2. We use the DFA, the Wavelet and R/S to estimate H. In particular, these estimators are going to be performed on time-changed Brownian motion since it has an exact simulation. We will not test these estimators on time-changed fractional Brownian motion. The interpolation method to get such a process on a regular time grid induce some error on the Hurst index estimate. The linear interpolation tends to smooth the path of the process, and so the Hurst index can be overestimated. This study on the time change effect on the estimators makes no sense in that case. We define the function θ as follows

$$\theta(t) = \frac{\cos(\pi/2 + 4\pi t) - 3/4\sin(\pi/2 + 8\pi t) + (10+\epsilon)\pi t + 3/4}{(10+\epsilon)\pi} \tag{3.18}$$

The quantity ϵ is to ensure that the derivative of the function θ is strictly positive for any $t \in [0,1]$. The function θ satisfies $\theta(0) = 0$, $\theta(1) = 1$ and for all $t \in [0,1]$, $\theta'(t) > 0$. We compare the DFA, Wavelet and R/S estimators before and after time changing a Brownian motion. We perform a Monte-Carlo experiment: we simulate 1000 paths of time-changed Browanian motion $B(\theta(t))$. Because the estimators perform better when the number of observations is a power of 2, we simulate 2^{12} observations for each path. For comparison, we also estimate the Hurst of Brownian motion B'. The mean and twice the standard deviation of the experiment are given in Table 3.1.

	DFA	WAV	R/S
$B'(t)$ with regular time space	0.4971 (0.0686)	0.4990 (0.0429)	0.5637 (0.0216)
$B(\theta(t))$ with regular time space	0.5039 (0.0729)	0.5006 (0.0485)	0.5615 (0.0304)

Table 3.1: Hurst estimation on time-changed Brownian motion. The number in brackets, represent twice the standard error.

Note that $B' \overset{(fdd)}{=} B$, and so it is normal that we do not obtain the same estimation for both processes. Note that the estimators presented here seem to not be affected by the time change function theta.

Chapter 4

Time-change estimation using path variation

We consider a process X of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is locally self-similar with stationary increments and index H (H-lsssi), and $\theta \in \Theta(\mathrm{T})$, where for a compact subset of the positive real line T, the set $\Theta(\mathrm{T})$ is defined by (4.1).

$$\Theta(\mathrm{T}) = \left\{\theta : \mathrm{T} \to \mathrm{T} \in \mathcal{C}^1, \quad s.t. \quad \forall t \in \mathrm{T},\, \theta'(t) > 0 \text{ and } \theta(0) = 0\right\} \tag{4.1}$$

Having the process X, our aim is to estimate the function θ. In this Chapter, we present two methods based on the path variation of the process to estimate the time-change function θ. The first method uses the p-variation ($p > 0$) of the process, this last is adapted to continuous processes such as the fractional Brownian motion. The second method uses the log-variation, which is suited to heavy-tailed processes (for example if $\tilde{X}$ is an α-stable Lévy process). In fact our estimators do not estimate θ directly but its derivative θ'.

Our chapter will be divided into six Sections. In the first section, we present new notations and some preliminary results, which are going to be used in our work. In the second and the third section, we introduce respectively the p-variation and the log-variation time-change estimators. For both estimators, we study their convergence, which will depend on the structure of dependency of the time-changed process. In the fourth and fifth Section, we analyse the bias and the consistency of the estimators. This study will be especially focused on time-changed H-sssi processes. We finish then by applying our estimators on some H-sssi processes to test their performances.

4.1 Notation and preliminary results

4.1.1 Definition of the time-change function θ

We consider the stepwise function $\theta'^{(n)}$ defined in (4.2). Let $\{t_0, \ldots, t_i, \ldots, t_n\}$ be a sequence of $\mathrm{T} = [0, T]$, such that $t_0 = 0$, $t_n = T$ and $\forall (i, j) \in \{0, 1, \ldots, n\}^2$ ($n \in \mathbb{N}$)

and $i < j$, $t_i < t_j$. The stepwise function $\theta'^{(n)}$ is defined so that for $t \in \mathrm{T}$

$$\theta'^{(n)}(t) = \sum_{j=0}^{n-1} \theta'(t_j^*)\mathbb{I}_{[t_j,t_{j+1})}(t) \tag{4.2}$$

where $\mathbb{I}_{[t_j,t_{j+1})}(t) = 1$ if $t \in [t_j, t_{j+1})$, $\mathbb{I}_{[t_j,t_{j+1})}(t) = 0$ otherwise; and $t_j^* \in [t_j, t_{j+1})$, such that

$$\theta'(t_j^*) = \frac{\theta(t_{j+1}) - \theta(t_j)}{t_{j+1} - t_j}$$

Such θ_j^* exists by the mean value theorem. Since the function θ' is continuous on the compact T, it is easy to show that $\lim_{n\to+\infty} \theta'^{(n)}(t) = \theta'(t)$, $\forall t \in \mathrm{T}$.
Both estimators presented in this chapter, estimate the derivative function $\theta'^{(n)}$, which is equivalent to estimate the sequence[1] of the $\{\theta'(t_j^*)\}_{j\in[0,n-1]}$, for $t_j^* \in [t_j, t_{j+1})$.

4.1.2 Notation

We set for any random process Y the two following quantities respectively called the p-variation ($p > 0$) and the log-variation of the process Y

$$V_n^{H,p}(Y, t, \delta t) = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \left|\Delta_{n,k}^H Y(t, \delta t)\right|^p \tag{4.3}$$

and

$$U_n^H(Y, t, \delta t) = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \log \left|\Delta_{n,k}^H Y(t, \delta t)\right| \tag{4.4}$$

Where $\Delta_{n,k}^H Y(t, \delta t) = \dfrac{Y\left(t + (k+1)\delta_n\right) - Y\left(t + k\delta_n\right)}{\delta_n^H}$, $\delta t > 0$, $\delta_n = \dfrac{\delta t}{2^n}$ and H the index of self-similarity estimated using the crossing tree method described in Section 1.1.4.
For simplicity of notation and also to avoid big expressions, while we prove the convergence of our estimators, we define a certain number of expressions. We denote by $c_Y^V(k, \gamma, p)$ and $c_Y^U(k, \gamma)$ for $\gamma > 0$, the following covariance functions

$$c_Y^V(k, \gamma, p) = Cov\left(\left|\frac{Y(\gamma) - Y(0)}{\gamma^H}\right|^p, \left|\frac{Y((k+1)\gamma) - Y(k\gamma)}{\gamma^H}\right|^p\right) \tag{4.5}$$

and

$$c_Y^U(k, \gamma) = Cov\left(\log\left|\frac{Y(\gamma) - Y(0)}{\gamma^H}\right|, \log\left|\frac{Y((k+1)\gamma) - Y(k\gamma)}{\gamma^H}\right|\right) \tag{4.6}$$

Note that when Y is an H-sssi process, covariances (4.5) and (4.6) become independent of the constant γ. One gets $c_Y^V(k, \gamma, p) = Cov\left(|Y(1)|^p, |Y(k+1) - Y(k)|^p\right)$ and

[1] Note that this sequence represents the "averaged" derivative in the intervals $\{[t_j, t_{j+1})\}$, which means later for our financial application, that the estimators will estimate the averaged activity of a given asset between two dates.

$c_Y^U(k,\gamma) = Cov\left(\log|Y(1)|, \log|Y(k+1) - Y(k)|\right)$. We set also

$$v_Y^V(k,\gamma,p) = \mathbb{V}ar\left(\left|\frac{Y((k+1)\gamma) - Y(k\gamma)}{\gamma^H}\right|^p\right) \tag{4.7}$$

and

$$v_Y^U(k,\gamma) = \mathbb{V}ar\left(\log\left|\frac{Y((k+1)\gamma) - Y(k\gamma)}{\gamma^H}\right|\right) \tag{4.8}$$

As for the covariance, when the process Y is H-sssi , (4.7) and (4.8) are reduced to $v_Y^V(k,\gamma,p) = \mathbb{V}ar\left(|Y(1)|^p\right)$ and $v_Y^U(k,\gamma) = \mathbb{V}ar\left(\log|Y(1)|\right)$. For simplicity we define the following means and the standard deviations: $\mu_Y^V(p) = \mathbb{E}|Y(1)|^p$, $\mu_Y^U = \mathbb{E}\log|Y(1)|$, $\sigma_Y^V(p) = \sqrt{\mathbb{V}ar\left(|Y(1)|^p\right)}$ and $\sigma_Y^U = \sqrt{\mathbb{V}ar\left(\log|Y(1)|\right)}$.

4.1.3 Preliminary result

We introduce the following Lemma, concerning time-changed locally self-similar process.

Lemma 4.1.1. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi and $\theta \in \Theta(\mathrm{T})$. Let $B(t_0,\eta)$ the open ball centered in t_0 and of radius η so that $B(t_0,\eta) \subset \mathrm{T}$. Then for all $t, t' \in B(t_0,\eta)$*

$$\lim_{t,t'\to t_0} \frac{X(t) - X(t')}{|t-t'|^H} \overset{(d)}{=} \theta'(t_0)^H T_{\tilde{X},0}(1) \tag{4.9}$$

Proof. One has

$$\begin{aligned}\lim_{t,t'\to t_0} \frac{X(t) - X(t')}{|t-t'|^H} &= \lim_{t,t'\to t_0} \frac{\tilde{X}(\theta(t)) - \tilde{X}(\theta(t'))}{|t-t'|^H} \\ &= \lim_{t,t'\to t_0} \frac{\tilde{X}(\theta(t)) - \tilde{X}(\theta(t'))}{|\theta(t)-\theta(t')|^H} \left|\frac{\theta(t)-\theta(t')}{t-t'}\right|^H\end{aligned}$$

Since $\theta'(t) > 0$ for any $t \in \mathrm{T}$,and θ is $\mathcal{C}^1$, one has

$$\lim_{t,t'\to t_0} \left|\frac{\theta(t)-\theta(t')}{t-t'}\right|^H = \theta'(t_0)^H$$

Moreover as $(t,t') \to (t_0,t_0)$ uniformly, $(\theta(t),\theta(t')) \to (\theta(t_0),\theta(t_0))$ uniformly so by stationarity of the increments of $\tilde{X}$

$$\lim_{t,t'\to t_0} \frac{\tilde{X}(\theta(t)) - \tilde{X}(\theta(t'))}{|\theta(t)-\theta(t')|^H} \overset{(d)}{=} \lim_{t,t'\to t_0} \frac{\tilde{X}(\theta(t)-\theta(t')) - \tilde{X}(0)}{|\theta(t)-\theta(t')|^H} \overset{(d)}{=} T_{\tilde{X},0}(1)$$

Whence

$$\lim_{t,t'\to t_0} \frac{X(t) - X(t')}{|t-t'|^H} \overset{(d)}{=} \theta'(t_0)^H T_{\tilde{X},0}(1)$$

□

4.2 Time-changed estimation using the p-varition

The p-variation method is used to estimate the time change function θ from the process $X(t) = \tilde{X}(\theta(t))$. Our estimator depends on the nature of the locally self-similar process $\tilde{X}$. We decide then to separate two cases. We first consider the process $\tilde{X}$ having stationary increments, and such that the covariance function of the absolute value of its increments power p satisfies

$$\sup_{\gamma\in(0,1)} c_{\tilde{X}}^{V}(k,\gamma,p) \leq O\left(\frac{1}{k^{4(1-H)}}\right) \text{ as } k \to +\infty$$

For the second case, we consider the process $\tilde{X}$ having stationary and ergodic increments. We will show that in the first case, the estimator has an almost surely conveniences while in the second case the estimator will converge in probability to the result. Let $\theta \in \Theta(\mathrm{T})$, we set

$$\theta(t) = \lambda_{\lfloor t/\delta t\rfloor,\delta t}(t - \lfloor t/\delta t\rfloor\delta t) + \sum_{j=0}^{\lfloor t/\delta t\rfloor-1} \lambda_{j,\delta t}\delta t \tag{4.10}$$

where $\lambda_{\lfloor t/\delta t\rfloor,\delta t}$ is a positif constant depending on t and δt. (4.10) can be rewritten as

$$\theta(t) = \lambda_{\lfloor t/\delta t\rfloor,\delta t}(t - \lfloor t/\delta t\rfloor\delta t) + \theta(\lfloor t/\delta t\rfloor\delta t) \tag{4.11}$$

Our estimator will estimate the sequence of $\{\lambda_{j,\delta t}\}_j$. For short, we set $\lambda_{t,\delta t} = \lambda_{\lfloor t/\delta t\rfloor,\delta t}$. The following Lemma will be useful later.

Lemma 4.2.1. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi and θ defined by (4.10). Then*

$$V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t) \overset{(d)}{=} {\lambda_{t,\delta t}}^{pH} V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t) \tag{4.12}$$

Proof. We use the fact that for two sequences of random variables $\{X_i\}_i$ and $\{Y_i\}_i$ so that $\{X_i\}_i \overset{(fdd)}{=} \{Y_i\}_i$, then $\sum_i X_i \overset{(d)}{=} \sum_i Y_i$.

$$\begin{aligned}
&V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t) \\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1} \left|\frac{X(\lfloor t/\delta t\rfloor\delta t + (k+1)\delta_n) - X(\lfloor t/\delta t\rfloor\delta t + k\delta_n)}{{\delta_n}^H}\right|^p \\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1} \left|\frac{\tilde{X}(\theta(\lfloor t/\delta t\rfloor\delta t) + \lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\theta(\lfloor t/\delta t\rfloor\delta t) + \lambda_{t,\delta t}k\delta_n)}{{\delta_n}^H}\right|^p \\
&\overset{(d)}{=} \frac{1}{2^n}\sum_{k=0}^{2^n-1} {\lambda_{t,\delta t}}^{pH} \left|\frac{\tilde{X}(\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\lambda_{t,\delta t}k\delta_n)}{\lambda_{t,\delta t}^H{\delta_n}^H}\right|^p \\
&= {\lambda_{t,\delta t}}^{pH} V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)
\end{aligned}$$

□

4.2.1 Time-changed H-lsssi process with condition on covariance

Theorem 4.2.2. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi and θ defined by (4.10). We assume for some $\alpha > 2p$ that for $(t, t') \in \mathrm{T}$*

$$\sup_{t,t' \in \mathrm{T}} \mathbb{E}\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t - t'|^H}\right)^{\alpha} < +\infty \tag{4.13}$$

and

$$\sup_{\gamma \in (0,1)} c_{\tilde{X}}^{V}(k, \gamma, p) \leq O\left(\frac{1}{k^{4(1-H)}}\right) \text{ as } k \to +\infty \tag{4.14}$$

Then

$$\left(\frac{V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)}{\mu^{V}_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}} \overset{(a.s.)}{\to} \lambda_{t,\delta t}, \text{ as } n \to +\infty \tag{4.15}$$

Proof. By Lemma 4.2.1, the expected value of $V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)$ is given by

$$\mathbb{E}\left[V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right] = \lambda_{t,\delta t}^{pH} \mathbb{E}\left[V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right]$$

By assumption (4.13), for $p < 2p < \alpha$, $\forall t, t' \in \mathrm{T}$, $\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t - t'|^H}\right)^{p}$ is uniformly integrable, one has by Lemma 4.1.1

$$\lim_{n \to +\infty} \mathbb{E}\left|\frac{\tilde{X}(\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\lambda_{t,\delta t} k \delta_n)}{\lambda_{t,\delta t}^{H} {\delta_n}^{H}}\right|^{p} = \mathbb{E}\left|T_{\tilde{X},0}(1)\right|^{p}$$

the limit is uniformly in k, so

$$\lim_{n \to +\infty} \mathbb{E}\left[V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right] = \lim_{n \to +\infty} \lambda_{t,\delta t}^{pH} \mathbb{E}\left[V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right] = \lambda_{t,\delta t}^{pH} \mathbb{E}|T_{\tilde{X},0}(1)|^{p} \tag{4.16}$$

We now show the following limit to complete the proof

$$\lim_{n \to +\infty} \left|V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t) - \mathbb{E}\left[V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right]\right| \overset{(a.s.)}{=} 0 \tag{4.17}$$

By Lemma 4.2.1, the variance of $V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)$ is given by

$$\mathbb{V}ar\left(V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right) = \lambda_{t,\delta t}^{2pH} \mathbb{V}ar\left(V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right)$$

where

$$\mathbb{V}ar\left(V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right) = \frac{1}{2^{2n}} \sum_{k=0}^{2^n - 1} v_{\tilde{X}}^{V}(k, \lambda_{t,\delta t}\delta_n, p) + \frac{2}{2^{2n}} \sum_{k=1}^{2^n - 1} (2^n - k) c_{\tilde{X}}^{V}(k, \lambda_{t,\delta t}\delta_n, p)$$

By assumption (4.13), one has for $p < \alpha/2$, $\sup_{n,k} v_{\tilde{X}}^V(k, \lambda_{t,\delta t}\delta_n, p) < +\infty$. We set for $\epsilon > 0$

$$p_n = \mathbb{P}\left(|V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t) - \mathbb{E}[V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)]| \geq \epsilon\right)$$

By the Tchebychev inequality, one has

$$\begin{aligned}
p_n &\leq \frac{\lambda_{t,\delta t}^{2pH} \mathbb{V}ar\left(V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right)}{\epsilon^2} \\
&\leq \lambda_{t,\delta t}^{2pH}\left(\frac{1}{2^{2n}\epsilon^2}\sum_{k=0}^{2^n-1} v_X^V(k, \lambda_{t,\delta t}\delta_n, p) + \frac{2}{2^{2n}\epsilon^2}\sum_{k=1}^{2^n-1}(2^n - k)c_X^V(k, \lambda_{t,\delta t}\delta_n, p)\right) \\
&\leq \lambda_{t,\delta t}^{2pH}\left(\frac{1}{2^n\epsilon^2}\sup_{n,k} v_X^V(k, \lambda_{t,\delta t}\delta_n, p) + \frac{2}{2^{2n}\epsilon^2}\sum_{k=1}^{2^n-1}(2^n - k)O\left(\frac{1}{k^{4(1-H)}}\right)\right) \\
&\leq \lambda_{t,\delta t}^{2pH}\left(\frac{1}{2^n\epsilon^2}\sup_{n,k} v_X^V(k, \lambda_{t,\delta t}\delta_n, p) + \frac{2}{2^n\epsilon^2}\sum_{k=1}^{2^n-1} O\left(\frac{1}{k^{4(1-H)}}\right) - \frac{2}{2^{2n}\epsilon^2}\sum_{k=1}^{2^n-1} O\left(\frac{1}{k^{3-4H}}\right)\right)
\end{aligned}$$

We denote $a_n \sim b_n$ when $a_n/b_n \to C \in (0, +\infty)$ as $n \to +\infty$. One has

$$\sum_{k=0}^{2^n-1} 1/k^\alpha \sim \begin{cases} 2^{n(1-\alpha)} & \text{if } \alpha < 1 \\ n\log(2) & \text{if } \alpha = 1 \\ 1 & \text{if } \alpha > 1 \end{cases}$$

hence knowing $\lambda_{t,\delta t}^{2pH} < \infty$

- For $H \in (0, 1/2)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{1}{2^{2n}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H = 1/2$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{n}{2^{2n}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H \in (1/2, 3/4)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{1}{2^{4n(1-H)}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H = 3/4$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{n}{2^n}) + O(\frac{1}{2^n})$. This implies $p_n = O(\frac{n}{2^n})$
- For $H \in (3/4, 1)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^{4n(1-H)}}) + O(\frac{1}{2^{4n(1-H)}})$. This implies $p_n = O(\frac{1}{2^{4n(1-H)}})$

Whence for each case, $\sum_{n=1}^{\infty} p_n < +\infty$, which implies (4.17) by the Borel-Cantelli Lemma. Hence using (4.16) and by continuity of the function $x \mapsto \left(\frac{x}{\mathbb{E}|T_{\tilde{X},0}(1)|^p}\right)^{\frac{1}{pH}}$, the continuous mapping Theorem yields

$$\left(\frac{V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)}{\mathbb{E}|T_{\tilde{X},0}(1)|^p}\right)^{\frac{1}{pH}} \xrightarrow{(a.s.)} \lambda_{t,\delta t}, \text{ as } n \to +\infty$$

□

Note for $\tilde{X}$ self-similar processes, assumption (4.13) of Theorem 4.2.2 is reduced to

$$\mathbb{E}|\tilde{X}(1)|^{\alpha} \sup_{t,t' \in \mathrm{T}} \left(\frac{\theta(t) - \theta(t')}{t - t'} \right)^{\alpha H} < +\infty, \text{ for } \alpha > 2p$$

which is equivalent to assume $\mathbb{E}|\tilde{X}(1)|^{\alpha} < +\infty$, as the function θ is bounded on the compact T. Similarly, assumption (4.14) is reduced to

$$Cov\left(\left| \tilde{X}(1) \right|^{p}, \left| \tilde{X}(k+1) - \tilde{X}(k) \right|^{p} \right) \leq O\left(\frac{1}{k^{4(1-H)}} \right) \text{ as } k \to +\infty$$

4.2.2 Time-changed H-lsssi process with ergodic increments

For a process $\tilde{X}$, which does not satisfy assumptions (4.13) of Theorem 4.2.2, we consider instead the ergodicity of its increments. An example of a process that has stationary and ergodic increments is the fractional Brownian motion (see [66] for the ergodicity of the fBm). For $\tilde{X}$ H-lsssi process, we assume that its tangent process $T_{\tilde{X}}$ has ergodic increments. We split the case where $\tilde{X}$ is a H-sssi with ergodic increments (see Proposition 4.2.3), and the case where $\tilde{X}$ is a H-lsssi with tangent process $T_{\tilde{X}}$ having ergodic increments (see Proposition 4.2.4).

Proposition 4.2.3. *Let $\tilde{X}$ be a H-sssi with ergodic increments such that for some $\alpha \geq p$, $\mathbb{E}\left|\tilde{X}(1)\right|^{\alpha} < +\infty$. Then*

$$\left(\frac{V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)}{\mu_{\tilde{X}}^{V}(p)} \right)^{\frac{1}{pH}} \xrightarrow{(\mathcal{P})} \lambda_{t,\delta t}, \text{ as } n \to +\infty$$

Proof. By Lemma4.2.1 and the self-similarity of the process $\tilde{X}$, one has

$$\begin{aligned} V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t) &\overset{(d)}{=} \lambda_{t,\delta t}^{pH} V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t) \\ &\overset{(d)}{=} \frac{\lambda_{t,\delta t}^{pH}}{2^n} \sum_{k=0}^{2^n-1} |\tilde{X}(k+1) - \tilde{X}(k)|^{p} \end{aligned}$$

By the stationarity and the ergodicity of the increments of the process $\tilde{X}$, one gets by the ergodic Theorem [67]

$$\frac{1}{2^n} \sum_{k=0}^{2^n-1} |\tilde{X}(k+1) - \tilde{X}(k)|^{p} \xrightarrow{(a.s.)} \mathbb{E}|\tilde{X}(1)|^{p}, \text{ as } n \to +\infty$$

Whence

$$\lim_{n \to +\infty} V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t) \overset{(\mathcal{P})}{=} \lambda_{t,\delta t}^{pH} \mathbb{E}|\tilde{X}(1)|^{p}$$

Hence by continuity of the function $x \to \left(\frac{x}{\mathbb{E}|\tilde{X}(1)|^{p}} \right)^{\frac{1}{pH}}$, Corollary 2 page 31 in [68]

yields

$$\left(\frac{V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)}{\mathbb{E}|\tilde{X}(1)|^p}\right)^{\frac{1}{pH}} \overset{(\mathcal{P})}{\to} \lambda_{t,\delta t}, \text{ as } n \to +\infty$$

□

Proposition 4.2.4. *Let $\tilde{X}$ be H-lass process with stationary increments so that its tangent process $T_{\tilde{X},0}$ has stationary and ergodic increments and for some $\alpha \geq p$, $\mathbb{E}|T_{\tilde{X},0}(1)|^\alpha < +\infty$. Then*

$$\lim_{n\to+\infty}\lim_{\delta t\to 0}\left(\frac{V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}} \overset{(\mathcal{P})}{=} \lambda_{t,\delta t}$$

Before proving Proposition 4.2.4, we introduce the following Lemma.

Lemma 4.2.5. *For any $t \geq 0$*

$$\lim_{\epsilon\to 0^+}\left\{\frac{\tilde{X}(t+\epsilon(u+1)) - \tilde{X}(t+\epsilon u)}{\epsilon^H}\right\}_{u\geq 0} \overset{(fdd)}{=} \{T_{\tilde{X},0}(u+1) - T_{\tilde{X},0}(u)\}_{u\geq 0} \tag{4.18}$$

Proof.

$$\begin{aligned}
&\lim_{\epsilon\to 0^+}\left\{\frac{\tilde{X}(t+\epsilon(u+1)) - \tilde{X}(t+\epsilon u)}{\epsilon^H}\right\}_{u\geq 0}\\
&\overset{(fdd)}{=} \lim_{\epsilon\to 0^+}\left\{\frac{\tilde{X}(\epsilon(u+1)) - \tilde{X}(\epsilon u)}{\epsilon^H}\right\}_{u\geq 0}\\
&= \lim_{\epsilon\to 0^+}\left\{\frac{\tilde{X}(\epsilon(u+1)) - \tilde{X}(0)}{\epsilon^H} - \frac{\tilde{X}(\epsilon u) - \tilde{X}(0)}{\epsilon^H}\right\}_{u\geq 0}\\
&\overset{(fdd)}{=} \{T_{\tilde{X},0}(u+1) - T_{\tilde{X},0}(u)\}_{u\geq 0}
\end{aligned}$$

□

Proof. (of Proposition 4.2.4). Here we use the fact that for a finite sequence of random variables $\{X_i^{(n)}\}_i \overset{(fdd)}{\to} \{X_i\}_i$ as $n \to +\infty$, then $\sum_i X_i^{(n)} \overset{(d)}{\to} \sum_i X_i$ as $n \to +\infty$. By Lemma 4.2.5, one has

$$\begin{aligned}
\lim_{\delta t\to 0} V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t) &\overset{(d)}{=} \lim_{\delta t\to 0} \lambda_{t,\delta t}^{pH} V_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\\
&= \frac{\lambda_{t,\delta t}^{pH}}{2^n}\sum_{k=0}^{2^n-1}\lim_{\delta t\to 0}\left|\frac{\tilde{X}(\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\lambda_{t,\delta t}k\delta_n)}{(\lambda_{t,\delta t}\delta_n)^H}\right|^p\\
&\overset{(d)}{=} \frac{\lambda_{t,\delta t}^{pH}}{2^n}\sum_{k=0}^{2^n-1}|T_{\tilde{X},0}(k+1) - T_{\tilde{X},0}(k)|^p
\end{aligned}$$

Moreover, by the stationarity and the ergodicity of the increments of the process $T_{\tilde{X},0}$, one gets by the ergodic Theorem [67],

$$\frac{1}{2^n}\sum_{k=0}^{2^n-1}|T_{\tilde{X},0}(k+1)-T_{\tilde{X},0}(k)|^p \overset{(a.s.)}{\to} \mathbb{E}|T_{\tilde{X},0}(1)|^p \text{ as } n\to+\infty$$

Hence by continuity of the function $x\to x^{\frac{1}{pH}}$, Corollary 1 page 31 in [68] implies

$$\lim_{n\to+\infty}\lim_{\delta t\to 0}\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{1/pH} \overset{(d)}{=} \lambda_{t,\delta t}\left(\mathbb{E}|T_{\tilde{X},0}(1)|^p\right)^{1/pH}$$

Whence

$$\lim_{n\to+\infty}\lim_{\delta t\to 0}\left(\frac{V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)}{\mathbb{E}|T_{\tilde{X},0}(1)|^p}\right)^{\frac{1}{pH}} \overset{(\mathbb{P})}{=} \lambda_{t,\delta t}$$

□

4.2.3 Trend effect on the estimator

Our motivation for considering the effect of trend on the estimator comes from the fact that, in practice, time series exhibit trends. Our aim is to investigate the effect of this trend on the p-variation estimators. We consider the process $X^f(t)=X(t)+f(t)$, where the process X is of the form $\tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi , θ as defined in (4.10), and f is Hölder continuous function with a global index $\alpha_f > H$, so that for $(t,s)\in \mathrm{I},\mathrm{I}$ a compact of $\mathbb{R}$, there exists a positive constant C such that $|f(t)-f(s)|\leq C|t-s|^{\alpha_f}$.

Proposition 4.2.6. *Assuming the process $X^f(t)=X(t)+f(t)$, where $X(t)$ is of the form $X(t)=\tilde{X}(\theta(t))$, f an α_f-Hölder continuous function on $\mathbb{R}$ such that $\alpha_f > H$ and that $V_n^{H,p}(X,t,\delta t)$ converges. Then*

$$\left|\left(V_n^{H,p}(X^f,t,\delta t)\right)^{1/pH}-\left(V_n^{H,p}(X,t,\delta t)\right)^{1/pH}\right| \overset{(a.s.)}{\to} 0 \text{ as } n\to+\infty \tag{4.19}$$

Proof. One needs to show that

$$\left|V_n^{H,p}(X^f,\lfloor t/\delta t\rfloor\delta t,\delta t)-V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right| \overset{(a.s.)}{\to} 0 \tag{4.20}$$

and then by the Theorem of continuous mapping see for example [68], we complete the

proof. Using the triangular inequality for $p \leq 1$, one gets for n large enough

$$\begin{aligned}
&\left|V_n^{H,p}(X^f, \lfloor t/\delta t \rfloor \delta t, \delta t) - V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right| \\
&= \left| \frac{1}{2^n} \sum_{k=1}^{2^n-1} \left|\Delta_{n,k}^H X(\lfloor t/\delta t \rfloor \delta t, \delta t) + \Delta_{n,k}^H f(\lfloor t/\delta t \rfloor \delta t, \delta t)\right|^p - \frac{1}{2^n} \sum_{k=1}^{2^n-1} \left|\Delta_{n,k}^H X(\lfloor t/\delta t \rfloor \delta t, \delta t)\right|^p \right| \\
&\leq \frac{1}{2^n} \sum_{k=1}^{2^n-1} \left| \frac{f\left(\lfloor t/\delta t \rfloor \delta t + (k+1)\delta_n\right) - f\left(\lfloor t/\delta t \rfloor \delta t + k\delta_n\right)}{\delta_n^H} \right|^p \\
&\leq \frac{C}{2^n} \sum_{k=1}^{2^n-1} \left| \frac{\delta_n^{\alpha_f}}{\delta_n^H} \right|^p \\
&= C \left|\delta_n\right|^{p(\alpha_f - H)} \to 0 \text{ as } n \to +\infty \quad (\alpha_f > H)
\end{aligned}$$

This implies (4.20). For $p > 1$, we use the Minkowski inequality, one gets

$$\begin{aligned}
&\left| \left(V_n^{H,p}(X^f, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)^{\frac{1}{p}} - \left(V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)^{\frac{1}{p}} \right|^p \\
&\qquad = \frac{1}{2^n} \left| \left(\sum_{k=1}^{2^n-1} \left|\Delta_{n,k}^H X(\lfloor t/\delta t \rfloor \delta t, \delta t) + \Delta_{n,k}^H f(\lfloor t/\delta t \rfloor \delta t, \delta t)\right|^p \right)^{\frac{1}{p}} \right. \\
&\qquad \left. - \left(\sum_{k=1}^{2^n-1} \left|\Delta_{n,k}^H X(\lfloor t/\delta t \rfloor \delta t, \delta t)\right|^p \right)^{\frac{1}{p}} \right|^p \\
&\qquad \leq \frac{1}{2^n} \sum_{k=1}^{2^n-1} \left| \frac{f\left(\lfloor t/\delta t \rfloor \delta t + (k+1)\delta_n\right) - f\left(\lfloor t/\delta t \rfloor \delta t + k\delta_n\right)}{\delta_n^H} \right|^p \\
&\qquad \leq C \left|\delta_n\right|^{p(\alpha_f - H)} \to 0 \text{ as } n \to +\infty \quad (\alpha_f > H)
\end{aligned}$$

This implies $\left| \left(V_n^{H,p}(X^f, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)^{\frac{1}{p}} - \left(V_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)^{\frac{1}{p}} \right| \overset{(a.s.)}{\to} 0$, whence (4.19). □

Note that in practice, Proposition 4.2.6 shows that the p-variation estimator is not affected by a trend or seasonality effect added to the process X provided that it is smoother than X. We may also consider the function f as a stochastic process. Choosing for example f as locally self-similar process, independent of the process X, with local index of self-similarity $h_f(t)$ is such that $\inf_{t\in\mathbb{R}}\{h_f(t)\} > H$, then Proposition 4.2.6 still holds.

4.3 Time changed estimation using the log-varition

We present a method to estimate $\lambda_{t,\delta t}$, from the process $X(t) = \tilde{X}(\theta(t))$. As in Section 4.2, we separate two cases. We first consider the process $\tilde{X}$ having stationary increments and such that the covariance function of the logarithm of the absolute value of

its increments satisfies

$$\sup_{\gamma\in(0,1)} c_{\tilde{X}}^{U}(k,\gamma) \leq O\left(\frac{1}{k^{4(1-H)}}\right) \text{ as } k \to +\infty$$

For the second case, we consider the process $\tilde{X}$ having stationary and ergodic increments. We then study the robustness of the estimator by adding to the process X, a function f which is Hölder continuous with Hölder exponent bigger than the local index of self-similarity of the process $\tilde{X}$.

Lemma 4.3.1. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi and θ defined by (4.10). Then*

$$U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t) \overset{(d)}{=} H\log(\lambda_{t,\delta t}) + U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t) \tag{4.21}$$

Proof.

$$\begin{aligned}
&U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t)\\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1} \log\left|\frac{X(\lfloor t/\delta t\rfloor \delta t + (k+1)\delta_n) - X(\lfloor t/\delta t\rfloor \delta t + k\delta_n)}{{\delta_n}^H}\right|\\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1} \log\left|\frac{\tilde{X}(\theta(\lfloor t/\delta t\rfloor \delta t) + \lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\theta(\lfloor t/\delta t\rfloor \delta t) + \lambda_{t,\delta t}k\delta_n)}{{\delta_n}^H}\right|\\
&\overset{(d)}{=} \frac{1}{2^n}\sum_{k=0}^{2^n-1} \log\left({\lambda_{t,\delta t}}^H\left|\frac{\tilde{X}(\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\lambda_{t,\delta t}k\delta_n)}{\lambda_{t,\delta t}^H{\delta_n}^H}\right|\right)\\
&= H\log(\lambda_{t,\delta t}) + U_n^{H,p}(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)
\end{aligned}$$

□

4.3.1 Time-changed H-lsssi process with condition on covariance

Theorem 4.3.2. *Let $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi and θ defined by (4.10). We assume that*

$$\left\{\log^2\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t-t'|^H}\right) : (t,t') \in \mathrm{T}\right\}, \text{ is uniformly integrable} \tag{4.22}$$

and

$$\sup_{\gamma\in(0,1)} c_{\tilde{X}}^{U}(k,\gamma) \leq O\left(\frac{1}{k^{4(1-H)}}\right) \text{ as } k \to +\infty \tag{4.23}$$

then

$$\exp\left(\frac{U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t) - \mu^U_{T_{\tilde{X},0}}}{H}\right) \overset{(a.s.)}{\to} \lambda_{t,\delta t}, \text{ as } n \to +\infty \tag{4.24}$$

Proof. By Lemma 4.3.1, the expected value of $U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t)$ is given by

$$\mathbb{E}\left[U_n^{H,p}(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right] = H \log(\lambda_{t,\delta t}) + \mathbb{E}\left[U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right]$$

By assumption (4.22), $\log\left|\frac{\tilde{X}(t) - \tilde{X}(t')}{(t-t')^H}\right|$ is uniformly integrable, one has by Lemma 4.1.1

$$\lim_{n\to+\infty} \mathbb{E}\left[\log\left|\frac{\tilde{X}(\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(\lambda_{t,\delta t}k\delta_n)}{\lambda_{t,\delta t}^H {\delta_n}^H}\right|\right] = \mathbb{E}\left[\log\left|T_{\tilde{X},0}(1)\right|\right]$$

the limit is uniform in k, so

$$\begin{aligned}\lim_{n\to+\infty} \mathbb{E}\left[U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right] &= \lim_{n\to+\infty} H \log(\lambda_{t,\delta t}) + \mathbb{E}\left[U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right] \\ &= H \log(\lambda_{t,\delta t}) + \mu^U_{T_{\tilde{X},0}} \end{aligned} \tag{4.25}$$

We now show the following limit to complete the proof

$$\lim_{n\to+\infty} \left|U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t) - \mathbb{E}\left[U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right]\right| \overset{(a.s.)}{=} 0 \tag{4.26}$$

Which is equivalent to show $\lim_{n\to+\infty} \left|U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t) - \mathbb{E}\left[U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right]\right| \overset{(a.s.)}{=} 0$. The variance of $U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)$ is given by

$$\mathbb{V}ar\left(U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right) = \frac{1}{2^{2n}} \sum_{k=0}^{2^n-1} v^U_{\tilde{X}}(k, \lambda_{t,\delta t}\delta_n) + \frac{2}{2^{2n}} \sum_{k=1}^{2^n-1} (2^n - k) c^U_{\tilde{X}}(k, \lambda_{t,\delta t}\delta_n)$$

From assumption (4.22), one has $\sup_{t,t'\in \mathrm{T}} \mathbb{E}\left[\log^2\left|\frac{\tilde{X}(t) - \tilde{X}(t')}{(t-t')^H}\right|\right] < +\infty$, this yields,

$$\sup_{n,k} v^U_X(k, \lambda_{t,\delta t}\delta_n) < +\infty$$

Setting $p_n = \mathbb{P}\left(|U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t) - \mathbb{E}[U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)]| \geq \epsilon\right)$, for $\epsilon > 0$. By the Tchebychev inequality, one gets

$$\begin{aligned} p_n &\leq \frac{\mathbb{V}ar\left(U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)\right)}{\epsilon^2} \\ &\leq \frac{1}{2^{2n}\epsilon^2} \sum_{k=0}^{2^n-1} v^U_X(k, \lambda_{t,\delta t}\delta_n) + \frac{2}{2^{2n}\epsilon^2} \sum_{k=1}^{2^n-1} (2^n - k) c^U_X(k, \lambda_{t,\delta t}\delta_n) \\ &\leq \frac{1}{2^n\epsilon^2} \sup_{n,k} v^U_X(k, \lambda_{t,\delta t}\delta_n) + \frac{2}{2^{2n}\epsilon^2} \sum_{k=1}^{2^n-1} (2^n - k) O\left(\frac{1}{k^{4(1-H)}}\right) \\ &\leq \frac{1}{2^n\epsilon^2} \sup_{n,k} v^U_X(k, \lambda_{t,\delta t}\delta_n) + \frac{2}{2^n\epsilon^2} \sum_{k=1}^{2^n-1} O\left(\frac{1}{k^{4(1-H)}}\right) - \frac{2}{2^{2n}\epsilon^2} \sum_{k=1}^{2^n-1} O\left(\frac{1}{k^{3-4H}}\right) \end{aligned}$$

Whence as in proof of Theorem 4.2.2

- For $H \in (0, 1/2)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{1}{2^{2n}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H = 1/2$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{n}{2^{2n}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H \in (1/2, 3/4)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^n}) + O(\frac{1}{2^{4n(1-H)}})$. This implies $p_n = O(\frac{1}{2^n})$
- For $H = 3/4$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{n}{2^n}) + O(\frac{1}{2^n})$. This implies $p_n = O(\frac{n}{2^n})$
- For $H \in (3/4, 1)$, $p_n \leq O(\frac{1}{2^n}) + O(\frac{1}{2^{4n(1-H)}}) + O(\frac{1}{2^{4n(1-H)}})$. This implies $p_n = O(\frac{1}{2^{4n(1-H)}})$

whence for each case $\sum_{n=1}^{\infty} p_n < +\infty$, which implies (4.26) by the Borel-Cantelli Lemma. Hence using (4.25) and by continuity of the function

$$x \to \exp\left(\frac{x - \mathbb{E}[\log|T_{\tilde{X},0}(1)|]}{H}\right)$$

the continuous mapping Theorem yields

$$\exp\left(\frac{U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t) - \mu^U_{T_{\tilde{X},0}}}{H}\right) \overset{(a.s.)}{\to} \lambda_{t,\delta t}, \text{ as } n \to +\infty$$

□

We now present a way to check assumption (4.22) of Theorem 4.3.2. We start by the two following Lemmas.

Lemma 4.3.3. *Let Y_n be a sequence of random variables. If for some $\alpha \in (0, 1)$*

$$\sup_{n\in\mathbb{N}} \mathbb{E}|Y_n|^{\alpha} < \infty \text{ and } \sup_{n\in\mathbb{N}} \mathbb{E}|Y_n|^{-\alpha} < \infty$$

then $\log^2|Y_n|$ is uniformly integrable

Proof. To show that $\log^2|Y_n|$ is uniformly integrable, we show that there exists a convex function ϕ_α (see for example [68]), such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\Big[\phi_\alpha(\log^2|Y_n|)\Big] < +\infty$$

We take the following convex function (as in [65])

$$\phi_\alpha(x) = e^{\alpha} x \mathbb{I}_{x\leq 1} + e^{\alpha\sqrt{x}} \mathbb{I}_{x>1}$$

One can check easily that the function ϕ_α is convex. One has

$$\mathbb{E}\Big[\phi_\alpha(\log^2|Y_n|)\Big] \leq e^{\alpha} + \mathbb{E}|Y_n|^{\alpha} + \mathbb{E}|Y_n|^{-\alpha}$$

hence $\sup_{n\in\mathbb{N}} \mathbb{E}\left[\phi_\alpha(\log^2|Y_n|)\right]$ is bounded, if for some $\alpha \in (0,1)$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|Y_n|^\alpha < \infty \text{ and } \sup_{n\in\mathbb{N}} \mathbb{E}|Y_n|^{-\alpha} < \infty$$

□

The following Lemma gives us an easy way to check when the $\mathbb{E}|Y|^{-\alpha}$ is finite for $\alpha \in (0,1)$for a given random variable Y.

Lemma 4.3.4. *Let Y be a random variable , let f be the density function of Y. If for some $\epsilon > 0$ the function f is bounded on $(-\epsilon,\epsilon)$. Then for $\alpha \in (0,1)$, $\mathbb{E}|Y|^{-\alpha} < +\infty$.*

Proof. There exists a positive constant M which bound the function f on $(-\epsilon,\epsilon)$, so that

$$\begin{aligned}
\mathbb{E}|Y|^{-\alpha} &= \int_{-\infty}^{+\infty} \frac{1}{|y|^\alpha} f(y)dy \\
&\leq 2M \int_0^\epsilon \frac{1}{y^\alpha} dy + \int_\epsilon^{+\infty} \frac{1}{y^\alpha}(f(y)+f(-y))dy \\
&\leq 2M\epsilon^{1-\alpha} + \frac{1}{\epsilon^\alpha}\int_\epsilon^{+\infty}(f(y)+f(-y))dy \\
&\leq 2M\epsilon^{1-\alpha} + \frac{1}{\epsilon^\alpha} \\
&< +\infty
\end{aligned}$$

□

Note using Lemma 4.3.3, the uniformly integrability of (4.22) is satisfied if for $0 < \alpha < 1$

$$\sup_{t,t'\in \mathrm{T}} \mathbb{E}\left(\frac{|\tilde{X}(t)-\tilde{X}(t')|}{|t-t'|^H}\right)^\alpha < \infty \text{ and } \sup_{t,t'\in \mathrm{T}} \mathbb{E}\left(\frac{|\tilde{X}(t)-\tilde{X}(t')|}{|t-t'|^H}\right)^{-\alpha} < \infty \qquad (4.27)$$

For self-similar processes, condition (4.27) is reduced to

$$\mathbb{E}|\tilde{X}(1)|^\alpha < \infty \text{ and } \mathbb{E}|\tilde{X}(1)|^{-\alpha} < \infty$$

In the following section we present a similar result to Theorem 4.3.2 for processes that do not satisfy condition (4.23). We replace this condition by ergodicity of the increments of the process $\tilde{X}$.

4.3.2 Time-changed H-lsssi process with ergodic increments

Proposition 4.3.5. *Let $\tilde{X}$ be a H-sssi with ergodic increments such that*

$$\left|\mathbb{E}\left[\log|\tilde{X}(1)|\right]\right| < +\infty$$

Then

$$\exp\left(\frac{U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)-\mu_{\tilde{X}}^U}{H}\right)\overset{(\mathcal{P})}{\to}\lambda_{t,\delta t},\ \ as\ n\to+\infty$$

Proof. By H-sssi property of the process $\tilde{X}$, one has

$$\begin{aligned}U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)&\overset{(d)}{=}H\log(\lambda_{t,\delta t})+U_n^H(\tilde{X},0,\lambda_{t,\delta t}\delta t)\\&\overset{(d)}{=}H\log(\lambda_{t,\delta t})+\frac{1}{2^n}\sum_{k=0}^{2^n-1}\log|\tilde{X}(k+1)-\tilde{X}(k)|\end{aligned}$$

By the stationarity and the ergodicity of the increments of the process $\tilde{X}$, one gets by the ergodic Theorem [67]

$$\frac{1}{2^n}\sum_{k=0}^{2^n-1}\log|\tilde{X}(k+1)-\tilde{X}(k)|\overset{(a.s.)}{\to}\mathbb{E}[\log|\tilde{X}(1)|]\text{ as }n\to+\infty$$

Whence

$$\lim_{n\to+\infty}U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\overset{(\mathcal{P})}{=}H\log(\lambda_{t,\delta t})+\mu_{\tilde{X}}^U$$

Hence by continuity of the function $x\to\exp\left(\frac{x-\mu_{\tilde{X}}^U}{H}\right)$, Corollary 2 page 31 in [68] yields

$$\exp\left(\frac{U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)-\mu_{\tilde{X}}^U}{H}\right)\overset{(\mathcal{P})}{\to}\lambda_{t,\delta t},\ \text{as}\ n\to+\infty$$

□

Note for Proposition 4.3.5, one can show that $\left|\mathbb{E}\log\left|\tilde{X}(1)\right|\right|<+\infty$, using the fact that

$$\left|\mathbb{E}\log\left|\tilde{X}(1)\right|\right|\le\mathbb{E}\left|\log\left|\tilde{X}(1)\right|\right|\le\sqrt{\mathbb{E}\log^2\left|\tilde{X}(1)\right|}$$

and then use Lemmas 4.3.3 and 4.3.4 to show $\mathbb{E}\log^2\left|\tilde{X}(1)\right|<+\infty$. For locally self-similar processes, we present a similar result in the following Propositions.

Proposition 4.3.6. *Let $\tilde{X}$ be a H-lass process with stationary increments so that its tangent process $T_{\tilde{X},0}$ has stationary and ergodic increments and $\left|\mathbb{E}\log\left|T_{\tilde{X},0}(1)\right|\right|<+\infty$. Then*

$$\lim_{n\to+\infty}\lim_{\delta t\to0}\exp\left(\frac{U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)-\mu_{T_{\tilde{X},0}}^U}{H}\right)\overset{(\mathcal{P})}{=}\theta'(t)$$

Proof. Using Lemma 4.2.5, one has

$$\lim_{\delta t\to 0} U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t) \overset{(d)}{=} \lim_{\delta t\to 0} H\log(\lambda_{t,\delta t}) + U_n^H(\tilde{X}, 0, \lambda_{t,\delta t}\delta t)$$
$$\overset{(d)}{=} H\log(\theta'(t)) + \frac{1}{2^n}\sum_{k=0}^{2^n-1}\log|T_{\tilde{X},0}(k+1) - T_{\tilde{X},0}(k)|$$

Hence, by continuity of the function $x \to \exp(x)$, Corollary 1 page 31 in [68] implies

$$\lim_{\delta t\to 0}\exp\left(U_n^H(X, \lfloor t/\delta t\rfloor\delta t, \delta t)/H\right) \overset{(d)}{=} \theta'(t)\exp\left(\frac{1}{2^n H}\sum_{k=0}^{2^n-1}\log|T_{\tilde{X},0}(k+1) - T_{\tilde{X},0}(k)|\right)$$

Moreover, by the stationarity and the ergodicity of the increments of the process $T_{\tilde{X},0}$, one gets by the ergodic Theorem [67]

$$\frac{1}{2^n}\sum_{k=0}^{2^n-1}\log|T_{\tilde{X},0}(k+1) - T_{\tilde{X},0}(k)| \overset{(a.s.)}{\to} \mathbb{E}[\log|T_{\tilde{X},0}(1)|] \text{ as } n \to +\infty$$

Hence by continuity of the function $x \to \exp(x)$

$$\lim_{n\to+\infty}\lim_{\delta t\to 0}\exp\left(U_n^H(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right) \overset{(d)}{=} \lambda_{t,\delta t}\exp\left(\mu^U_{T_{\tilde{X},0}}/H\right)$$

Whence

$$\lim_{n\to+\infty}\lim_{\delta t\to 0}\exp\left(\frac{U_n^H(X, \lfloor t/\delta t\rfloor\delta t, \delta t) - \mu^U_{T_{\tilde{X},0}}}{H}\right) \overset{(\mathbb{P})}{=} \lambda_{t,\delta t}$$

□

4.3.3 Trend effect on the estimator

We consider the following process $X^f(t) = X(t) + f(t)$, where the process X is of the form $\tilde{X}(\theta(t))$, where $\tilde{X}$ is H-lsssi , θ as defined in (4.10), and f is Hölder continuous function with a global index $\alpha_f \in (H, 1]$, so that for $(t, s) \in$ I, I a compact subset of $\mathbb{R}$, there exists a positive constant C such that $|f(t) - f(s)| \leq C|t - s|^{\alpha_f}$.

Proposition 4.3.7. *Assuming the process $X^f(t) = X(t) + f(t)$, where $X(t)$ satisfying conditions of Theorem 4.3.2 and f an α_f-Hölder continuous function on $\mathbb{R}$ such that $\alpha_f > H$. Then*

$$\left|\exp\left(U_n^H(X^f, \lfloor t/\delta t\rfloor\delta t, \delta t)\right) - \exp\left(U_n^H(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right)\right| \overset{(a.s.)}{\to} 0, \quad as\ n \to +\infty \quad (4.28)$$

Proof. The process X^f and X have the same tangent process $T_{\tilde{X},0}$, since

$$\lim_{t,t'\to t_0} \frac{X^f(t) - X^f(t')}{|t-t'|^H} = \lim_{t,t'\to t_0} \frac{X(t) - X(t')}{|t-t'|^H} + \frac{f(t) - f(t')}{|t-t'|^H}$$
$$= \lim_{t,t'\to t_0} \frac{X(t) - X(t')}{|t-t'|^H}$$

Hence $\exp\left(U_n^H(X^f, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)$ and $\exp\left(U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t)\right)$ converge to the same limit

$$\lambda_{t,\delta t}^H \exp\left(E[\log|T_{\tilde{X},0}|]\right)$$

almost surely by Theorem 4.3.2.

□

Note that in practice, Proposition 4.3.7 shows that the log-variation estimators is not affected by any relatively smooth trend or seasonality effect added to the process X. We may also consider the function f as a stochastic process. Choosing for example f locally self-similar, with local index of self-similarity $h_f(t)$ such that $\inf_{t\in\mathbb{R}}\{h_f(t)\} > H$, and f is independent of the process X, then Proposition 4.3.7 still holds.

4.4 Consistency of the p-variation estimator for known H

Let m be a strictly positive integer. We define $\Lambda \subset \mathbb{R}^m$ the set of λ's estimators, where $\lambda = \{\lambda_{j\delta t,\delta t}\}_{j\in\{0,\dots,m-1\}}$. We note $\hat{\lambda}^{V,n}(p) = \left\{\hat{\lambda}^{V,n}_{j\delta t,\delta t}(p)\right\}_{j\in\{0,\dots,m-1\}}$ and $\hat{\lambda}^{U,n} = \left\{\hat{\lambda}^{U,n}_{j\delta t,\delta t}\right\}_{j\in\{0,\dots,m-1\}}$ estimators of λ respectively using the p-variation and the log-variation, for $\delta t = T/m$. Both estimators are defined such that for every $j \in \{0,\dots,m-1\}$

$$\hat{\lambda}^{V,n}_{j\delta t,\delta t}(p) = \left(\frac{V_n^{H,p}(X, j\delta t, \delta t)}{\mu^V_{T_{\tilde{X},0}}}\right)^{\frac{1}{pH}} \quad \text{and} \quad \hat{\lambda}^{U,n}_{j\delta t,\delta t} = \exp\left(\frac{U_n^H(X, j\delta t, \delta t) - \mu^U_{T_{\tilde{X},0}}}{H}\right)$$

We analyse the bias and the mean square error (MSE) for $\hat{\lambda}^{V,n}(p)$ in this Section and $\hat{\lambda}^{U,n}$ in the next Section. We also check the asymptotic normality of the estimators in some cases. We define the normalised bias and the normalised MSE of an estimator $\hat{\lambda} \in \Lambda$ as

$$\text{bias}(\hat{\lambda}) = \mathbb{E}\left(\frac{\hat{\lambda} - \lambda}{\lambda}\right) \quad \text{and} \quad \text{MSE}(\hat{\lambda}) = \mathbb{E}\left|\frac{\hat{\lambda} - \lambda}{\lambda}\right|^2 = \mathbb{V}ar\left(\frac{\hat{\lambda}}{\lambda}\right) + (\text{bias}(\hat{\lambda}))^2$$

Note that as $\hat{\lambda}$ is a vector, which can be written as $\hat{\lambda} = \left\{\hat{\lambda}_j\right\}_{j\in\{1,\dots,m\}}$, the normilised

bias and the normalised MSE of $\hat{\lambda}$ can be written as follow

$$\text{bias}(\hat{\lambda}) = \left\{\text{bias}\left(\hat{\lambda}_j\right)\right\}_{j\in\{1,\dots,m\}} \quad \text{and} \quad \text{MSE}(\hat{\lambda}) = \left\{\text{MSE}\left(\hat{\lambda}_j\right)\right\}_{j\in\{1,\dots,m\}}$$

We define the integrated normalised MSE I_{MSE} by

$$I_{\text{MSE}}(\hat{\lambda}) = \frac{1}{m}\sum_{j=1}^{m}\text{MSE}\left(\hat{\lambda}_j\right)$$

4.4.1 Bias and MSE of the estimator on H-lsssi processes

We study the asymptotical behaviour of the normalised bias and the normilised MSE of $\hat{\lambda}^{V,n}(p) = \left\{\hat{\lambda}^{V,n}_{(j-1)\delta t,\delta t}(p)\right\}_{j\in\{1,\dots,m\}}$ as n goes to infinity. We start by the following Lemma

Lemma 4.4.1. *Let $X = \tilde{X}(\theta(t))$ where $\tilde{X}$ is locally self-similar and has stationary increments, and $\theta \in \Theta(\mathrm{T})$. We assume that there exists $\epsilon > 0$, so that*

$$\sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|\tilde{X}(t)-\tilde{X}(t')|}{|t-t'|^H}\right)^{2(1+\epsilon)/H} < \infty$$

Then for $p \leq 2/H$, the random variable $\left(V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right)^{2/pH}$ is uniformly integrable.

Proof. Note first since the function θ is bounded and differentiable on T and T is closed, there exist a constant $M > 0$ so that

$$\sup_{t,t'\in\mathrm{T}}\left|\frac{\theta(t)-\theta(t')}{t-t'}\right|^{2(1+\epsilon)} \leq M$$

One has

$$\sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|X(t)-X(t')|}{|t-t'|^H}\right)^{2(1+\epsilon)/H} \leq M\sup_{t,t'\in\mathrm{T}}\left(\frac{|\tilde{X}(t)-\tilde{X}(t')|}{|t-t'|^H}\right)^{2(1+\epsilon)/H} < \infty$$

To show that the random variable $\left(V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right)^{2/pH}$ is uniformly integrable, it suffices to show that for $\epsilon > 0$, $\sup_n \mathbb{E}\left[\left(V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right)^{2/pH+\epsilon}\right] < \infty$. Since $2/(pH)+\epsilon > 1$, whence by Jensen's inequality, one gets

$$\left(V_n^{H,p}(X, \lfloor t/\delta t\rfloor\delta t, \delta t)\right)^{2/pH+\epsilon} \leq \left(\frac{1}{2^n}\right)^{2/pH+\epsilon}\sum_{k=1}^{2^n-1}\left|\Delta^H_{n,k}X(t,\delta t)\right|^{2/H+p\epsilon}$$

Whence,

$$\mathbb{E}\left[\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{2/pH+\epsilon}\right] \leq \left(\frac{1}{2^n}\right)^{2/pH+\epsilon}\sum_{k=1}^{2^n-1}\mathbb{E}\left[\left|\Delta_{n,k}^H X(t,\delta t)\right|^{2/H+p\epsilon}\right]$$
$$\leq \left(\frac{1}{2^n}\right)^{2/pH+\epsilon-1}\sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|X(t)-X(t')|}{|t-t'|^H}\right)^{2/H+p\epsilon}$$

Using the fact that $p \leq 2/H$, one has

$$\sup_n \mathbb{E}\left[\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{2/pH+\epsilon}\right] \leq \sup_n\left(\frac{1}{2^n}\right)^{2/pH+\epsilon-1}\sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|X(t)-X(t')|}{|t-t'|^H}\right)^{2/H+p\epsilon}$$
$$\leq \sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|X(t)-X(t')|}{|t-t'|^H}\right)^{2(1+\epsilon)/H}$$
$$< \infty$$

□

Without lose of generality, we show in Proposition 4.4.2 that for $p \in (0, 2/H]$, the estimator $\hat{\lambda}_{t,\delta t}^{V,n}(p)$ is consistent as n goes to infinity .

Proposition 4.4.2. *We assume either Theorem 4.2.2 or Proposition 4.2.3 hold. Let* $X = \tilde{X}(\theta(t))$ *where* $\tilde{X}$ *is locally self-similar and has stationary increments, and* $\theta \in \Theta(\mathrm{T})$. *We assume that there exists* $\epsilon > 0$, *so that*

$$\sup_{t,t'\in\mathrm{T}}\mathbb{E}\left(\frac{|\tilde{X}(t)-\tilde{X}(t')|}{|t-t'|^H}\right)^{2(1+\epsilon)/H} < \infty$$

Then for $p \in (0, 2/H]$, *the estimator* $\hat{\lambda}_{t,\delta t}^{V,n}(p)$ *is consistent as* n *goes to infinity .*

Proof. One has

$$\mathbb{E}\left[\hat{\lambda}_{t,\delta t}^{V,n}(p)\right] = \mathbb{E}\left[\left(\frac{V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}}\right]$$
$$= \left(\frac{1}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}}\mathbb{E}\left[\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{1}{pH}}\right]$$

by Lemma 4.4.1, for $p \in (0, 2/H]$, $\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{1}{pH}}$ is uniformly integrable, thus one has

$$\lim_{n\to+\infty}\left(\frac{1}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}}\mathbb{E}\left[\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{1}{pH}}\right] = \lambda_{t,\delta t}$$

whence the estimator $\hat{\lambda}_{t,\delta t}^{V,n}(p)$ is asymptotically unbiased. For the MSE of the estimator,

it suffices to compute the variance of the estimator. One has

$$\mathbb{V}ar\left(\hat{\lambda}^{V,n}_{t,\delta t}(p)\right) = \mathbb{V}ar\left(\left(\frac{V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)}{\mu^V_{T_{\tilde{X},0}}(p)}(p)\right)^{\frac{1}{pH}}\right)$$
$$= \left(\frac{1}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{2}{pH}} \mathbb{V}ar\left(\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{1}{pH}}\right)$$

by Lemma 4.4.1, for $p \in (0, 2/H]$, $\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{2}{pH}}$ is uniformly integrable, thus one has

$$\lim_{n\to+\infty}\left(\frac{1}{\mu^V_{T_{\tilde{X},0}}(p)}\right)^{\frac{1}{pH}} \mathbb{V}ar\left(\left(V_n^{H,p}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right)^{\frac{1}{pH}}\right) = \mathbb{V}ar\left(\lambda_{t,\delta t}\right) = 0$$

whence $\lim_{n\to+\infty} \mathrm{MSE}(\hat{\lambda}^{V,n}_{t,\delta t}(p)) = 0$. The estimator $\hat{\lambda}^{V,n}_{t,\delta t}(p)$ is consistent. □

4.4.2 Bias of the estimator for H-sssi processes with $p = 1/H$

As our study is based in particular on self-similar processes, we show in Proposition 4.4.3 that for $\tilde{X}$ self-similar process with finite moment of order bigger or equal to $1/H$; the estimator $\lambda^{V,n}(1/H)$, that we note $\lambda^{V,n}$ for simplification, is unbiased.

Proposition 4.4.3. *Let $X = \tilde{X}(\theta(t))$ where $\tilde{X}$ is H-sssi and $\theta \in \Theta(\mathrm{T})$. We assume that $\mathbb{E}|\tilde{X}(1)|^{1/H} < \infty$, then the estimator $\hat{\lambda}^{V,n}_{t,\delta t}$ is unbiased.*

Proof. One has

$$\mathbb{E}[\hat{\lambda}^{V,n}_{t,\delta t}] = \mathbb{E}\left[\frac{V_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)}{\mu^V_{\tilde{X}}(1/H)}\right] = \frac{1}{\mu^V_{\tilde{X}}(1/H)}\mathbb{E}\left[V_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right]$$

Since

$$\mathbb{E}\left[V_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)\right] = \frac{1}{2^n}\sum_{k=1}^{2^n-1}\mathbb{E}\left|\frac{\tilde{X}(t+\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(t+\lambda_{t,\delta t}k\delta_n)}{\delta_n^H}\right|^{1/H}$$
$$= \lambda_{t,\delta t}\mathbb{E}\left|\tilde{X}(1)\right|^{1/H}$$

One has $\mathbb{E}[\hat{\lambda}^{V,n}_{t,\delta t}] = \lambda_{t,t+\delta t}$ and $\mathrm{bias}(\hat{\lambda}^{V,n}_{t,\delta t}) = 0$ □

4.4.3 MSE of the estimator for H-sssi processes with independent increments with $p = 1/H$

In Proposition 4.4.4 we compute the normalised MSE of the estimator $\lambda^{V,n}$, for self-similar processes with independent increments and finite moment bigger or equal to $2/H$.

Proposition 4.4.4. *Let $X = \tilde{X}(\theta(t))$ where $\tilde{X}$ is H-sssi with independent increments and $\theta \in \Theta(\mathrm{T})$. We assume that $\mathbb{E}|\tilde{X}(1)|^{2/H} < \infty$, the MSE is given by*

$$MSE(\hat{\lambda}^{V,n}_{t,\delta t}) = \frac{(\sigma^V_{\tilde{X}}(1/H))^2}{2^n(\mu^V_{\tilde{X}}(1/H))^2} \tag{4.29}$$

Proof. Since the estimator is unbiased, it suffices to compute the variance of $\hat{\lambda}^{V,n}_{t,\delta t}$

$$\begin{aligned}\mathbb{V}ar\left(\hat{\lambda}^{V,n}_{t,\delta t}\right) &= \mathbb{V}ar\left(\frac{V_n^{H,1/H}(X, \lfloor t/\delta t\rfloor \delta t, \delta t)}{\mu^V_{\tilde{X}}(1/H)}\right) \\ &= \left(\frac{1}{\mu^V_{\tilde{X}}(1/H)}\right)^2 \mathbb{V}ar\left(V_n^{H,1/H}(X, \lfloor t/\delta t\rfloor \delta t, \delta t)\right)\end{aligned}$$

Since the increments of the process $\tilde{X}$ are independent, one has

$$\begin{aligned}&\mathbb{V}ar\left(V_n^{H,1/H}(X, \lfloor t/\delta t\rfloor \delta t, \delta t)\right) \\ &\quad= \frac{1}{2^{2n}}\sum_{k=1}^{2^n-1}\mathbb{V}ar\left(\left|\frac{\tilde{X}(t+\lambda_{t,\delta t}(k+1)\delta_n) - \tilde{X}(t+\lambda_{t,\delta t}k\delta_n)}{\delta_n^H}\right|^{1/H}\right) \\ &\quad= \frac{(\lambda_{t,\delta t})^2}{2^n}\mathbb{V}ar\left(|\tilde{X}(1)|^{1/H}\right)\end{aligned}$$

Having $\sigma^V_{\tilde{X}}(1/H) = \sqrt{\mathbb{V}ar\left(|\tilde{X}(1)|^{1/H}\right)}$ and $\mathrm{MSE}(\hat{\lambda}^{V,n}_{t,\delta t}) = \mathbb{V}ar\left(\frac{\hat{\lambda}^{V,n}_{t,\delta t}}{\lambda_{t,\delta t}}\right) + \left(\mathrm{bias}(\hat{\lambda}^{V,n}_{t,\delta t})\right)^2$, one has

$$\mathrm{MSE}(\hat{\lambda}^{V,n}_{t,\delta t}) = \frac{(\sigma^V_{\tilde{X}}(1/H))^2}{2^n(\mu^V_{\tilde{X}}(1/H))^2}$$

□

4.4.4 MSE for H-sssi with independent and Gaussian increments ($p = 1/H$)

In the case where the increments of the H-sssi are Gaussian distributed, then one gets Corollary 4.4.5. Note that H-sssi process having stationary, independent and normal distributed increments, must have $H = 1/2$. This is due to the unicity of the Brownian motion. However, we keep $H \in (0,1)$ so one can have an idea of how the MSE reacts as a function of H.

Corollary 4.4.5. *We assume $\tilde{X}$ H-sssi with Gaussian increments with mean zero. Then*

$$MSE(\hat{\lambda}^{V,n}_{t,\delta t}) = \frac{e(H)}{2^n} \tag{4.30}$$

where $e(H) = \left(\frac{\sqrt{\pi}\Gamma\left(\frac{2+H}{2H}\right)}{\Gamma\left(\frac{H+1}{2H}\right)^2} - 1\right)$

Proof. We can limit our case to the standard Gaussian process (mean 0 and variance 1). Indeed assuming the process $X = \sigma Z$ where Z is a standard Gaussian process and σ a positive constant, one has for a real $\alpha > 0$

$$\frac{\mathbb{V}ar\left(|\tilde{X}(1)|^\alpha\right)}{\left(\mathbb{E}|\tilde{X}(1)|^\alpha\right)^2} = \frac{\mathbb{V}ar\left(|\tilde{Z}(1)|^\alpha\right)}{\left(\mathbb{E}|\tilde{Z}(1)|^\alpha\right)^2}$$

One has

$$\mathrm{MSE}(\hat{\lambda}^{V,n}_{t,\delta t}) = \frac{1}{2^n}\left(\frac{\sigma^V_{\tilde{X}}(1/H)}{\mu^V_{\tilde{X}}(1/H)}\right)^2 = \frac{1}{2^n}\left(\frac{\mathbb{E}|\tilde{X}(1)|^{2/H}}{(\mathbb{E}|\tilde{X}(1)|^{1/H})^2} - 1\right)$$

for a real $\alpha > 0$, one gets the following result

$$\mathbb{E}|\tilde{X}(1)|^\alpha = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x|^\alpha \exp\left(-x^2/2\right)dx = \frac{2(\sqrt{2})^{1+\alpha}}{\sqrt{2\pi}}\int_0^{+\infty} x^\alpha \exp\left(-x^2\right)dx = \frac{2^{\frac{\alpha}{2}}\Gamma\left(\frac{\alpha+1}{2}\right)}{\sqrt{\pi}} \tag{4.31}$$

Whence by substituting α by $2/H$ and $1/H$, one gets

$$\mathrm{MSE}(\hat{\lambda}^{V,n}_{t,\delta t}) = \frac{1}{2^n}\left(\frac{\sqrt{\pi}\Gamma\left(\frac{2+H}{2H}\right)}{\Gamma\left(\frac{H+1}{2H}\right)^2} - 1\right)$$

□

We draw the function e in Figure 4.1. From this we deduce that for a given n, the estimator perform better on smoother processes. In the case of a time-changed Brownian motion, one has $e(1/2) = 2$, and the normalised MSE of the $\hat{\lambda}^{V,n}_{t,\delta t}$ is $1/2^{n-1}$. So assuming that we want an error to be less or equal to $\alpha \in (0,1)$, then n has to be bigger or equal to $1 - \log_2(\alpha)$, (this means for example for $\alpha = 0.01$, $n \geq 8$). Figure 4.2 illustrates the MSE of the estimator when applied to time-changed Brownian motion. We study next the asymptotic normality of the estimator.

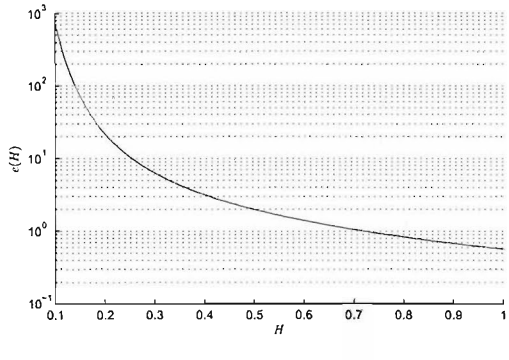

Figure 4.1: Function $\log_{10}(e(H))$.

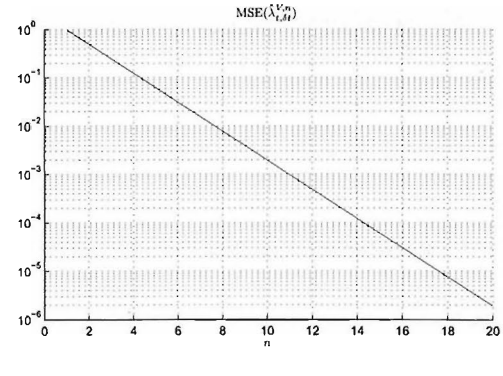

Figure 4.2: MSE of the estimator on time-changed Brownian motion.

4.4.5 Asymptotic normality for H-sssi processes with independent increments ($p = 1/H$)

We study the asymptotic normality of the estimator.

Proposition 4.4.6. *Under the conditions of Proposition 4.4.4, the asymptotic normality of the estimator $\hat{\lambda}^{V,n}_{t,\delta t}$ is given by*

$$\sqrt{2^n}\left(\frac{\hat{\lambda}^{V,n}_{t,\delta t}-\lambda_{t,\delta t}}{\lambda_{t,\delta t}}\right)\sim\mathcal{N}\left(0,\left(\frac{\sigma^V_{\tilde{X}}(1/H)}{\mu^V_{\tilde{X}}(1/H)}\right)^2\right)\quad as\ n\to+\infty \tag{4.32}$$

Proof.

$$\begin{aligned}
&\sqrt{2^n}\left(\hat{\lambda}^{V,n}_{t,\delta t}-\lambda_{t,\delta t}\right)\\
&\qquad=\sqrt{2^n}\left(\frac{V_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)}{\mu^V_{\tilde{X}}(1/H)}-\lambda_{t,\delta t}\right)\\
&\qquad\overset{(d)}{=}\lambda_{t,\delta t}\frac{\sigma^V_{\tilde{X}}(1/H)}{\mu^V_{\tilde{X}}(1/H)}\left(\frac{1}{\sigma^V_{\tilde{X}}(1/H)\sqrt{2^n}}\sum_{k=0}^{2^n-1}\left(|\tilde{X}(k+1)-\tilde{X}(k)|^{1/H}-\mu^V_{\tilde{X}}(1/H)\right)\right)
\end{aligned}$$

Hence having assumed $\tilde{X}$ with independent increments, the Central Limit Theorem yields

$$\sqrt{2^n}\left(\frac{\hat{\lambda}^{V,n}_{t,\delta t}-\lambda_{t,\delta t}}{\lambda_{t,\delta t}}\right)\sim\mathcal{N}\left(0,\left(\frac{\sigma^V_{\tilde{X}}(1/H)}{\mu^V_{\tilde{X}}(1/H)}\right)^2\right)\quad\text{as } n\to+\infty$$

□

Note that substituting H by its estimator $\hat{H}$ in the asymptotic normality expression 4.32, this last remains still valid.

4.5 Consistency of the log-variation estimators for known H

We study the asymptotical behavior of the normilised bias and the normilised MSE of $\hat{\lambda}^{U,n}=\left\{\hat{\lambda}^{U,n}_{[(j-1)\delta t,\delta t)}\right\}_{j\in\{1,\dots,m\}}$ as n goes to infinity. Here m is fixed and $\delta t = T/m$.

4.5.1 Bias and MSE of H-lsssi processes with independent increments

Lemma 4.5.1. *Let X_n be a sequence of strictly positive random variables, such that $X_n \overset{(d)}{\to} X$ as $n\to\infty$, and*

$$\mathbb{E}\left[X_n^{\alpha}\right]\le C \text{ and } \mathbb{E}\left[X_n^{-\alpha}\right]\le C \tag{4.33}$$

for some $\alpha > 0$ and C a finite real. Then

$$\lim_{n\to\infty} \left(\mathbb{E}\left[X_n^{1/n}\right]\right)^n = \exp\left(\mathbb{E}\log(X)\right) \tag{4.34}$$

Proof. From O.D. Jones
We set $Y_n = \log(X_n)$ and $Y = \log(X)$. Then (4.34) is equivalent to

$$\lim_{n\to\infty} \left(\mathbb{E}\left[\exp\left(Y_n/n\right)\right]\right)^n = \exp\left(\mathbb{E}Y\right)$$

We define

$$m_n(t) = \mathbb{E}\left[\exp\left(tY_n\right)\right] = \mathbb{E}\left[X_n^t\right]$$

The function m_n exists $\forall n$, and $t \in (-\alpha, \alpha)$. Thus $m_n'(0) = \mathbb{E}\left[Y_n\right]$ exists and we have

$$m_n(t) = 1 + m_n'(0) + o(t)$$

Condition implies (4.33) $\dfrac{o(t)}{t} \to 0$ as $t \to 0$ uniformly. This last combined with Lemma 4.3.3, implies that Y_n is uniformly integrable. So

$$\mathbb{E}\left[Y_n\right] = m_n'(0) \to m'(0) = \mathbb{E}\left[Y\right] \text{ as } n \to \infty$$

Thus we can write $m_n(t) = 1 + tm'(0) + o(t)$. And so

$$\log\left(\mathbb{E}\left[\exp\left(Y_n/n\right)\right]\right)^n = \log(1 + tm'(0) + o(t))^n \to \log(\exp(m'(0))) = \mathbb{E}\left[Y\right]$$

as $n \to \infty$. This end the proof. □

Proposition 4.5.2. *Assuming there exists $\alpha \in (0,1)$, so that*

$$\sup_{t,t'\in \mathrm{T}} \mathbb{E}\left[\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t-t'|^H}\right)^{\alpha}\right] < \infty \text{ and } \sup_{t,t'\in \mathrm{T}} \mathbb{E}\left[\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t-t'|^H}\right)^{-\alpha}\right] < \infty$$

Then the estimator $\hat{\lambda}_{t,\delta t}^{U,n}$ is asymptotically unbiased.

Proof. First, one can rewrite

$$\begin{aligned}\exp\left(U_n^H(X, \lfloor t/\delta t \rfloor \delta t, \delta t)/H\right) &= \exp\left(\frac{1}{2^n}\sum_{k=0}^{2^n-1} \log\left|\Delta_{n,k}^H X(t,\delta t)\right|/H\right) \\ &= \prod_{k=0}^{2^n-1} \left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{1}{2^n H}}\end{aligned}$$

Whence by the independency of the increments of X, one gets

$$\mathbb{E}\left[\exp\left(U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t)/H\right)\right] = \mathbb{E}\left[\prod_{k=0}^{2^n-1} \left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{1}{2^n H}}\right]$$
$$= \prod_{k=0}^{2^n-1} \mathbb{E}\left[\left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{1}{2^n H}}\right]$$

By stationarity of the increments of $\tilde{X}$, one has

$$\mathbb{E}\left[\exp\left(U_n^H(X, \lfloor t/\delta t\rfloor \delta t, \delta t)/H\right)\right] = \lambda_{t,\delta t}\left(\mathbb{E}\left[\left|\frac{\tilde{X}(\lambda_{t,\delta t}\delta_n) - \tilde{X}(0)}{(\lambda_{t,\delta t}\delta_n)^H}\right|^{\frac{1}{2^n H}}\right]\right)^{2^n}$$

The bias of the estimator being given by the expected value of $\hat{\lambda}_{t,\delta t}^{U,n}$, one has

$$\mathbb{E}[\hat{\lambda}_{t,\delta t}^{U,n}] = \mathbb{E}\left[\exp\left(\frac{U_n^{H,1/H}(X, \lfloor t/\delta t\rfloor \delta t, \delta t) - \mu_{\tilde{X}}^U}{H}\right)\right]$$
$$= \lambda_{t,\delta t}\exp\left(\frac{-\mu_{T_{\tilde{X},0}}^U}{H}\right)\left(\mathbb{E}\left[\left|\frac{\tilde{X}(\lambda_{t,\delta t}\delta_n) - \tilde{X}(0)}{(\lambda_{t,\delta t}\delta_n)^H}\right|^{\frac{1}{2^n H}}\right]\right)^{2^n} \tag{4.35}$$

Since $\left|\frac{\tilde{X}(\lambda_{t,\delta t}\delta_n) - \tilde{X}(0)}{(\lambda_{t,\delta t}\delta_n)^H}\right|^{\frac{1}{H}} \xrightarrow{(d)} |T_{\tilde{X},0}|^{\frac{1}{H}}$, one concludes by Lemma 4.5.1, $\mathbb{E}[\hat{\lambda}_{t,\delta t}^{U,n}] \to \lambda_{t,\delta t}$ as $n \to +\infty$ □

Proposition 4.5.3. *Assuming there exists $\alpha \in (0,1)$, so that*

$$\sup_{t,t'\in \mathrm{T}} \mathbb{E}\left[\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t-t'|^H}\right)^{\alpha}\right] < \infty \text{ and } \sup_{t,t'\in \mathrm{T}} \mathbb{E}\left[\left(\frac{|\tilde{X}(t) - \tilde{X}(t')|}{|t-t'|^H}\right)^{-\alpha}\right] < \infty$$

Then the estimator $\hat{\lambda}_{t,\delta t}^{U,n}$ is asymptotically consistent.

Proof. The MSE of the estimator is the sum of its variance and square bias. As the bias converge to 0 as n tends to infinity, it is sufficient to show that the variance of $\hat{\lambda}_{t,\delta t}^{U,n}$ tends to 0 as n tends to infinity. Using proof of Proposition 4.5.2 and setting $\Delta_{n,0}^H \tilde{X}(0, \lambda_{t,\delta t}\delta t) = \frac{\tilde{X}(\lambda_{t,\delta t}\delta_n) - \tilde{X}(0)}{(\lambda_{t,\delta t}\delta_n)^H}$, one has by the independency and the stationar-

ity of the increments of $\tilde{X}$

$$\begin{aligned}
&\mathbb{V}ar\left(\exp\left(U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)/H\right)\right)\\
&\quad=\mathbb{V}ar\left(\prod_{k=0}^{2^n-1}\left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{1}{2^nH}}\right)\\
&\quad=\mathbb{E}\left[\prod_{k=0}^{2^n-1}\left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{2}{2^nH}}\right]-\left(\mathbb{E}\left[\prod_{k=0}^{2^n-1}\left|\Delta_{n,k}^H X(t,\delta t)\right|^{\frac{1}{2^nH}}\right]\right)^2\\
&\quad=\lambda_{t,\delta t}^2\left(\left(\mathbb{E}\left[\left|\Delta_{n,0}^H\tilde{X}(0,\lambda_{t,\delta t}\delta t)\right|^{\frac{2}{2^nH}}\right]\right)^{2^n}-\left(\mathbb{E}\left[\left|\Delta_{n,0}^H\tilde{X}(0,\lambda_{t,\delta t}\delta t)\right|^{\frac{1}{2^nH}}\right]\right)^{2^{n+1}}\right)
\end{aligned}$$

The variance of the estimator is given by

$$\begin{aligned}
\mathbb{V}ar\left(\hat{\lambda}_{t,\delta t}^{U,n}\right)&=\mathbb{V}ar\left(\exp\left(\frac{U_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)-\mu_{\tilde{X}}^U}{H}\right)\right)\\
&=\exp\left(\frac{-2\mu_{T_{\tilde{X},0}}^U}{H}\right)\mathbb{V}ar\left(\exp\left(U_n^H(X,\lfloor t/\delta t\rfloor\delta t,\delta t)/H\right)\right)\\
&=\lambda_{t,\delta t}^2\exp\left(\frac{-2\mu_{T_{\tilde{X},0}}^U}{H}\right)\left(\left(\mathbb{E}\left[\left|\Delta_{n,0}^H\tilde{X}(0,\lambda_{t,\delta t}\delta t)\right|^{\frac{2}{2^nH}}\right]\right)^{2^n}\right.\\
&\quad\left.-\left(\mathbb{E}\left[\left|\Delta_{n,0}^H\tilde{X}(0,\lambda_{t,\delta t}\delta t)\right|^{\frac{1}{2^nH}}\right]\right)^{2^{n+1}}\right)
\end{aligned}\tag{4.36}$$

Since $\left|\Delta_{n,0}^H\tilde{X}(0,\lambda_{t,\delta t}\delta t)\right|^{\frac{1}{H}}\xrightarrow{(d)}|T_{\tilde{X},0}|^{\frac{1}{H}}$, one concludes by Lemma 4.5.1, $\mathbb{V}ar\left(\hat{\lambda}_{t,\delta t}^{U,n}\right)\to 0$ as $n\to+\infty$ □

4.5.2 Case of H-sssi processes with independent increments

For H-sssi processes with independent increments, the normalised bias and MSE are deduced respectively from (4.35) and (4.36), and are defined by

$$\text{bias}(\hat{\lambda}_{t,\delta t}^{U,n})=\exp\left(\frac{-\mu_{\tilde{X}}^U}{H}\right)\left(\mathbb{E}\left[\left|\tilde{X}(1)\right|^{\frac{1}{2^nH}}\right]\right)^{2^n}-1\tag{4.37}$$

$$\text{MSE}(\hat{\lambda}_{t,\delta t}^{U,n})=\mathbb{V}ar\left(\frac{\hat{\lambda}_{t,\delta t}^{U,n}}{\lambda_{t,\delta t}}\right)+\left(\text{bias}(\hat{\lambda}_{t,\delta t}^{U,n})\right)^2\tag{4.38}$$

where $\mathbb{V}ar\left(\dfrac{\hat{\lambda}_{t,\delta t}^{U,n}}{\lambda_{t,\delta t}}\right)=\exp\left(\dfrac{-2\mu_{\tilde{X}}^U}{H}\right)\left(\left(\mathbb{E}\left[\left|\tilde{X}(1)\right|^{\frac{2}{2^nH}}\right]\right)^{2^n}-\left(\mathbb{E}\left[\left|\tilde{X}(1)\right|^{\frac{1}{2^nH}}\right]\right)^{2^{n+1}}\right)$.

4.5.2.1 Case of H-sssi Gaussian processes

For Gaussian processes, the bias and the MSE can be given explicitly. We first present the following Lemma and then we compute the bias and the MSE of an H-sssi process with Gaussian and independent increments. Note that Gaussian H-sssi processes with $H \neq 1/2$ are correlated, and so there is no point to study the case where $H \neq 1/2$. However, we decide to study for all $H \in (0,1)$.

Lemma 4.5.4.

$$\int_0^{+\infty} \frac{\log(x)\exp(-x^2/2)}{\sqrt{2\pi}}dx = -\frac{1}{4}(\gamma + \log(2))$$

$$\int_0^{+\infty} \frac{(\log(x))^2\exp(-x^2/2)}{\sqrt{2\pi}}dx = \frac{1}{16}\left(\pi^2 + 2\log^2(2) + \gamma(2\gamma + \log(16))\right)$$

where $\gamma = -\int_0^{+\infty}\exp(-x)\log(x)dx \approx 0.57721$ *is the Euler-Mascheroni constant.*

Proposition 4.5.5. *Let* $\tilde{X}$ *be an* H*-sssi process with independent increments of mean zero. The bias and the MSE of the estimator* $\hat{\lambda}_{t,\delta t}^{U,n}$ *are given by*

$$bias(\hat{\lambda}_{t,\delta t}^{U,n}) = 2^{1/H}\exp\left(\frac{\gamma}{2H}\right)\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^{n+1}H}\right)}{\sqrt{\pi}}\right)^{2^n} - 1 \tag{4.39}$$

and

$$MSE\left(\hat{\lambda}_{t,\delta t}^{U,n}\right) = 1+2^{2/H}\exp\left(\frac{\gamma}{H}\right)\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^{n}H}\right)}{\sqrt{\pi}}\right)^{2^n} - 2^{1+1/H}\exp\left(\frac{\gamma}{2H}\right)\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^{n+1}H}\right)}{\sqrt{\pi}}\right)^{2^n} \tag{4.40}$$

Proof. Assuming $\tilde{X}(t) = \sigma_t Z(t)$, where Z is H-sssi process and $\sigma_t = t^H\sigma$, one has

$$\begin{aligned}\exp\left(-\mathbb{E}\left[\log\left|\tilde{X}(1)\right|/H\right]\right)\left(\mathbb{E}|\tilde{X}(1)|^{\frac{1}{2^nH}}\right)^{2^n} &= \exp\left(-\mathbb{E}\left[\log|\sigma Z(1)|/H\right]\right)\left(\mathbb{E}|\sigma Z(1)|^{\frac{1}{2^nH}}\right)^{2^n}\\ &= \exp\left(-\mathbb{E}\left[\log|Z(1)|/H\right]\right)\left(\mathbb{E}|Z(1)|^{\frac{1}{2^nH}}\right)^{2^n}\end{aligned}$$

Note that the bias of the estimator $\hat{\lambda}_{t,\delta t}^{U,n}$ is not affected by the variance σ^2 of the increments of $\tilde{X}$. The computation are then done for $\sigma = 1$. One has

$$\mathbb{E}\left[\log\left|\tilde{X}(1)\right|/H\right] = \frac{2}{H}\int_0^{+\infty}\frac{\log(x)\exp(-x^2/2)dx}{\sqrt{2\pi}} = -\frac{1}{2H}\left(\log(2)+\gamma\right) \tag{4.41}$$

and by (4.31)

$$\left(\mathbb{E}|\tilde{X}(1)|^{\frac{1}{2^nH}}\right)^{2^n} = (\sqrt{2})^{\frac{1}{H}}\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^{n+1}H}\right)}{\sqrt{\pi}}\right)^{2^n}$$

Whence using (4.37), one gets (4.39). The MSE of $\hat{\lambda}_{t,\delta t}^{U,n}$ is defined by (4.38). Simple

computation yields to

$$\mathrm{MSE}\left(\hat{\lambda}^{U,n}_{t,\delta t}\right) = 1 + \exp\left(\frac{-2\mu^U_{\tilde{X}}}{H}\right)\left(\mathbb{E}\left[\left|\tilde{X}(1)\right|^{\frac{2}{2^n H}}\right]\right)^{2^n} - 2\exp\left(\frac{-\mu^U_{\tilde{X}}}{H}\right)\left(\mathbb{E}\left[\left|\tilde{X}(1)\right|^{\frac{1}{2^n H}}\right]\right)^{2}$$

Using (4.41) and (4.31), one has

$$\mathrm{MSE}\left(\hat{\lambda}^{U,n}_{t,\delta t}\right) = 1+2^{2/H}\exp\left(\frac{\gamma}{H}\right)\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^n H}\right)}{\sqrt{\pi}}\right)^{2^n} - 2^{1+1/H}\exp\left(\frac{\gamma}{2H}\right)\left(\frac{\Gamma\left(\frac{1}{2}+\frac{1}{2^{n+1}H}\right)}{\sqrt{\pi}}\right)^{2^n}$$

□

We illustrate in Figure 4.3 the bias and the MSE of the estimator in the case $\tilde{X}$ is a Brownian motion.

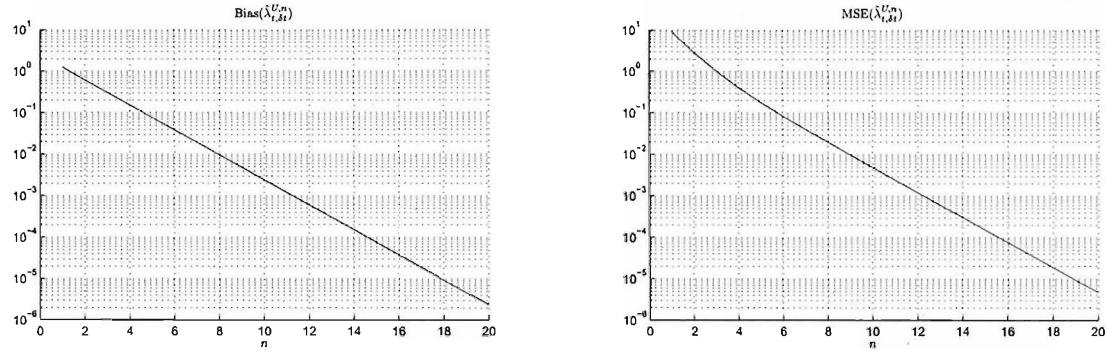

Figure 4.3: Bias (left panel) and the MSE (right panel) of $\hat{\lambda}^{U,n}_{t,\delta t}$ of a time-changed Brownian motion

4.5.2.2 Asymptotic normality of the estimator

We study the asymptotic normality of the estimator. Let us start with the following Lemma.

Lemma 4.5.6. *Let $\hat{\lambda}^n$ be an estimator of λ, such that for $v > 0$*

$$\sqrt{2^n}(\hat{\lambda}^n - \lambda) \sim \mathcal{N}(0, v) \text{ as } n \to +\infty$$

Let f a $\mathcal{C}^1$ function such that f is differentiable in λ. Then

$$\sqrt{2^n}(f(\hat{\lambda}^n) - f(\lambda)) \sim \mathcal{N}(0, f'(\lambda)^2 v) \text{ as } n \to +\infty$$

Proof. First, we note that $\hat{\lambda}^n \xrightarrow{(\mathcal{P})} \lambda$. Moreover, there exists a real $\hat{\lambda}^n_*$, in the open ball of radius $|\hat{\lambda}^n - \lambda|$ centered in λ, so that $f(\hat{\lambda}^n) = f(\lambda) + f'(\hat{\lambda}^n_*)(\hat{\lambda}^n - \lambda)$; by the Continuous Mapping Theorem one has $f(\hat{\lambda}^n) \xrightarrow{(\mathcal{P})} f(\lambda)$ and $f'(\hat{\lambda}^n_*) \xrightarrow{(\mathcal{P})} f'(\lambda)$. Hence

$$\sqrt{2^n}(f(\hat{\lambda}^n) - f(\lambda)) = \sqrt{2^n}(\hat{\lambda}^n - \lambda) f'(\hat{\lambda}^n_*) \xrightarrow{(d)} \mathcal{N}(0, f'(\lambda)^2 v)$$

□

Proposition 4.5.7. *The asymptotic normality of the estimator* $\hat{\lambda}^{U,n}_{t,\delta t}$ *is given by*

$$\sqrt{2^n}\left(\frac{\hat{\lambda}^{U,n}_{t,\delta t}-\lambda_{t,\delta t}}{\lambda_{t,\delta t}}\right)\sim\mathcal{N}\left(0,\left(\frac{\sigma^U_{\tilde{X}}}{H}\right)^2\right)\quad as\ n\to+\infty \tag{4.42}$$

Proof. One has

$$\begin{aligned}\sqrt{2^n}\left(\log(\hat{\lambda}^{U,n}_{t,\delta t})-\log(\lambda_{t,\delta t})\right)&=\sqrt{2^n}\left(\frac{U_n^{H,1/H}(X,\lfloor t/\delta t\rfloor\delta t,\delta t)-\mu^U_{\tilde{X}}}{H}-\log(\lambda_{t,\delta t})\right)\\&\overset{(d)}{=}\frac{\sigma^U_{\tilde{X}}}{H}\left(\frac{1}{\sigma^U_{\tilde{X}}\sqrt{2^n}}\sum_{k=0}^{2^n-1}\left(\log|\tilde{X}(k+1)-\tilde{X}(k)|-\mu^U_{\tilde{X}}\right)\right)\end{aligned}$$

Having assumed $\tilde{X}$ with independent increments, the Central Limit Theorem yields to

$$\sqrt{2^n}\left(\log(\hat{\lambda}^{U,n}_{t,\delta t})-\log(\lambda_{t,\delta t})\right)\sim\mathcal{N}\left(0,\left(\frac{\sigma^U_{\tilde{X}}}{H}\right)^2\right)\ \text{as}\ n\to+\infty$$

By Lemma 4.5.6 using function $g:x\mapsto\exp(x)$ for $x\in\mathbb{R}$, one has $g':x\mapsto\exp(x)$,

$$\sqrt{2^n}\left(\hat{\lambda}^{U,n}_{t,\delta t}-\lambda_{t,\delta t}\right)\sim\mathcal{N}\left(0,\lambda^2_{t,\delta t}\left(\frac{\sigma^U_{\tilde{X}}}{H}\right)^2\right)\ \text{as}\ n\to+\infty$$

□

Note for Gaussian H-sssi processes $\sigma^U_{\tilde{X}}=\dfrac{\pi}{2\sqrt{2}}$. Indeed, by Lemma 4.5.4, one gets

$$\begin{aligned}\mathbb{V}ar\left(\log|\tilde{X}(1)|\right)&=\int_R\frac{(\log|x|)^2\exp(-x^2/2)}{\sqrt{2\pi}}dx-\left(\int_R\frac{\log|x|\exp(-x^2/2)}{\sqrt{2\pi}}dx\right)^2\\&=2\int_0^{+\infty}\frac{(\log(x))^2\exp(-x^2/2)}{\sqrt{2\pi}}dx-\left(2\int_0^{+\infty}\frac{\log(x)\exp(-x^2/2)}{\sqrt{2\pi}}dx\right)^2\\&=\frac{1}{8}\left(\pi^2+2\log^2(2)+\gamma(2\gamma+\log(16))\right)-\frac{1}{4}(\gamma+\log(2))^2\\&=\pi^2/8\end{aligned}$$

4.6 Experimental application to some H-sssi processes

In this section, we test the performances of our estimators on time-changed self-similar processes; In particular we use the following H-sssi processes: the Brownian motion, the Lévy α-stable motion and the fractional Brownian motion. We set the following time-change function θ. For $t\in\mathrm{T}=[0,1]$ and $\epsilon=10^{-7}$, we define θ as follows (see

Figure 4.4)

$$\theta(t) = \frac{\cos(\pi/2 + 4\pi t) - 3/4\sin(\pi/2 + 8\pi t) + (10+\epsilon)\pi t + 3/4}{(10+\epsilon)\pi} \tag{4.43}$$

The quantity ϵ is to ensure that the derivative of the function θ is strictly positive for any $t \in [0,1]$. The function θ satisfies $\theta(0) = 0$, $\theta(1) = 1$ and for all $t \in \mathrm{T}$, $\theta'(t) > 0$. However, our function θ here is not piecewise continuous, but this is not so important since in practice θ is unknown. Figure 4.4 illustrates the curve of θ and its derivative θ'.

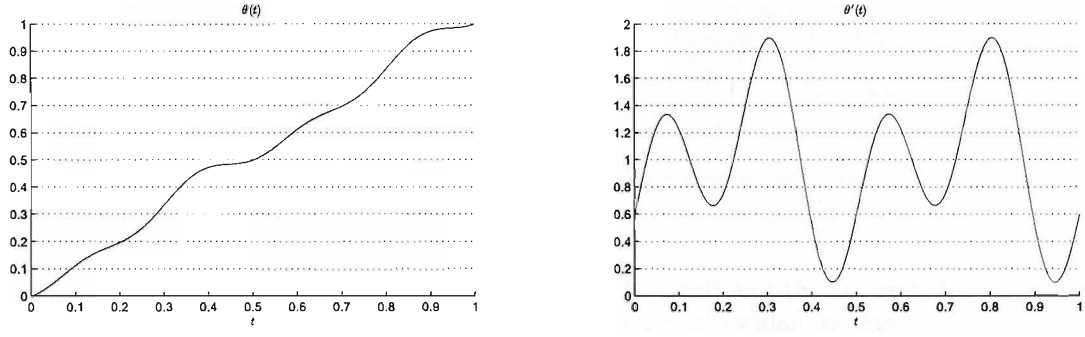

Figure 4.4: The function θ (left panel) and its derivative θ' (right panel)

4.6.1 Time-changed Brownian motion

The time changed Brownian motion is given by (4.44). The justification of this representation is explained in Chapter 3.

$$B(\theta(t)) \overset{(fdd)}{=} \int_0^t \sqrt{\theta'(s)} dB(s) \tag{4.44}$$

For these tests, the theoretical self-similarity index is used ($H = 1/2$). Indirectly in the previous subsection, we already checked that the Brownian motion does satisfy Theorems 4.2.2 and 4.3.2. However we remind that for $\epsilon > 0$, $\mathbb{E}|B(1)|^{2+\epsilon} < +\infty$ and for $\alpha \in (0,1)$, $\mathbb{E}|B(1)|^{-\alpha} < +\infty$, so condition (4.13) and (4.22) are also satisfied. Moreover the Brownian motion has independent increments so condition (4.14) and (4.23) are satisfied too. We simulate a Brownian motion on the interval $[0,1]$, since the variance does not affect the estimators, we set $\mathbb{V}ar\,(B(1)) = 1$. The estimators are set for $\delta t = 0.005$ and $n = 10$. The time change estimate for both estimators are shown in Figure 4.5. The green line on the graphs represent the theoretical θ' and the two dashed red lines the 95% confidence interval separated from the green line by twice σ_n, where $\sigma_n = \dfrac{\sigma^V_{B(1)}(1/H)}{\sqrt{2^n}\mu^V_{\tilde{X}}(1/H)}$ for the $1/H$-variation and $\sigma_n = \dfrac{\sigma^U_{B(1)}}{H\sqrt{2^n}}$ for the log-variation. Note that in Figure 4.5, the $1/H$-variation performs better than the log-variation estimator, which is expected.

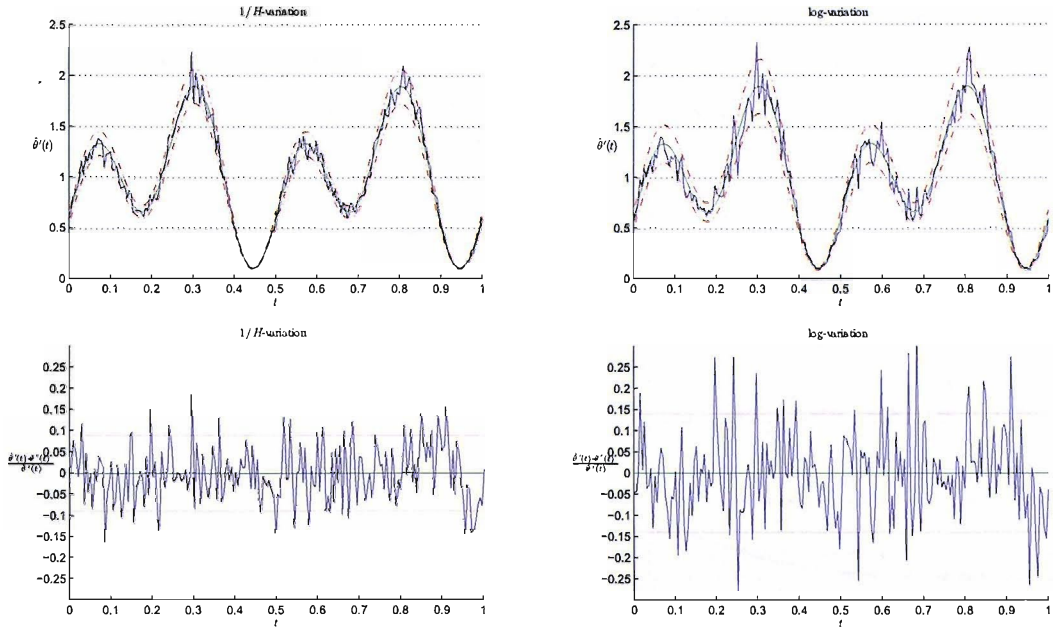

Figure 4.5: Estimate of θ' using the $1/H$-variation (Top left panel) and the log-variation (Top right panel) and the normalized bias bottom panel of a time-changed Brownian motion

In order to understand the effect of the number of observations in a sample, we simulate time changed Brownian motion so that the number of observations used to estimate the time change function in the interval of length δt is $N_{\text{obs}} \in \{10, 50, 100, 500, 1000\}$. We compute the MSE of our estimators for different values of δt ($\in \{0.1, 0.05, 0.001, 0.005\}$) for $n = 1, \ldots, 12$.

The results for 10000 runs are shown in Figures 4.6 and 4.7 respectively for the $1/H$-variation and the log-variation estimators. In both cases, we consider two types of interpolation; one taking the previous data as interpolation and the second one using a linear interpolation of the data. Note that we could have taken N_{obs} as a power of 2 in the interval of length δt, but we decided not to do so in order to get closer to a real situation in practice, where the number of observations are not specially a power of two. The dashed green line represent the theoretical MSE error in the case of a continuous time Brownian motion. The theoretical MSE are given by (4.30) for the $1/H$-variation method and by (4.40) for the log-variation estimator. The blue vertical dashed lines represents the logarithm in base 2 of N_{obs}.

We note that for $n > N_{\text{obs}}$, the MSE is more or less constant. That is showing that the choice of n depend on the number of observations in the interval of length δt. The optimum n noted n_{op} is given by $n_{op} = \max_{n \in \mathbb{N}}(n < \log_2(N_{\text{obs}}))$. In fact this last it is only applicable for the log-variation method taking as interpolation the previous data. In the other situation, when using a linear interpolation, a bigger n does not affect the estimator. (We will see later that in fact for irregularly spaced data choosing $n > n_{op}$ may affect the estimator in some case). Note also the choice of the interpolation is

important. While a previous data interpolation is favorable for the $1/H$-variation estimate, a linear interpolation is preferred for the log-variation estimate so one can avoid increments with value zero which affect the estimator.

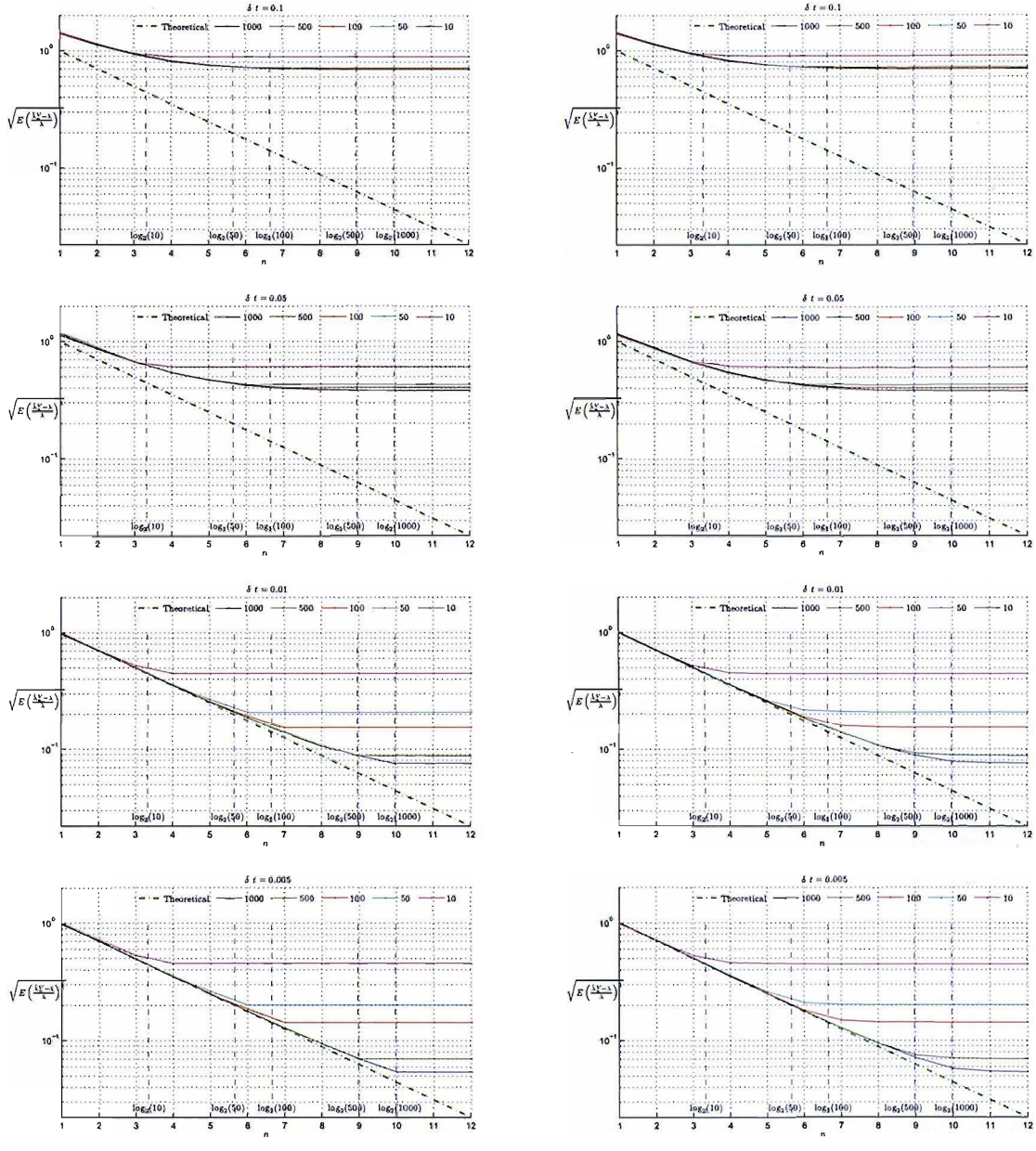

Figure 4.6: MSE of $\hat{\lambda}^V_{t,\delta t}$ from a time-changed Brownian motion for different values of δt and N_{obs}. Left panels it is when using a previous data interpolation and right panels is when using a linear interpolation of the data.

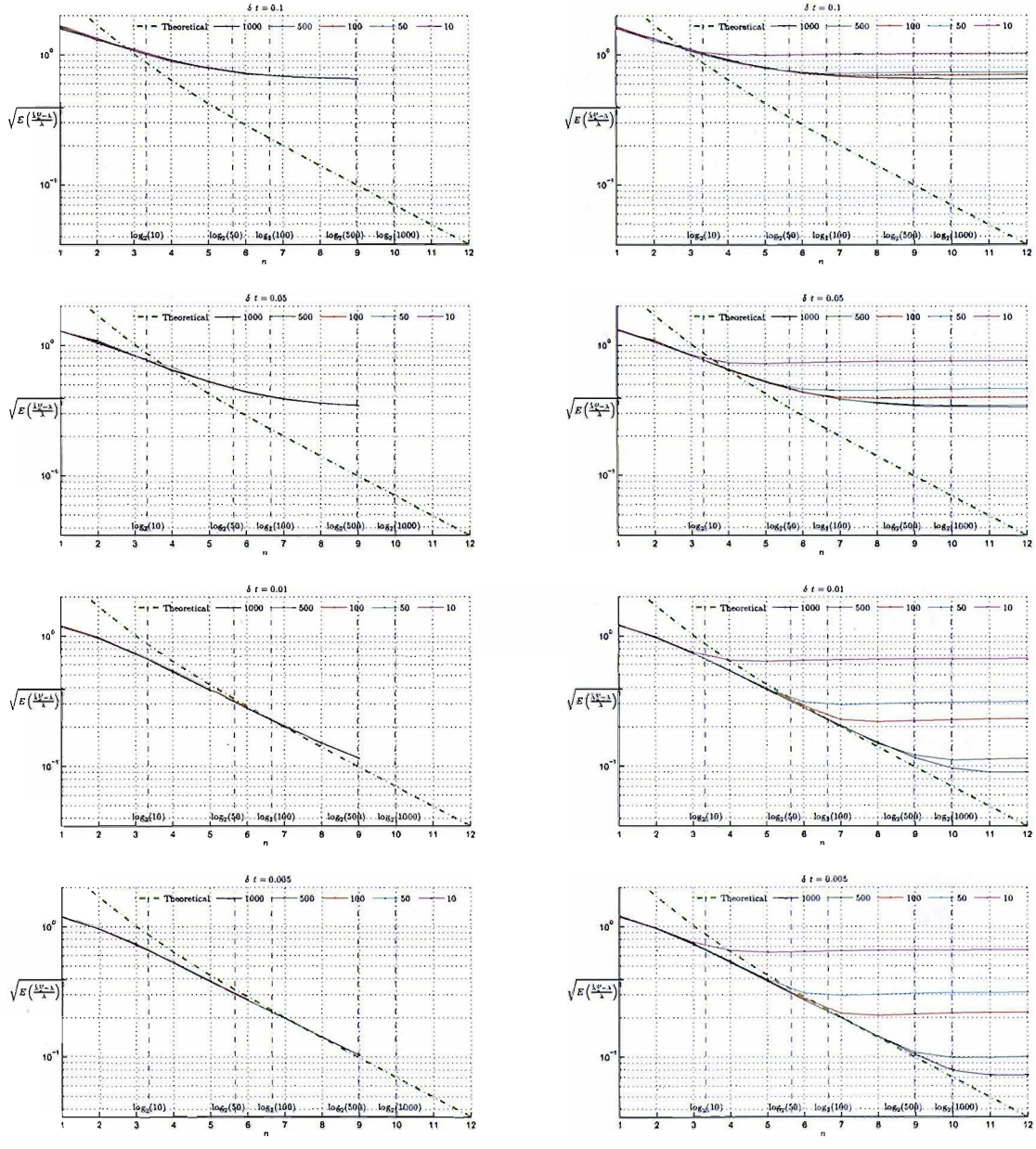

Figure 4.7: MSE of $\hat{\lambda}^U_{t,\delta t}$ from a time-changed Brownian motion for different values of δt and N_{obs}. Left panels it is when using a previous data interpolation and right panels is when using a linear interpolation of the data.

4.6.2 Time-changed Lévy stable motion

An interesting process to study is the Lévy stable motion $L_\alpha(t)$. This process has infinite variance for $\alpha \in (0, 2)$ and has infinite jumps. This study is interesting in the sense that we will test the performance of our estimators on a process with jumps. For $\alpha < 2$, $\mathbb{E}[|L_\alpha(1)|^\alpha] = +\infty$ (see [11]), so condition (4.13) is not satisfied and Theorem 4.2.2 can not be applied. We then discard the $1/H$-variation estimator for this study. For $0 < a < 1$ and $\alpha > a$, $\mathbb{E}|L_\alpha(1)|^a < +\infty$ and $\mathbb{E}|L_\alpha(1)|^{-a} < +\infty$, so condition (4.22) is satisfied and condition (4.23) also since Lévy processes has independent increments, Theorem 4.3.2 can be applied.
As for the time-changed Brownian motion, we study the time-changed Lévy motion. A way of simulating time change Lévy stable motion $L_\alpha(\theta(t))$ is to use the following equality in distribution

$$L_\alpha(\theta(t)) \overset{(d)}{=} \int_0^t (\theta'(s))^{1/\alpha} dL_\alpha(s) \tag{4.45}$$

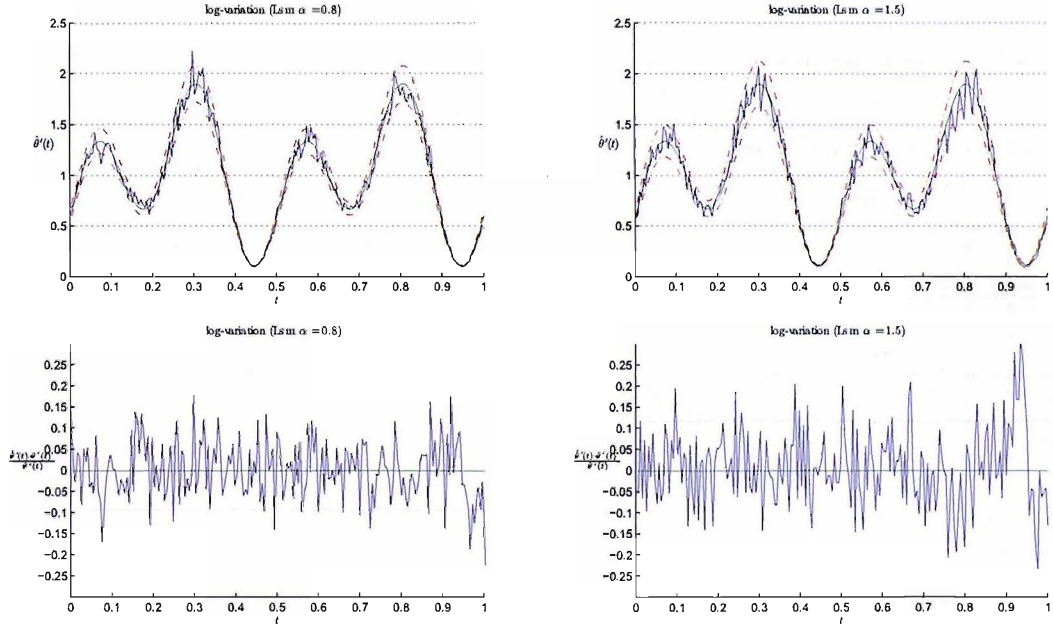

Figure 4.8: Estimate of θ' using the log-variation of a time-changed Lévy stable motion for $\alpha = 0.8$ (left panels) and $\alpha = 1.5$ (right panels). The top panels represent an estimate of θ' and bottom panels the normalized bias.

We perform a similar study as for the time-changed Brownian motion; we simulate $L_\alpha(\theta(t))$ on the interval $[0, 1]$, with parameters $\alpha = \{0.8, 1.5\}$. The estimator given by the log-variation are set for $\delta t = 0.005$ and $n = 10$. The results of the time-change estimation are shown in Figure 4.8. The green line on the graphs represent the theoretical θ' function and the two dashed red lines the 95% confidence interval

separated from the green line by twice $\sigma_n = \dfrac{\alpha\sigma^U_{L_\alpha(1)}}{\sqrt{2^n}}$. Unlike the Brownian motion, $\sigma^U_{L_\alpha(1)}$ is not given explicitly so we estimate it using a Monte Carlo simulation.

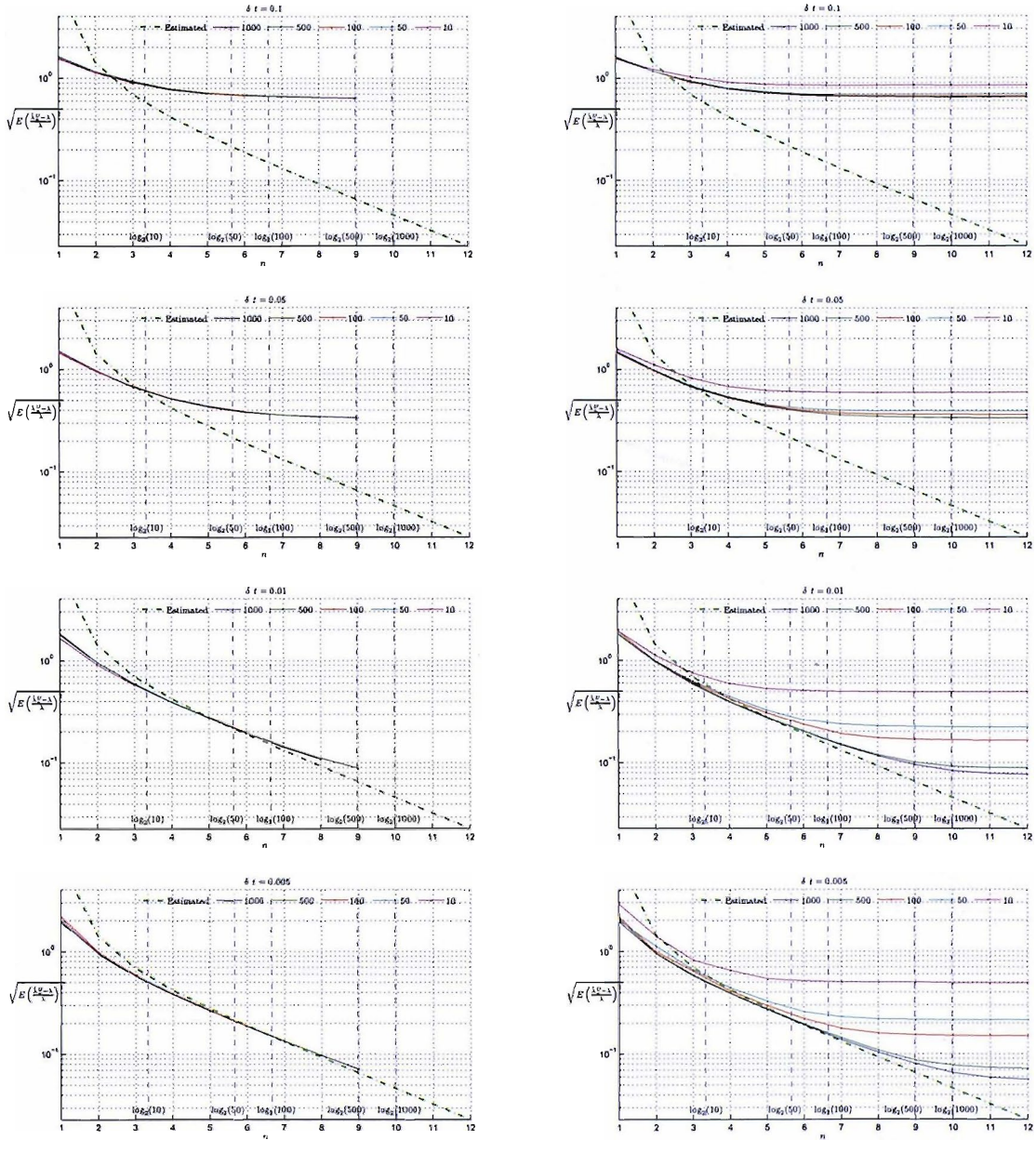

Figure 4.9: MSE of $\hat{\lambda}^U_{t,\delta t}$ from a time-changed Lévy stable motion with $\alpha = 0.8$, for different values of δt and N_{obs}. Left panels it is when using a previous data interpolation and right panels is when using a linear interpolation of the data.

One can see that the estimator performs better as α gets smaller; in other words when H gets bigger. This may appear paradoxical since for small values of α, the Lévy stable motion has big sizes jumps. We mean by paradoxical counter-intuitive, since we estimate the time change function using the path variation of the process, we would think that for a process with jumps, the estimation of the time change function would be affected by the huge size of process jumps. In fact, the equation 4.38 shows that

the MSE decreases as α decreases.

The MSE for different value of δt and the number of observations are shown in Figures 4.9 and 4.10. As expected, the log-variation estimator performs well on H-sssi process with jumps as long as this last satisfies the log-variation method hypothesis. As for the time-changed Brownian motion, a similar commentary can be made.

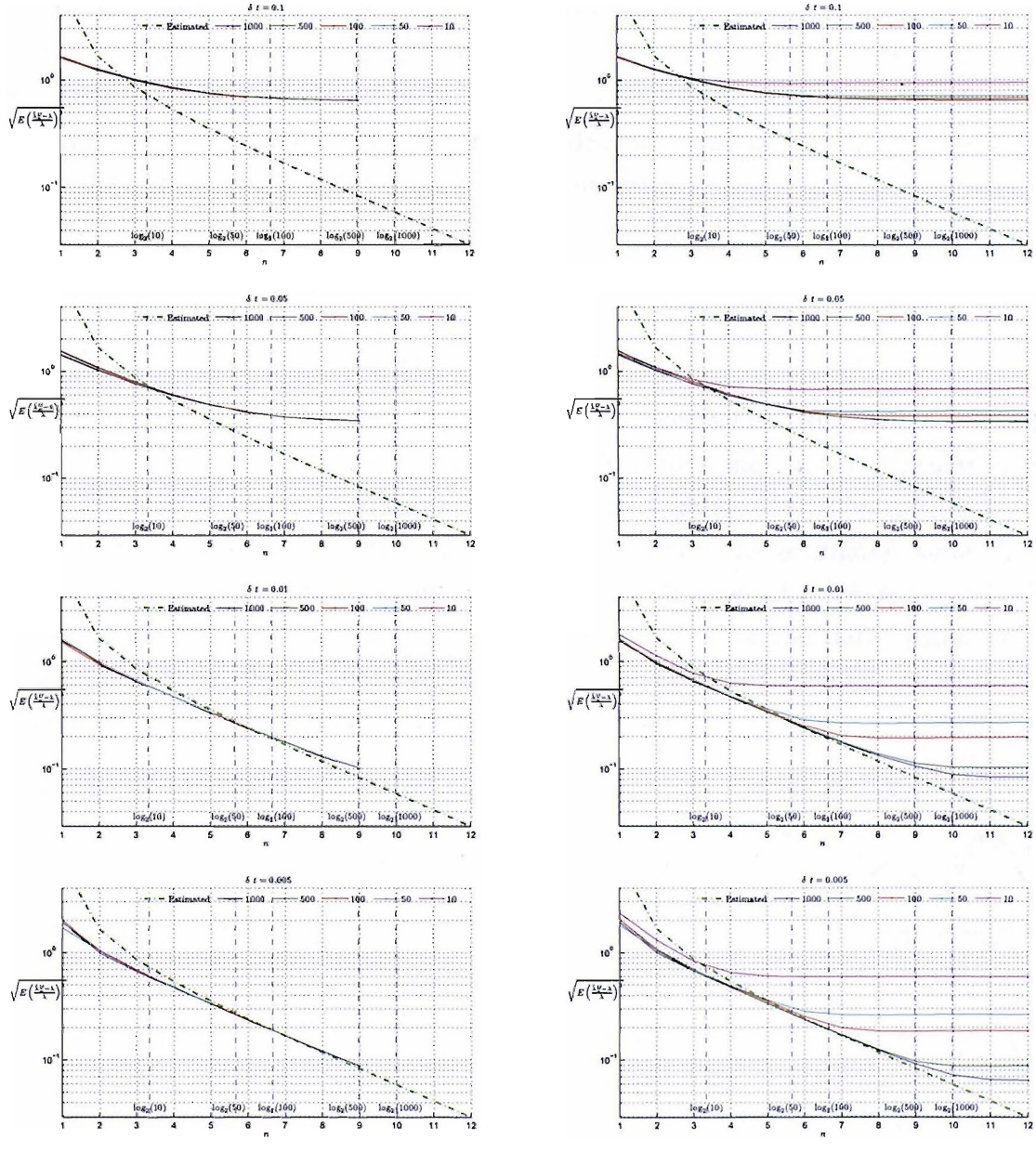

Figure 4.10: MSE of $\hat{\lambda}^{U}_{t,\delta t}$ from a time-changed Lévy stable motion with $\alpha = 1.5$, for different values of δt and N_{obs}. Left panels it is when using a previous data interpolation and right panels is when using a linear interpolation of the data.

4.6.3 Time-changed fractional Brownian motion

An interesting case of time-changed H-sssi process, here we can test the performance of our estimators in the case where the increments of the process are dependent, stationary and ergodic. The fractional Brownian motion satisfies conditions of Propositions 4.2.3 and 4.3.5. The $1/H$-variation and the log-variation estimators can be applied.
Keeping the same setting as for the Lévy stable motion for our estimators, we test the performance of our estimators for fBm with $H = 0.3$ and $H = 0.8$. The results are shown in Figures 4.12 and 4.14 for $H = 0.3$, Figures 4.13 and 4.15 for $H = 0.8$. The dashed red lines are the theoretical MSE when assuming the process having independent and Gaussian distributed increments with $H \in \{0.3, 0.8\}$. The green line is the MSE for continuous fBm with their respective H. This was estimated using a Monte carlo simulation over 1000 runs.

4.6.4 Comment irregularly spaced time data

For irregularly spaced time data, the estimator is biased when using a linear interpolation of the data. We simulate a time-changed Brownian motion on a regular spatial grid with 100000 observations. We estimate the time-changed function from this sample. Setting $T = 1$, $\delta t = 0.01$ and $n = 10$, we obtain the estimated time change function using the $1/H$-variation and log-variation in Figure 4.11. We see clearly that the estimator is biased when using a linear interpolation of the data, while for the previous data interpolation, the $1/H$-variation is not biased and the log-variation does not perform as there are many increments with value 0.

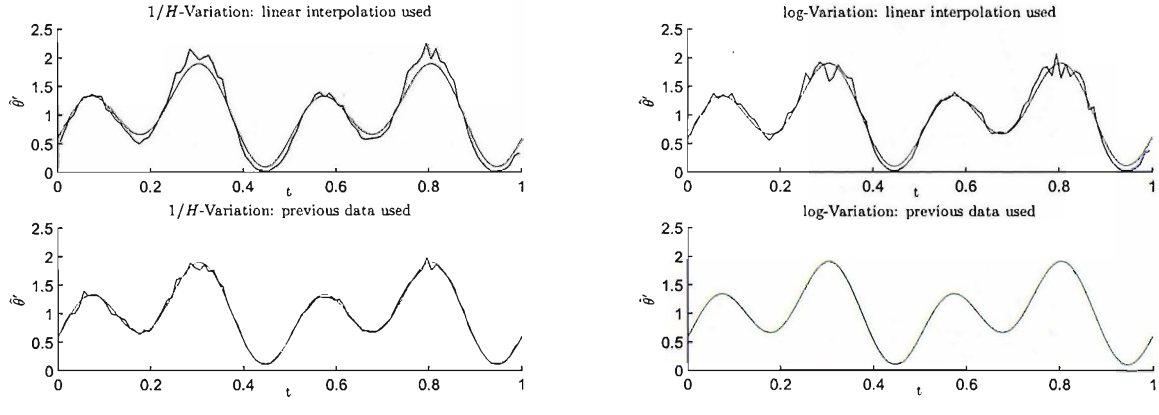

Figure 4.11: Time change estimate from a Brownian motion simulated on a regular spacial grid. The top panels when using the linear interpolation and the bottom panels when using a previous data interpolation. The green lines represents the theoretical time-change applied and the blue lines the estimated time-change function.

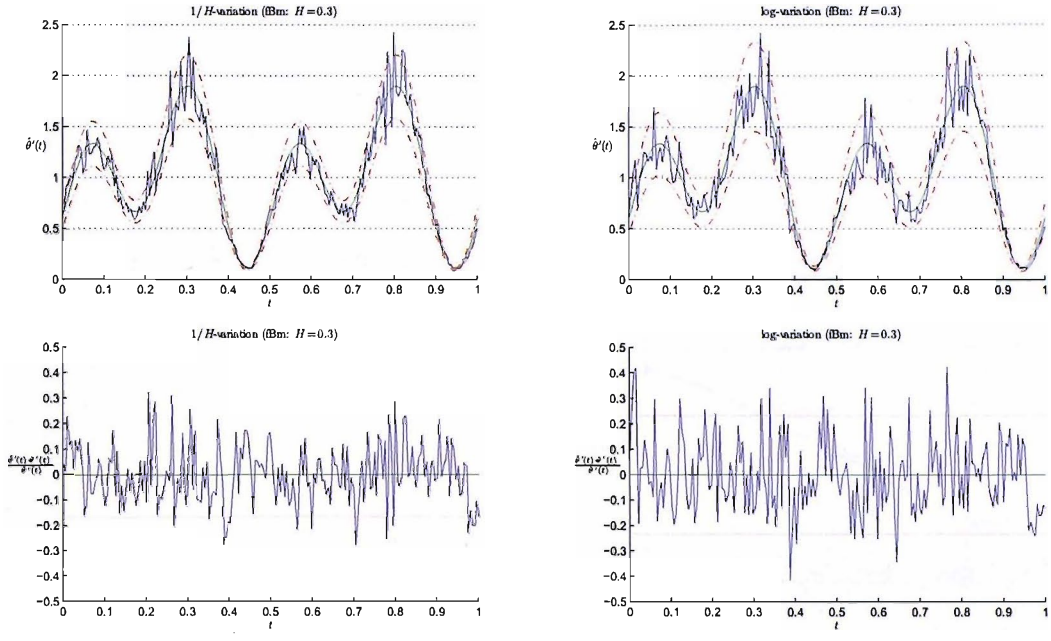

Figure 4.12: Estimate of θ' from a time-changed fractional Brownian motion with $H = 0.3$ using the $1/H$-variation (left panels) and the log-variation (right panels). The top panels represent an estimate of θ' and bottom panels the normalized bias of it.

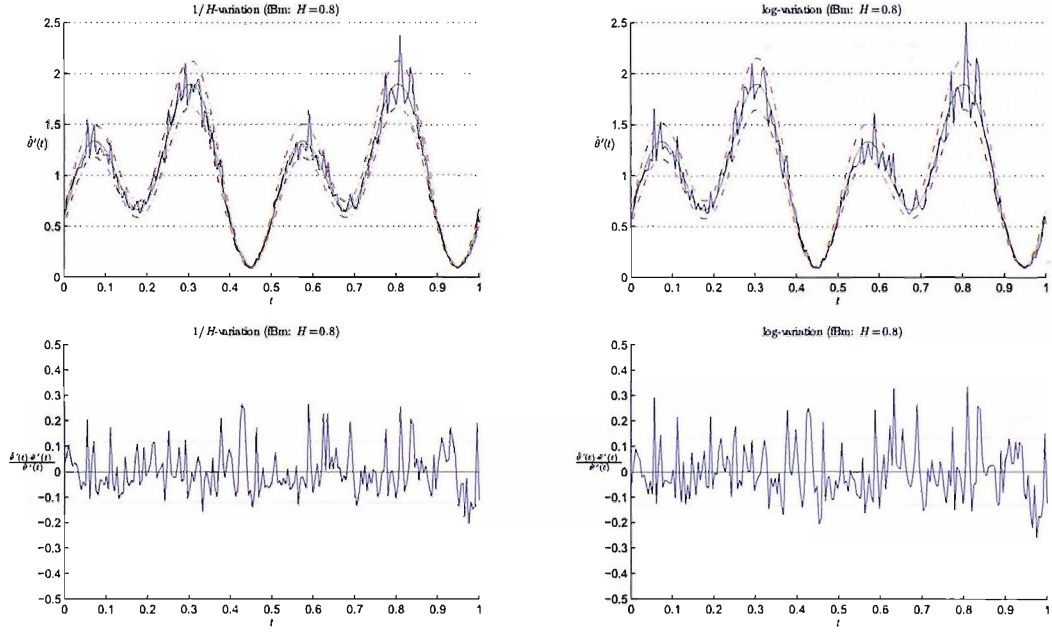

Figure 4.13: Estimate of θ' from a time-changed fractional Brownian motion with $H = 0.8$ using the $1/H$-variation (left panels) and the log-variation (right panels). The top panels represent an estimate of θ' and bottom panels the normalized bias of it.

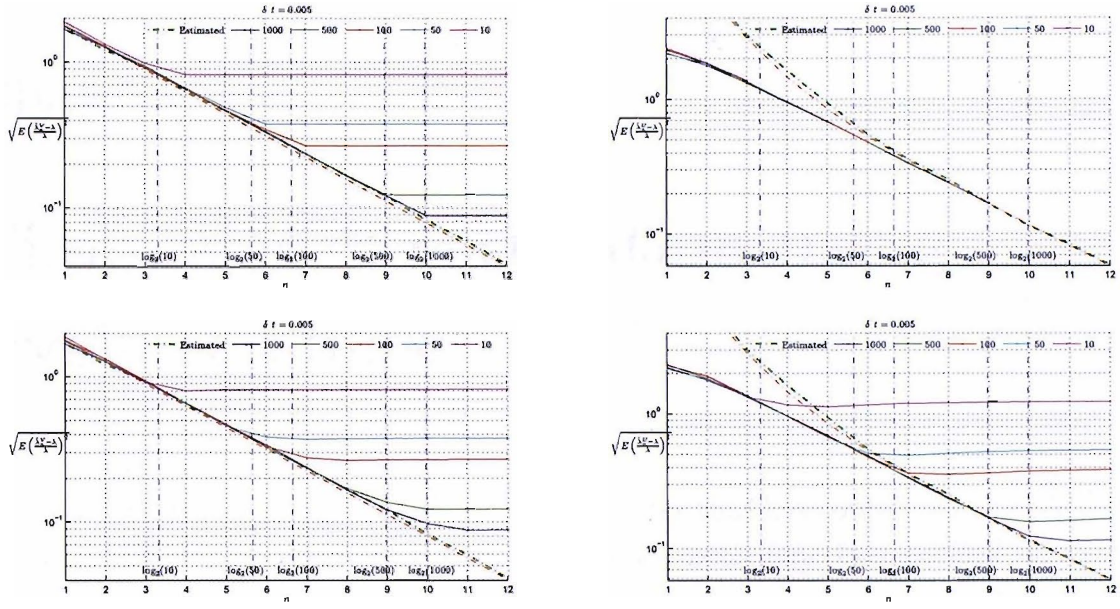

Figure 4.14: MSE of $\hat{\lambda}^{V}_{t,\delta t}$ (left panels) and $\hat{\lambda}^{u}_{t,\delta t}$ (right panels) from a time-changed fBm with $H = 0.3$, $\delta t = 0.005$ and $N_{\text{obs}} = \{10, 50, 100, 500, 1000\}$. Top panels shows the MSE when using a previous data interpolation and bottom panels when performing a linear interpolation of the data.

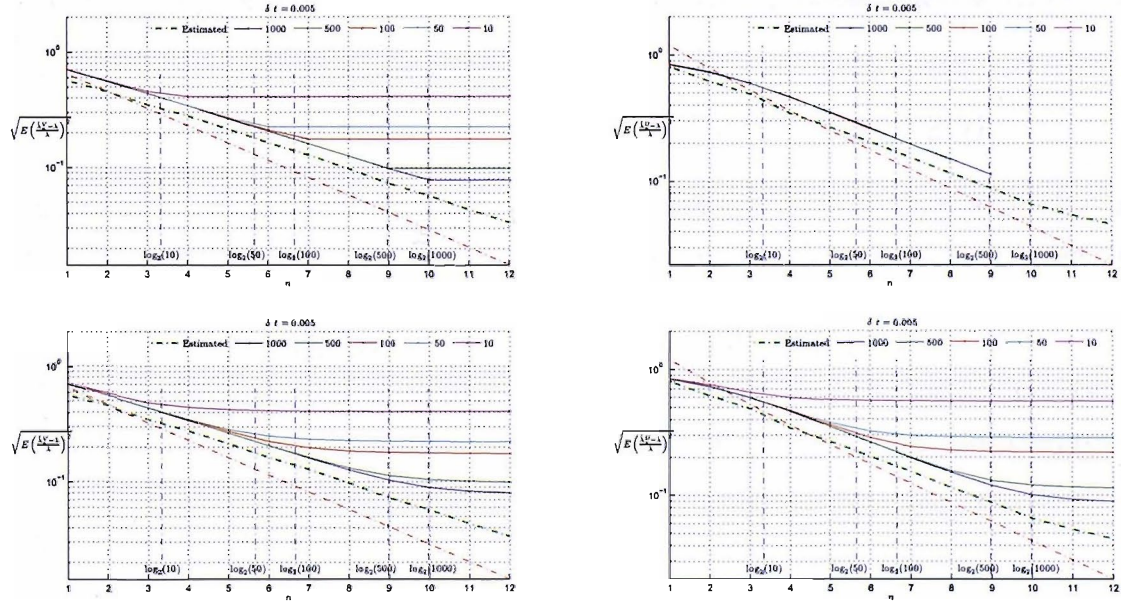

Figure 4.15: MSE of $\hat{\lambda}^{V}_{t,\delta t}$ (left panels) and $\hat{\lambda}^{u}_{t,\delta t}$ (right panels) from a time-changed fBm with $H = 0.8$, $\delta t = 0.005$ and $N_{\text{obs}} = \{10, 50, 100, 500, 1000\}$. Top panels shows the MSE when using a previous data interpolation and bottom panels when performing a linear interpolation of the data.

Chapter 5

Time-change estimation using level-crossings

In all the Chapter, we consider crossings of processes that has continuous sample path. We note $C([0,T])$ the space of continuous sample paths functions on $[0,T]$ for some $T > 0$.

5.1 Preliminaries

Let $X = \{X(t)\}_{\{t\geq 0\}}$ be a stochastic process so that $X(0) \overset{(a.s.)}{=} 0$ and X has continuous sample path. We define $T^\delta = \{T_i^\delta\}_{\{i\in\mathbb{N}\}}$ the crossing time process of levels $\delta\mathbb{Z}$ ($\delta > 0$), such that

$$T_0^\delta = 0 \text{ and } \forall i \in \mathbb{N}, \quad T_{i+1}^\delta = \inf\{t \geq T_i^\delta : |X(t) - X(T_i^\delta)| \geq \delta\}$$

We define a crossing of level $k\delta$ for $k \in \mathbb{Z}$, if the previous crossing is of level $(k-1)\delta$ or $(k+1)\delta$. We note $\mathcal{X}(\delta\mathbb{Z})$ the sets of crossing points. This type of crossing has been already defined in [7] in order to estimate the Hurst index. An illustration of the crossing points for a particular sample path is presented in Figure 5.1.
We denote by $\mathcal{N}_{(s,t)}^\delta(X)$ the number of crossing points of levels $\delta\mathbb{Z}$ of the process X in the interval (s,t) by the number of crossing points made by the shifted process $\breve{X}_s = \{\breve{X}_s(t) = X(t+s) - X(s)\}_{\{t\geq 0\}}$ in the interval $(0, t-s)$. The first crossing point, $\breve{X}_s(s) = 0$ almost surely, is not considered as crossing. note that

$$\mathcal{N}_{(s,t)}^\delta(X) = \mathcal{N}_{(0,t-s)}^\delta(\breve{X}_s) \tag{5.1}$$

Four types of crossing points can be considered (see [69]): up-crossing, down-crossing, tangency from below and tangency from above[1]. We respectively note their corre-

[1] We will show later that the tangencies are not likely to happen

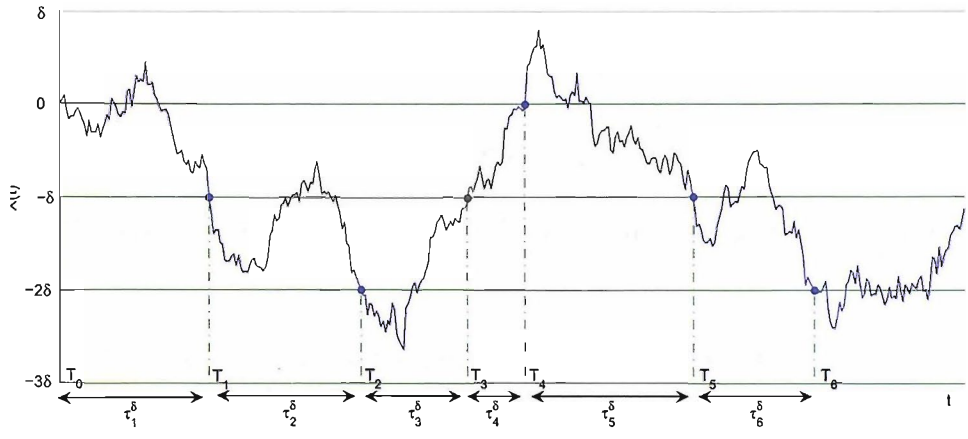

Figure 5.1: Crossing times.

sponding set $\mathcal{X}^U(\delta\mathbb{Z})$, $\mathcal{X}^D(\delta\mathbb{Z})$, $\mathcal{X}^B(\delta\mathbb{Z})$ and $\mathcal{X}^A(\delta\mathbb{Z})$. One has

$$\mathcal{X}(\delta\mathbb{Z}) = \mathcal{X}^U(\delta\mathbb{Z}) \cup \mathcal{X}^D(\delta\mathbb{Z}) \cup \mathcal{X}^B(\delta\mathbb{Z}) \cup \mathcal{X}^A(\delta\mathbb{Z})$$

We present the different crossing types stated by Cramer and Leadbetter in [70]. Let f be a continuous function on $\mathbb{R}^+$ and $k \in \mathbb{Z}$.

- f is said to have an up-crossing of level $k\delta$ at time t_0, if the previous crossing point or tangency[2] was of level $(k-1)\delta$ and there exists $\epsilon > 0$ such that $f(t) > k\delta$ for $t \in (t_0, t_0 + \epsilon)$.

- f is said to have a down-crossing of level $k\delta$ at time t_0, if the previous crossing point or tangency was of level $(k+1)\delta$ and there exists $\epsilon > 0$ such that $f(t) < k\delta$ for $t \in (t_0, t_0 + \epsilon)$.

- f is said to have a tangency from above of level $k\delta$ at time t_0, if $f(t_0) = k\delta$, the previous crossing point or tangency was of level $(k+1)\delta$ and there exists $\epsilon > 0$ such that $f(t) \geq k\delta$ for $t \in (t_0, t_0 + \epsilon)$.

- f is said to have a tangency from below of level $k\delta$ at time t_0, if $f(t_0) = k\delta$, the previous crossing point or tangency was of level $(k-1)\delta$ and there exists $\epsilon > 0$ such that $f(t) \leq k\delta$ for $t \in (t_0, t_0 + \epsilon)$.

Ylvisaker, [69] and [71], shows that for Gaussian sample path, tangencies does not occur almost surely. We show in the next Lemma that for our process X, tangencies does not occur almost surely under some assumption.
We note first that the number of crossings in an interval $(s, s+\epsilon)$, where s is a positive constant and $\epsilon > 0$, is finite for processes with continuous sample path. Indeed, if the number of crossings is infinite, then there is an accumulation point of crossings,

[2] As it will be shown that tangencies does not occur almost surely, only crossing point of type down or up-crossing will be considered.

that is a point with infinitely many crossings in the neighborhood of s, so X must be discontinuous.

Lemma 5.1.1. *Let $X \in C([0,+\infty))$, so that $\mathcal{N}^{\delta}_{(0,1)}(X) < +\infty$ almost surely. We assume that for every $t \geq 0$, $X(t)$ is a continuous random variable, and*

$$\mathbb{P}(X \text{ has a upcrossing of } \delta \text{ in } (s,t)) \text{ is a continuous function of } \delta \tag{5.2}$$

Then

$$\mathbb{P}(X \text{ has a tangent crossing from below of } \delta \in (s,t)) = 0$$

Proof. We fix a real δ, and s and t two positive real such that $s < t$. We note first that as $X(u)$ is a continuous random variable for all u, one has

$$\mathbb{P}(\delta \in \{X(u) : u \in Q\}) = 0$$

Thus

$$\mathbb{P}(X(u) = \delta \text{ for all } u \in (s,t)) = 0$$

So X has no flat spot almost surely.
Split the interval (s,t) in n segments of length Δ, as illustrated in Figure 5.2.

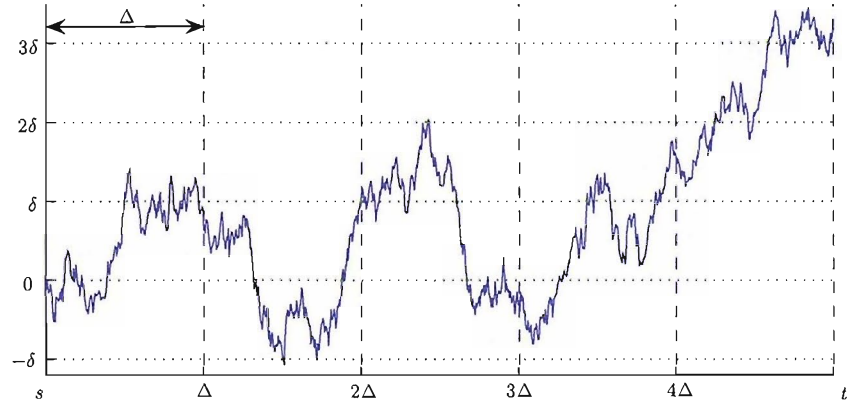

Figure 5.2: Subdivision of the interval (s,t).

Let $I^{\delta}_k = 1$ if X has an upcrossing at level δ in the k^{th} interval, and 0 otherwise. Similarly let $J^{\delta}_k = 1$ if X has a tangent crossing at level δ from below in the k^{th} interval, and 0 otherwise. We set

$$U^{\delta}_{\Delta} = \sum_{k=1}^{n} I^{\delta}_k$$

$$B^{\delta}_{\Delta} = \sum_{k=1}^{n} J^{\delta}_k$$

Then since any upcrossing of level δ will also be an upcrossing at level $\delta - \epsilon$ for $\epsilon > 0$ small enough, and any tangent crossing of δ from below will give an upcrossing of $\delta - \epsilon$ for $\epsilon > 0$ small enough, hence

$$U_\Delta^\delta + B_\Delta^\delta \leq \liminf_{\epsilon \to 0} U_\Delta^{\delta - \epsilon}$$

By taking the expectation and then using Fatou's lemma (since the number of crossing in the interval is assumed finite), one gets

$$\mathbb{E}\left[B_\Delta^\delta\right] \leq \liminf_{\epsilon \to 0} \mathbb{E}\left[U_\Delta^{\delta-\epsilon}\right] - \mathbb{E}\left[U_\Delta^\delta\right]$$

Since the right hand side is bounded and $\mathbb{P}(X$ has a upcrossing of δ in $(s,t))$ is a continuous function of δ, the right hand side of the inequality converges to 0 and we have $\mathbb{E}\left[B_\Delta^\delta\right] = 0$. As $B_\Delta^\delta \geq 0$, one has $B_\Delta^\delta \overset{(a.s.)}{=} 0$.
Since X has no flat spot, for any tangent crossing of δ, there must be an ϵ such that there are no other tangent crossings within distance ϵ. Thus $B_\Delta^\delta \to \#$ of tangent crossings from below of level δ in (s,t), as $\Delta \to 0$. This ends the proof. □

Remark 5.1.2. *An analogous argument works for tangent crossings from above, where $\mathbb{P}(X$ has a downcrossing of δ in $(s,t))$ is a continuous function of δ instead of assumption (5.2).*

Remark 5.1.3. *Proof of Lemma 5.1.1 is based on Leadbetter, Lindgren, and Rootzen in [72], Theorem 7.2.5.*

Since tangencies do not happen almost surely, the crossings are only defined by its up-crossings and down-crossings. We set C^δ the space of function $f \in C([0,+\infty))$ such that the tangencies of the function f at level δ does not occur.

5.2 Crossing points of process with stationary and ergodic increments

Let $X = \{X(t)\}_{t \geq 0}$ be a stochastic process with stationary increments and has continuous sample path. We show that for any $t \geq 0$, the process $\{\mathcal{N}_{(m,m+1)}^\delta(X)\}_{m \geq 1}$ is stationary, we also show that $\{\mathcal{N}_{(m,m+k)}^\delta(X)\}_{k \geq 0}$ and $\{\mathcal{N}_{(0,k)}^\delta(X)\}_{k \geq 1}$ have the same joint distribution. Moreover, if the process X has ergodic increments, we show that the process $\{\mathcal{N}_{(m,m+1)}^\delta(X)\}_{m \geq 1}$ is also ergodic. Since tangencies do not occur almost surely, we consider crossing points as being either up-crossing or down-crossing.
Having $C([0,+\infty))$ the space of continuous functions, we define the metric ρ from $C([0,+\infty)) \times C([0,+\infty))$ to $[0,+\infty)$ corresponding to the uniform convergence on compact subsets defined by $\forall (f,g) \in C([0,+\infty)) \times C([0,+\infty))$ and $I \subset [0,+\infty)$,

$$\rho(f,g) = \sup_{t \in I} |f(t) - g(t)|$$

Basically, the subset interval I will be chosen as the interval of time in which the number of crossings of the process X is counted.

5.2.1 Continuity of the function $f \mapsto \mathcal{N}^{\delta}_{(0,1)}(f)$

We define the function h from $C([0,+\infty))$ to $\mathbb{N}$ by

$$h : f \mapsto \mathcal{N}^{\delta}_{(0,1)}(f) \tag{5.3}$$

Lemma 5.2.1. *The function h, as defined in (5.3), is continuous on C^{δ}.*

Proof. Let $f \in C^{\delta}$ and f_{2^n} defined as in (5.6) which is also in C^{δ}, and $f_{2^n} \to f$. To prove the continuity of the function h from C^{δ} to $\mathbb{N}$, we show

$$\lim_{n \to +\infty} h(f_{2^n}) = h(f)$$

Clearly $\forall n \in \mathbb{N}$, $h(f_{2^n}) \leq h(f)$. One has

$$\lim_{n \to +\infty} h(f_{2^n}) \leq h(f) \tag{5.4}$$

Now let $h(f) = m > 0$, where m is a finite integer; that would mean, since the tangencies do not happen almost surely, that there is, say crossing points at $t_1 < \cdots < t_m$ in $(0,1)$ at level respectively $\delta k_1, \ldots, \delta k_m$, for $k_i \in \mathbb{Z}$. One can find, for $\epsilon_i > 0$, disjoint subintervals $(t_i - \epsilon_i, t_i + \epsilon_i)$, so that for n_0 large enough and $j_i \in \{0, \ldots, 2^{n_0}\}$, $j_i/2^{n_0} \in (t_i - \epsilon_i, t_i)$ and $(j_i+1)/2^{n_0} \in (t_i, t_i + \epsilon_i)$. As $h(f_{2^n})$ increases with n, it implies for $n \geq n_0$, $h(f_{2^n}) \geq m$, hence

$$\lim_{n \to +\infty} h(f_{2^n}) \geq h(f) \tag{5.5}$$

whence by (5.4) and (5.5), $\lim_{n \to +\infty} h(f_{2^n}) = h(f)$. □

Remark 5.2.2. *The last proof is based on Cramer and Leadbetter in [70].*

Lemma 5.2.3. *Let Z be a stochastic process so that the function $t \mapsto Z(t,\omega) \in C([0,+\infty))$. We define Z_n so that for all $t \in [0,+\infty)$*

$$Z_n(t) = \sum_{i=1}^{n} ((nt - (i-1))Z(i/n) + (i - nt)Z((i-1)/n))\, \mathbf{1}_{\{t \in ((i-1)/n, i/n]\}} \tag{5.6}$$

Then $Z_n \overset{(a.s.)}{\to} Z$ as $n \to +\infty$.

Proof. To show that $Z_n \overset{(a.s.)}{\to} Z$, we need to show that the sample path of Z_n converges to the one of Z. This is traduced by showing that for $\omega \in \Omega \backslash N$, where $N \subset \Omega$ and $\mathbb{P}(N) = 0$,

$$\sup_{t \in [0,1]} |Z_n(t,\omega) - Z(t,\omega)| \to 0, \quad \text{as } n \to +\infty \tag{5.7}$$

Since in (5.6) $i = \lfloor nt \rfloor + 1$, one has for all $t \in [0, +\infty)$

$$
\begin{aligned}
&|Z_n(t) - Z(t)| \\
&\quad \leq \sum_{i=1}^{n} ((nt - (i-1))|Z(i/n) - Z(t)| + (i - nt)|Z((i-1)/n) - Z(t)|)\, \mathbf{1}_{\{t \in ((i-1)/n, i/n]\}} \\
&\quad \leq \sup_t |Z((\lfloor nt \rfloor + 1)/n) - Z(t)| + \sup_t |Z(\lfloor nt \rfloor / n) - Z(t)|
\end{aligned}
$$

Since $\lfloor nt \rfloor / n \to t$ and $(\lfloor nt \rfloor + 1)/n \to t$ as $n \to +\infty$, and by continuity of the sample path of Z, one gets (5.7). By completeness, we have

$$Z_n \overset{(a.s.)}{\to} Z, \text{ as } n \to +\infty$$

□

5.2.2 Stationarity of the process $\{\mathcal{N}^{\delta}_{(0,m)}(X)\}_{m \geq 1}$

Lemma 5.2.4. *Let h be a continuous function from C^{δ} to $\mathbb{N}$.*
Let $X = \{X(t)\}_{t \geq 0}$ and $Y = \{Y(t)\}_{t \geq 0}$ two stochastic processes in C^{δ} so that $X \overset{(fdd)}{=} Y$.
Then $h(X) \overset{(d)}{=} h(Y)$.

Proof. It suffices to show for any bounded continuous function $f : \mathbb{R} \mapsto \mathbb{R}$,

$$\mathbb{E}\left[f(h(X))\right] = \mathbb{E}\left[f(h(Y))\right]$$

Let $X_n(t) = X(\lfloor nt \rfloor / n)$ and $Y_n(t) = Y(\lfloor nt \rfloor / n)$. One has for any $t \geq 0$,

$$\lfloor nt \rfloor / n \to t, \text{ as } n \to +\infty$$

and so by continuity of the sample path of X and Y, one has $X_n \overset{(a.s.)}{\to} X$ and $Y_n \overset{(a.s.)}{\to} Y$ as $n \to +\infty$. By continuity of the function h,

$$h(X_n) \overset{(a.s.)}{\to} h(X) \text{ and } h(Y_n) \overset{(a.s.)}{\to} h(Y), \text{ as } n \to +\infty$$

Let now ϕ be a continuous function from the space of random vector $\mathbb{R}^{n+1}$ to the space of sample path $C([0, +\infty))$, so that $X_n = \phi(\{X(0), X(1/n), \ldots, X(1)\})$ (similarly for Y_n). The composition of two continuous function is continuous, so since $X \overset{(fdd)}{=} Y$

$$h \circ \phi(\{X(0), X(1/n), \ldots, X(1)\}) \overset{(d)}{=} h \circ \phi(\{Y(0), Y(1/n), \ldots, Y(1)\})$$

Whence $h(X_n) \overset{(d)}{=} h(Y_n)$. Now let f be any a bounded continuous function, one has

$$\begin{aligned}|\mathbb{E}[f(h(X))] - \mathbb{E}[f(h(Y))]| \leq & |\mathbb{E}[f(h(X))] - \mathbb{E}[f(h(X_n)))]| \\ & + |\mathbb{E}[f(h(X_n))] - \mathbb{E}[f(h(Y_n))]| \\ & + |\mathbb{E}[f(h(Y_n))] - \mathbb{E}[f(h(Y))]|\end{aligned}$$

Tending n to infinity, one gets $\mathbb{E}[f(h(X))] = \mathbb{E}[f(h(Y))]$. □

Lemma 5.2.5. *Let H a function defined by*

$$H : Z \to \{h_1(Z), h_2(Z), \dots\}$$

where Z is a process with continuous sample path in C^δ and $\{h_i\}_{i\geq 1}$ is a sequence of continuous functions on C^δ to $\mathbb{N}$.
Let $X = \{X(t)\}_{t\geq 0}$ and $Y = \{Y(t)\}_{t\geq 0}$ two stochastic processes in C^δ so that $X \overset{(fdd)}{=} Y$. Then $H(X) \overset{(fdd)}{=} H(Y)$.

Proof. This is a generalisation of Lemma 5.2.4. □

Using Lemmas 5.2.4 and 5.2.5, we deduce the following Theorem.

Theorem 5.2.6. *Let X be a real-valued stochastic process with stationary increments and whose sample path is continuous. For $0 < s < t$, let $\mathcal{N}^\delta_{(s,t)}(X)$ the number of crossing points of the process X in the interval (s,t) as defined earlier. Then*

(i) $\{\mathcal{N}^\delta_{(m,m+k)}(X)\}_{k\geq 1} \overset{(fdd)}{=} \{\mathcal{N}^\delta_{(0,k)}(X)\}_{k\geq 1}$, *for any* $m \geq 1$

(ii) $\{\mathcal{N}^\delta_{(n+m,n+m+k)}(X)\}_{m\geq 0} \overset{(fdd)}{=} \{\mathcal{N}^\delta_{(m,m+k)}(X)\}_{m\geq 0}$, *for any* $k \geq 1$ *and* $n \geq 1$.

Proof. (i) For $k \geq 1$, we set

$$h_k(.) \to \mathcal{N}^\delta_{(0,k)}(.) \text{ and } H(.) \to \{h_1(.), h_2(.), \dots\}$$

Using the shifted process at m, one has

$$\breve{X}_m = \{X(t+m) - X(m)\}_{t\geq 0} \overset{(fdd)}{=} \{X(t) - X(0)\}_{t\geq 0} = \breve{X}_0$$

From Lemma 5.2.5, we have by continuity of the functions h_k on C^δ (see Lemma 5.2.1), $H(\breve{X}_m) \overset{(fdd)}{=} H(\breve{X}_0)$. using the equality (5.1), one gets (i).

(ii) For fixed $k \geq 1$ and $m \geq 0$, we set functions

$$g_m(.) \to \mathcal{N}^\delta_{(m,m+k)}(.) \text{ and } G(.) \to \{g_0(.), g_1(.), \dots\}$$

Using the shifted process at n, one has

$$\breve{X}_n = \{X(t+n) - X(n)\}_{t\geq 0} \overset{(fdd)}{=} \{X(t) - X(0)\}_{t\geq 0} = \breve{X}_0$$

From Lemma 5.2.5, we have by continuity of the functions g_m on C^δ (see Lemma 5.2.1), $G(\breve{X}_n) \overset{(fdd)}{=} G(\breve{X}_0)$. using the equality (5.1), one gets (ii).

□

5.2.3 Ergodicity

Definition 5.2.7 ([73]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A Transform T is called measure preserving if for any event $A \subset \mathcal{F}$, $\mathbb{P}(T^{-1}(A)) := \mathbb{P}(\{\omega \in \Omega : T(\omega) \in A\}) = \mathbb{P}(A)$. The event A is called invariant (with respect to T) if $T^{-1}(A) = A$. Moreover, T is ergodic if for any invariant set A, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.*

We define the shift transformation T so that

$$T(\{X(0)-X(0), X(1/n)-X(0), \dots\}) = \{X(1)-X(1), X(1/n+1)-X(1), \dots\} \quad (5.8)$$

Sine the process X has stationary and ergodic increments, T is measure preserving and ergodic. Using the shifted process, we can rewrite (5.8) as

$$T(\{\breve{X}_0(0), \breve{X}_0(1/n), \dots\}) = \{\breve{X}_1(0), \breve{X}_1(1/n), \dots\} \quad (5.9)$$

Lemma 5.2.8. *For some $t > 0$, the process $\{\mathcal{N}^\delta_{(m,m+1)}(X)\}_{m\geq 0}$ is ergodic.*

Proof. Let $h_m. \mapsto \mathcal{N}^\delta_{(m,m+1)}(.)$.
Since the process $\{\mathcal{N}^\delta_{(m,m+1)}(X)\}_{m\geq 0}$ is stationary, it is enough to show that for every continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ such that $\mathbb{E}[f(h_0(X))] < +\infty$

$$\frac{1}{m}\sum_{i=0}^{m-1} f(h_i(X)) \overset{(a.s.)}{\to} \mathbb{E}[f(h_0(X))] \quad (5.10)$$

Setting $h = h_0$ and $\breve{X}_i$ the shifted process of X at time i, equality (5.10) becomes

$$\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) \overset{(a.s.)}{\to} \mathbb{E}\left[f(h(\breve{X}_0))\right] \quad (5.11)$$

We now set $\breve{X}_i^{(n)}$ the linear interpolation of $\breve{X}_i$ at point $\{\breve{X}_i(0), \breve{X}_i(1/n), \dots\}$ and ψ a continuous function from $\mathbb{R}^n$ to the space of continuous sample path, so that

$$\breve{X}_i^{(n)} = \psi(\{\breve{X}_i(0), \breve{X}_i(1/n), \dots\})$$

Using (5.9), we have

$$h(\breve{X}_i^{(n)}) = h \circ \psi(\{\breve{X}_i(0), \breve{X}_i(1/n), \dots\}) = h \circ \psi(T^i\{\breve{X}_0(0), \breve{X}_0(1/n), \dots\})$$

By continuity of the function $h \circ \psi$ and knowing T ergodic, the sequence $\{h(\breve{X}_i^{(n)})\}$ is ergodic. We now show (5.11), that implies (5.10). For any $n > 1$, one has

$$\begin{aligned}\left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) - \mathbb{E}\left[f(h(\breve{X}_0))\right]\right| \leq & \left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) - \frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i^{(2^n)}))\right| \\ & + \left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i^{(2^n)})) - \mathbb{E}\left[f(h(\breve{X}_0^{(2^n)}))\right]\right| \\ & + \left|\mathbb{E}\left[f(h(\breve{X}_0^{(2^n)}))\right] - \mathbb{E}\left[f(h(\breve{X}_0))\right]\right|\end{aligned}$$

Since the number of crossings in an interval of time of finite length if finite almost surely, there exist $N \in \mathbb{N}$ such that for any $n \geq N$

$$\left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) - \frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i^{(2^n)}))\right| = 0$$

and

$$\left|\mathbb{E}\left[f(h(\breve{X}_0^{(n)}))\right] - \mathbb{E}\left[f(h(\breve{X}_0))\right]\right| = 0$$

For any $n \geq N$, the last inequality is reduced to

$$\left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) - \mathbb{E}\left[f(h(\breve{X}_0))\right]\right| \leq \left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i^{(n)})) - \mathbb{E}\left[f(h(\breve{X}_0^{(n)}))\right]\right|$$

By ergodicity of the sequence $\{h(\breve{X}_i^{(n)})\}$, one has for any $n \geq 1$

$$\lim_{m\to+\infty}\left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i^{(n)})) - \mathbb{E}\left[f(h(\breve{X}_0^{(n)}))\right]\right| \overset{(a.s.)}{=} 0$$

whence

$$\lim_{m\to+\infty}\left|\frac{1}{m}\sum_{i=0}^{m-1} f(h(\breve{X}_i)) - \mathbb{E}\left[f(h(\breve{X}_0))\right]\right| \overset{(a.s.)}{=} 0$$

an so the statement of the Lemma is shown. □

5.3 Almost sure and L^1 convergence of $(\mathcal{N}^{\delta}_{(0,t)}(X)/t)$

In this section, we show the convergence of $(\mathcal{N}^{\delta}_{(0,t)}(X)/t)$ almost surely and in L^1, as t goes to ∞, when X has stationary and ergodic increments and has continuous sample path. The proof uses Derriennic's almost subadditive ergodic Theorem [74]. For simplicity, we note $N_{(0,t)} = \mathcal{N}^{\delta}_{(0,t)}(X)$.

5.3.1 Almost subadditivity of $\mathcal{N}^{\delta}_{(0,t)}(X)$

We show in Lemma 5.3.1 that the level crossing process is almost sub-additive, term introduced by Derriennic [74], which means that it exists a process $\{U_{(0,n)}\}_{n>0}$ such that

$$N_{(0,n)} \leq N_{(0,m)} + N_{(m,n)} + U_{(m,n)}$$

Where $\lim_{n\to+\infty} \dfrac{\mathbb{E}\left[U_{(0,n)}\right]}{n} \to 0.$

Lemma 5.3.1. *It exists a positive process* $\{U_{(0,n)}\}_{n\geq 1}$ *such that* $\dfrac{\mathbb{E}\left[U_{(0,n)}\right]}{n} \to 0$ *and*

$$N_{(0,n)} \leq N_{(0,m)} + N_{(m,n)} + U_{(m,n)} \tag{5.12}$$

Proof. For $m < n$, we set $Z_{(m,n)} = N_{(0,n)} - N_{(0,m)}$. Let $\tau + m$ the time of the crossing point after time m of the process X, so that

$$Z_{(m,n)} = \begin{cases} 1 + N_{(\tau+m,n)} & \text{if } \tau + m < n \\ 0 & \text{if } \tau + m \geq n \end{cases}$$

One has

$$N_{(0,n)} = N_{(0,m)} + Z_{(m,n)} = N_{(0,m)} + N_{(m,n)} + (Z_{(m,n)} - N_{(m,n)})$$

where

$$Z_{(m,n)} - N_{(m,n)} = \begin{cases} 1 + N_{(\tau+m,n)} - N_{(m,n)} \leq 1 + N_{(\tau+m,\tau+n)} - N_{(m,n)} & \text{if } \tau + m < n \\ -N_{(m,n)} \leq 0 & \text{if } \tau + m \geq n \end{cases}$$

Setting $U_{(m,n)} = \max\left(1, 1 + N_{(\tau+m,\tau+n)} - N_{(m,n)}\right)$, one gets.

$$N_{(0,n)} \leq N_{(0,m)} + N_{(m,n)} + U_{(m,n)}$$

Since

$$\mathbb{E}\left[N_{(\tau+m,\tau+n)}\right] = \mathbb{E}\left[\mathbb{E}\left[N_{(u+m,u+n)}|\tau = u\right]\right] = \mathbb{E}\left[\mathbb{E}\left[N_{(m,n)}|\tau = u\right]\right] = \mathbb{E}\left[N_{(m,n)}\right]$$

one has for all n and m positive

$$\mathbb{E}\left[U_{(m,n)}\right] = 1 < +\infty$$

and so taking $m = 0$, as $n \to +\infty$

$$\frac{\mathbb{E}\left[U_{(0,n)}\right]}{n} \to 0$$

□

5.3.2 Convergence of $(\mathcal{N}^{\delta}_{(0,t)}(X)/t)$

We need the following Derrienic's Theorem to show the convergence of $(\mathcal{N}^{\delta}_{(0,t)}(X)/t)$.

Theorem 5.3.2. *Almost Subadditive Ergodic Theorem [74]*
Let $N_{(m,n)}$ where $(m < n)$ be a sequence of random variables satisfying the following properties

(i) $N_{(0,n)} \leq N_{(0,m)} + N_{(m,n)} + U_{(m,n)}$ *such that* $\lim_{n\to+\infty} \frac{\mathbb{E}\left[U_{(0,n)}\right]}{n} = 0.$

(ii) $\{N_{(m,m+n)}\}_{m\geq 0}$ *are stationary for every* $n \geq 1$.

(iii) $\mathbb{E}\left[N_{(0,1)}\right] < +\infty$ *and for each* n, $\mathbb{E}\left[N_{(0,n)}\right] \geq Cn$, *where* $C > -\infty$

Then,

(i) $\lim_{n\to+\infty} \frac{\mathbb{E}\left[N_{(0,n)}\right]}{n} = \inf_{m>0} \frac{\mathbb{E}\left[N_{(0,m)}\right]}{m} = \gamma$

(ii) $\lim_{n\to+\infty} \frac{N_{(0,n)}}{n}$ *exists a.s. and in* L^1

Moreover if $\{N_{(m,m+n)}\}_{m\geq 0}$ *are ergodic then*

$$\lim_{n\to+\infty} \frac{N_{(0,n)}}{n} = \gamma \text{ in } L^1 \text{ and a.s.}$$

Proof. see [74] □

Theorem 5.3.3. *Let $X = \{X(t)\}_{t\geq 0}$ be a stochastic process so that its sample path is in C^{δ}. We assume X having stationary and ergodic increments. Moreover the density function of the $X(t)$ are continuous. Then*

$$\frac{\mathcal{N}^{\delta}_{(0,t)}(X)}{t} \to \gamma, \quad \text{a.s. and in } L^1 \quad \text{as } t \to +\infty \tag{5.13}$$

Where $\gamma = \lim_{t\to+\infty} \frac{\mathbb{E}\left[\mathcal{N}^{\delta}_{(0,t)}(X)\right]}{t}$.

Proof. Using the fact that

$$\frac{\lfloor t \rfloor}{t} \frac{N_{(0,\lfloor t \rfloor)}}{\lfloor t \rfloor} \leq \frac{N_{(0,t)}}{t} \leq \frac{\lfloor t \rfloor + 1}{t} \frac{N_{(0,\lfloor t \rfloor+1)}}{\lfloor t \rfloor + 1}$$

where $\frac{\lfloor t \rfloor}{t}$ and $\frac{\lfloor t \rfloor + 1}{t}$ goes to 1 as t goes to infinity, and since $\{N_{(m,m+n)}\}_{m\geq 0}$ satisfies assumptions of Theorem 5.3.2, we conclude the proof. □

Note that we can show when X has independent increments, that $\gamma = \frac{1}{\mathbb{E}\left[\tau_{\delta}\right]}$, where the random variable τ_{δ} represents the time interval between two consecutive crossing point of size δ. This can be shown using the properties of renewal processes stated below (see [75]).

Definition 5.3.4 (Renewal process). *Let $\tau_1, \tau_2, \ldots$ be independent identically distributed and positive stochastic variables, and set $T_n = \tau_1 + \cdots + \tau_n$. Then the process $\{N(t)\}_{t\geq 0}$, where*

$$N(t) = \max\{n : T_n \leq t\}$$

is called a renewal process.

Theorem 5.3.5. *Let $\{N(t)\}_{t\geq 0}$ be a renewal process with durations τ_n such that and $\mu = \mathbb{E}[\tau_1]$, then*

$$\frac{N(t)}{t} \overset{(a.s.)}{\to} \frac{1}{\mu}, \qquad as\ t \to +\infty$$

5.4 Crossing level of self-similar processes

5.4.1 Preliminaries

$\tilde{X}$ is H-sssi if there exists $H > 0$ so that for any $a > 0$,

$$\{\tilde{X}(at)\}_{t\geq 0} \overset{(fdd)}{=} \{a^H \tilde{X}(t)\}_{t\geq 0}$$

So one has for $t \in (0, T)$

$$\{\tilde{X}(t) \text{ is a crossing of level } a^{-H}\delta\mathbb{Z}\} = \{\tilde{X}(at) \text{ is a crossing of level } \delta\mathbb{Z}\}$$

This is equivalent to say that for any $t > 0$ and $a > 0$,

$$\mathcal{N}^{a^{-H}\delta}_{(0,t)}(\tilde{X}) \overset{(d)}{=} \mathcal{N}^{\delta}_{(0,at)}(\tilde{X}) \tag{5.14}$$

This later equality in distribution yields to the following Proposition.

Lemma 5.4.1. *Let $X = \{X(t)\}_{t\geq 0}$ a H-ss process with continuous sample path, so that $X(1)$ is a continuous random variable. Then for a fixed $t > 0$,*

$$\mathbb{P}(X \text{ has a upcrossing of } \delta \text{ in } (0,t)) \tag{5.15}$$

and

$$\mathbb{P}(X \text{ has a downcrossing of } \delta \text{ in } (0,t)) \tag{5.16}$$

are a continuous functions of δ.

Proof. It suffices to show it for (5.15). The case for (5.16) is then straight forward. We set

$$p\{X(u), \delta, (0,t)\} = \mathbb{P}(X(u) \text{ has a upcrossing of } \delta \text{ for } u \in (0,t))$$

From the equality in distribution 5.14, one has for any $\epsilon > 0$,

$$\begin{aligned} p\{X(u), \delta - \epsilon, (0,t)\} &= p\{X(u), \delta, (0, (1-\epsilon/\delta)^{-1/H} t)\} \\ &= p\{X(u), \delta, (0,t)\} + p\{X(u), \delta, [t, (1-\epsilon/\delta)^{-1/H} t)\} \end{aligned}$$

Hence

$$|p\{X(u), \delta, (0,t)\} - p\{X(u), \delta - \epsilon, (0,t)\}| = p\{X(u), \delta, [t, (1-\epsilon/\delta)^{-1/H} t)\}$$

and

$$\lim_{\epsilon \to 0} p\{X(u), \delta, [t, (1-\epsilon/\delta)^{-1/H} t)\} = \mathbb{P}\left(X(t) = \delta\right)$$

Since $X(1)$ is a continuous random variable

$$\mathbb{P}\left(X(t) = \delta\right) = \mathbb{P}\left(X(1) = t^{-H}\delta\right) = 0$$

This end the proof. □

From Lemma 5.4.1 and Lemma 5.1.1, we deduce that for H-sssi process $X = \{X(t)\}_{t \geq 0}$ that has continuous sample path and such that $X(1)$ is a continuous random variable, tangencies does not occur almost surely, and Theorem 5.3.3 can be applied.

Corollary 5.4.2. *Let X be H-sssi and satisfying conditions of Theorem 5.3.3. Then for $\delta > 0$*

$$\frac{\mathcal{N}^{\delta}_{(0,t)}(X)}{t} \to \frac{\gamma_1}{\delta^{1/H}}, \qquad a.s. \text{ and in } L^1 \qquad \text{as } t \to +\infty \tag{5.17}$$

where $\gamma_1 = \lim_{t \to +\infty} \dfrac{\mathbb{E}\left[\mathcal{N}^{1}_{(0,t)}(X)\right]}{t}$

Proof. Using Theorem 5.3.3, we have $\lim_{t \to +\infty} \dfrac{\mathcal{N}^{\delta}_{(0,t)}(X)}{t} = \delta^{-1/H}\gamma_1$ a.s and in L^1 and by equality (5.14),

$$\gamma_\delta = \lim_{t \to +\infty} \frac{\mathbb{E}\left[\mathcal{N}^{\delta}_{(0,t)}(X)\right]}{t} = \lim_{t \to +\infty} \frac{\mathbb{E}\left[\mathcal{N}^{1}_{(0,\delta^{-1/H}t)}(X)\right]}{t} = \delta^{-1/H}\gamma_1$$

This ends the proof. □

Corollary 5.4.3. *Let X be H-sssi and satisfying conditions of Theorem 5.3.3. Then for $t > 0$*

$$\delta^{1/H}\mathcal{N}^{\delta}_{(0,t)}(X) \overset{(\mathcal{P})}{\to} t\gamma_1, \qquad \text{as } \delta \to 0 \tag{5.18}$$

Proof. By equality (5.14), there exist a bounded function f so that one has

$$\left|\mathbb{E}\left[f\left(\delta^{1/H}\mathcal{N}^{\delta}_{(0,t)}(X)\right)\right] - f\left(t\gamma_1\right)\right| = \left|\mathbb{E}\left[f\left(\delta^{1/H}\mathcal{N}^{1}_{(0,\delta^{-1/H}t)}(X)\right)\right] - f\left(t\gamma_1\right)\right|$$

Theorem 5.3.3 yields to $\lim_{\delta\to 0}\left|\mathbb{E}\left[f\left(\delta^{1/H}\mathcal{N}^1_{(0,\delta^{-1/H}t)}(X)\right)\right]-f\left(t\gamma_1\right)\right|=0$ and so

$$\lim_{\delta\to 0}\delta^{1/H}\mathcal{N}^\delta_{(0,t)}(X)\overset{(d)}{=}t\gamma_1$$

As $t\gamma_1$ is a constant, we deduce (5.18). □

5.4.2 Crossing points as an estimate of H

From Corollary 5.4.2, one has $\gamma_\delta=\delta^{-1/H}\gamma_1$. An estimate $\hat{\gamma}_\delta$ of γ_δ is given by $\lim_{t\to+\infty}\frac{\mathcal{N}^\delta_{(0,t)}(X)}{t}$ and so an estimate $\hat{H}$ of H is given by the slope α when using a linear regression of the following equality

$$\log(\hat{\gamma}_\delta)=-\alpha\log(\delta)+\log(\gamma_1) \tag{5.19}$$

An estimate of H is $\hat{H}=-1/\hat{\alpha}$.

We apply the estimator on a Brownian motion simulated on a equally space grid [28], from which we estimate the self-similarity index ($H=1/2$ for the Brownian motion). We simulate 10000 observed crossings of size 1. A sample of the path is represented in Figure 5.3. The linear regression of the equation (5.19) which estimate the slope $-\alpha$ is represented in Figure 5.4. The setting is as follow: $\delta\in\{1,2,3,4,5\}$.

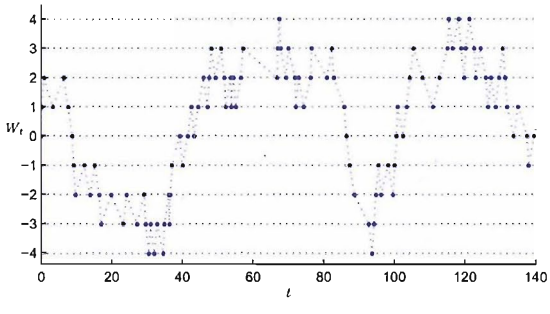

Figure 5.3: Brownian Motion.

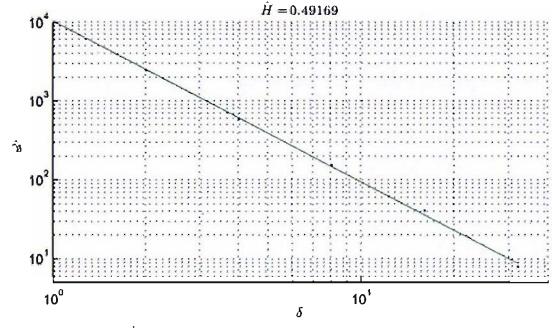

Figure 5.4: H estimation using a linear regression.

We compare the EBP estimator and our estimator performances. We simulate 100 Brownian motion giving each 10000 and 100000 observed crossings of size 1. Table 5.1 presents the mean of the estimate H and its standard deviation. The EBP estimator performs slightly better than the level crossing estimator.

number of observation	Level Crossings	EBP
10000	0.5005 (0.0110)	0.5004 (0.0079)
100000	0.5003 (0.0052)	0.5001 (0.0036)

Table 5.1: Self-similarity index estimation of a Brownian motion simulated in a fixed size grid using the level Crossings and the EBP methods.

5.4.3 H estimate from fractional Brownian motion

We compare the level crossings estimator and the EBP, on simulated fractional Brownian motion. First we note that there is no way to simulate the fractional Brownian motion on a regular grid space. The number of crossings computed from simulated fBm on a regular time space are an approximation of the number of true crossings. To show how this approximation is affected by the roughness of the process, we simulate samples of fractional Brownian motion $\{B^H(t)\}_{t\in[0,T]}$, where $T = 100$ and $\mathbb{V}ar\left(B^H(1)\right) = 1$. We simulate $2^n T$ observations of this process, where $n = \{0, 1, \ldots, 14\}$. We compute the number of crossings for different size of δ. The averaged number of crossings for 100 sample of fBm are illustrate in Figure 5.5 as a function of the number of observation. The dashed line represent the confidence interval taking twice the standard deviation on both side of the mean.

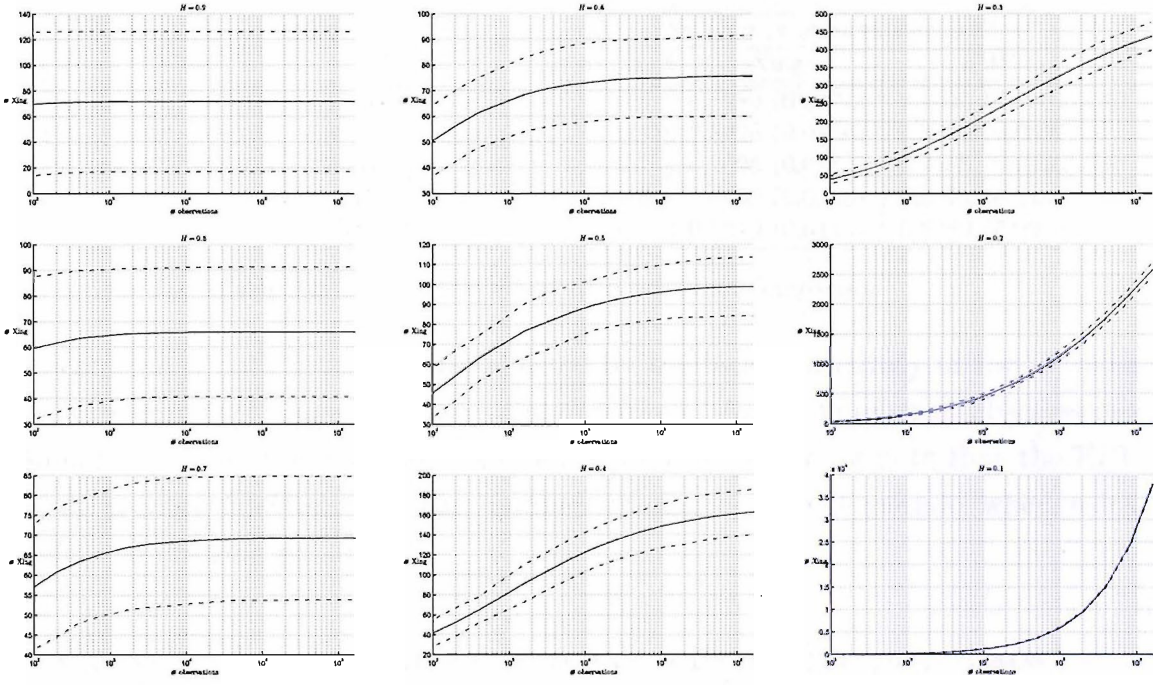

Figure 5.5: The average number of crossings for 100 sample of fBm as a function of the number of observations.

Setting $\delta_n = \sqrt{\mathbb{V}ar\left(B^H(1/2^n)\right)}$, the coresponding level of size δ_n for a sample with $2^n T$ observations when the crossings are taken at size $\delta = 1$, is given by $\delta = 2^{\text{level}}\delta_n$, whence

$$\text{level} = \log_2\left(\frac{\delta}{\sqrt{\mathbb{V}ar\left(B^H(1/2^n)\right)}}\right) = nH$$

We now compare the level crossing estimator and the EBP estimator. Due to the limited computational performance[3], we limit our study to $H \in [0.5 : 0.9]$. The results are given in Tables 5.2 for different number of observations. The mean and twice the

[3]The computer used do not alow us to simulate a huge number of observation per sample, and so for $H \leq 0.4$ the estimator will perform badly.

standard deviation of the estimated self-similar index H for 100 samples are given.

Number of observations 1000		
H	Xing	EBP
0.9	0.8445 (0.1047)	0.8513 (0.1264)
0.8	0.7860 (0.0951)	0.7894 (0.1337)
0.7	0.7075 (0.1040)	0.7110 (0.1440)
0.6	0.6346 (0.1186)	0.6620 (0.2236)
0.5	0.5498 (0.1131)	0.5651 (0.1769)

Number of observations 50000		
H	Xing	EBP
0.9	0.8833 (0.0515)	0.8910 (0.0391)
0.8	0.8003 (0.0340)	0.8107 (0.0216)
0.7	0.7068 (0.0329)	0.7180 (0.0266)
0.6	0.6160 (0.0381)	0.6271 (0.0310)
0.5	0.5191 (0.0344)	0.5340 (0.0390)

Number of observations 5000		
H	Xing	EBP
0.9	0.8766 (0.0749)	0.8798 (0.0720)
0.8	0.7943 (0.0645)	0.8040 (0.0616)
0.7	0.7097 (0.0769)	0.7196 (0.0862)
0.6	0.6285 (0.0655)	0.6347 (0.0987)
0.5	0.5385 (0.0790)	0.5509 (0.0980)

Number of observations 100000		
H	Xing	EBP
0.9	0.8958 (0.0479)	0.9009 (0.0360)
0.8	0.8020 (0.0252)	0.8109 (0.0182)
0.7	0.7074 (0.0210)	0.7175 (0.0190)
0.6	0.6145 (0.0280)	0.6289 (0.0240)
0.5	0.5194 (0.0269)	0.5364 (0.0336)

Number of observations 10000		
H	Xing	EBP
0.9	0.8791 (0.0706)	0.8824 (0.0570)
0.8	0.7962 (0.0642)	0.8013 (0.0563)
0.7	0.7153 (0.0611)	0.7188 (0.0522)
0.6	0.6192 (0.0706)	0.6335 (0.0571)
0.5	0.5297 (0.0716)	0.5335 (0.0877)

Number of observations 500000		
H	Xing	EBP
0.9	0.8980 (0.0374)	0.9020 (0.0301)
0.8	0.8026 (0.0166)	0.8105 (0.0097)
0.7	0.7095 (0.0114)	0.7185 (0.0130)
0.6	0.6152 (0.0106)	0.6271 (0.0195)
0.5	0.5210 (0.0114)	0.5382 (0.0241)

Table 5.2: Level crossing and EBP estimator comparison

We conclude that both estimator are affected by the lack of accuracy of the crossing points, for sample with small number of observations. As the number of observations in a sample increase, the estimators are less biased. However we note that the EBP estimator is more affected than the number of crossing points method when using approximated crossing points.

5.5 Crossing level of time-changed self-similar processes

5.5.1 Time change estimator

Let X be of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-sssi and satisfies conditions of Theorem 5.3.3, and θ a non-decreasing function, so that $\theta(0) = 0$.

Proposition 5.5.1. *Let the process X be of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is $H-$sssi and satisfies conditions of Theorem 5.3.3, and θ is a continuous non-decreasing function. Then*

$$\delta^{1/H}\mathcal{N}^{\delta}_{(t,t+\delta t)}(X) \overset{(\mathcal{P})}{\to} (\theta(t+\delta t) - \theta(t))\tilde{\gamma}_1, \qquad as\ \delta \to 0$$

where $\tilde{\gamma}_1 = \lim_{t\to+\infty} \dfrac{\mathbb{E}\left[\mathcal{N}^1_{(0,t)}(\tilde{X})\right]}{t}$.

Proof. It is obvious to see that $\mathcal{N}^{\delta}_{(t,t+\delta t)}(X) = \mathcal{N}^{\delta}_{(\theta(t),\theta(t+\delta t))}(\tilde{X}) \overset{(d)}{=} \mathcal{N}^{\delta}_{(0,\delta^{-1/H}(\theta(t+\delta t)-\theta(t)))}(\tilde{X})$, by Corollary 5.4.3, one has

$$\begin{aligned}\lim_{\delta\to 0}\delta^{1/H}\mathcal{N}^{\delta}_{(t,t+\delta t)}(X) &\overset{(d)}{=} \lim_{\delta\to 0}\delta^{1/H}\mathcal{N}^{\delta}_{(0,\delta^{-1/H}(\theta(t+\delta t)-\theta(t)))}(\tilde{X}) \\ &\overset{(d)}{=} (\theta(t+\delta t)-\theta(t))\lim_{\delta\to 0}\frac{\delta^{1/H}\mathcal{N}^{1}_{(0,\delta^{-1/H}(\theta(t+\delta t)-\theta(t)))}(\tilde{X})}{(\theta(t+\delta t)-\theta(t))} \\ &\overset{(d)}{=} (\theta(t+\delta t)-\theta(t))\tilde{\gamma}_1\end{aligned}$$

$(\theta(t+\delta t)-\theta(t))\tilde{\gamma}_1$ is a constant, whence the proof. □

Assuming the function θ as a simple function of the form

$$\theta(t) = \sum_{j\geq 0}(t-j\delta t)\frac{\theta((j+1)\delta t)-\theta(j\delta t)}{\delta t}I_{[j\delta,(j+1)\delta t)}(t)$$

By Proposition 5.5.1, one can estimate the θ by estimating $\lambda_{t,\delta t} = \dfrac{\theta(t+\delta t)-\theta(t)}{\delta t}$. An estimate of $\lambda_{t,\delta t}$ is given by

$$\hat{\lambda}_{t,\delta t} = C_\delta \frac{\mathcal{N}^{\delta}_{(t,t+\delta t)}(X)}{\delta t}$$

where $C_\delta = \dfrac{\delta^{1/H}}{\gamma_1}$. One has $\hat{\lambda}_{t,\delta t} \overset{(\mathcal{P})}{\to} \lambda_{t,\delta t}$ as $\delta \to 0$. In practice, assuming we have observation $\{X(t)\}_{t\in[0,T]}$, one set $\theta(0)=0$ and $\theta(T)=T$, and so C_δ can be determinate using the fact that

$$\sum_{j=0}^{T/\delta t-1}\lambda_{j\delta t,\delta t} = \frac{T}{\delta t}$$

where δt is chosen so that $T/\delta t$ is an integer. Note C_δ is a norming constant and so the index of self-similarity H of the process X does not need to be estimate as H vanishes in the norming constant. We apply next our time change estimator to time-changed Brownian motion and fractional Brownian motion. The theoretical time function is identical to the one used in Chapter 4.

5.5.2 Application to time-changed Brownian motion

We test the performance of the time change estimator on time-changed Brownian motion $B(\theta(t))$. There is three ways of simulating a time-changed Brownian motion as described in the previous Chapters: regularly space in time that we note B_1 (See Section 3.2.3), irregularly spaced in time that we note B_2 (this uses the time function θ in abscise. See Section 3.2.4) and regularly grid spaced that we note B_3, this last present the true crossing points of the process, we refer to [28]. For each of these ways, we simulate 100 samples of 100000 observations, from which we estimate θ'. A box plot is given in Figure 5.6.

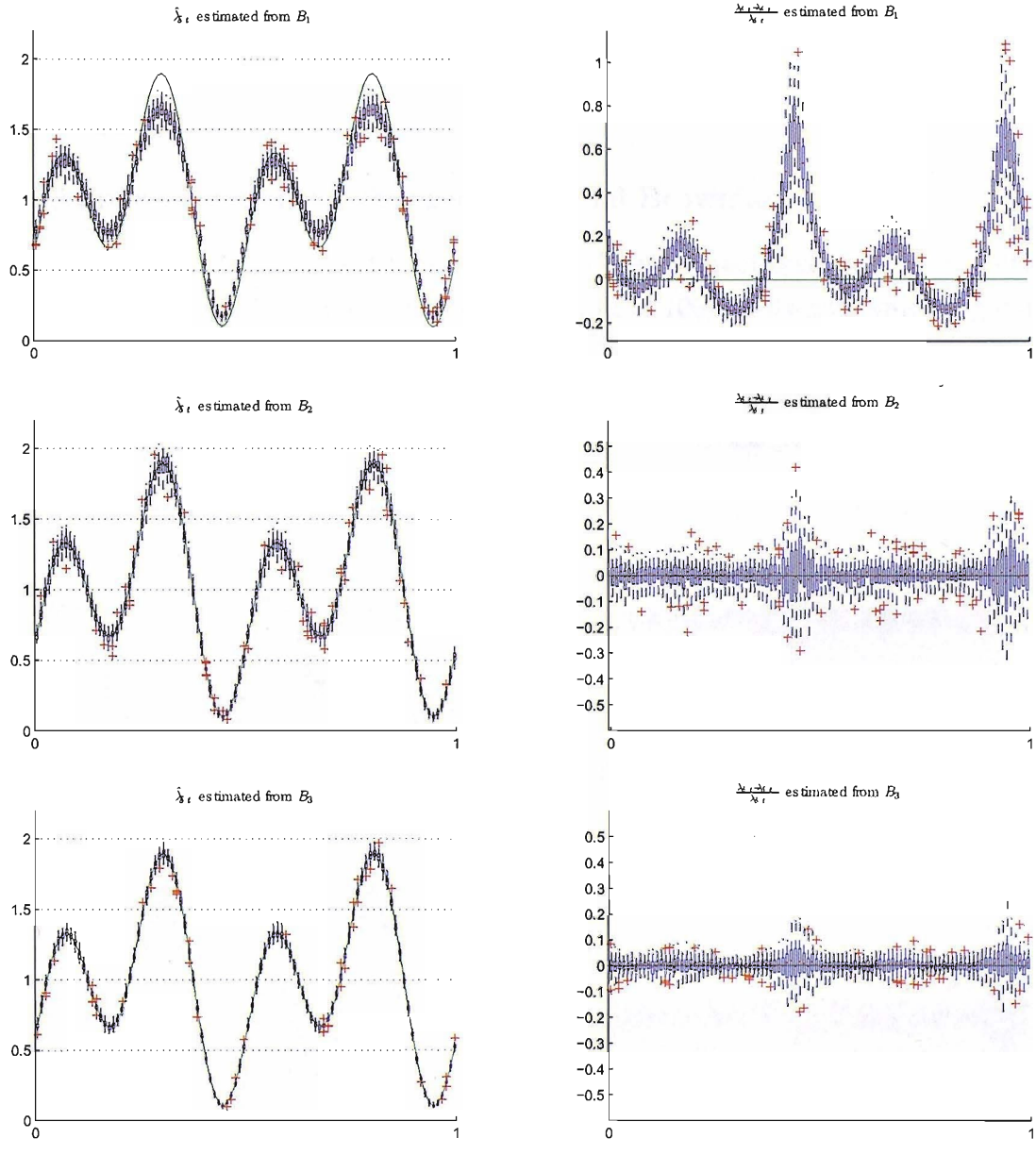

Figure 5.6: Time change function estimate.

Note that the estimated time change from B_1 is biased. This is due to the fact that there is in proportion more missing crossing points during period when θ' has big value than when θ' has small value. This proportion seem to equilibrate and becomes similar for any value of θ' in the case of B_2. From B_3, the estimator performs better when using true crossing points.
Note that for high volatile period, the estimator has a better estimate than low volatile period. This is due to the fact that their is more crossing points during high volatile period and so the estimate of the time change function at that period has better convergence.

5.5.3 Application to time-changed fractional Brownian motion

We test here the time change estimator on time-changed fractional Brownian motion $B_H(\theta(t))$ for $H = \{0.3, 0.7\}$. We simulate 100 sample of 100000 observations irregularly spaced in time as for B_2 in the previous section. The box plot is illustrate in Figure 5.7.

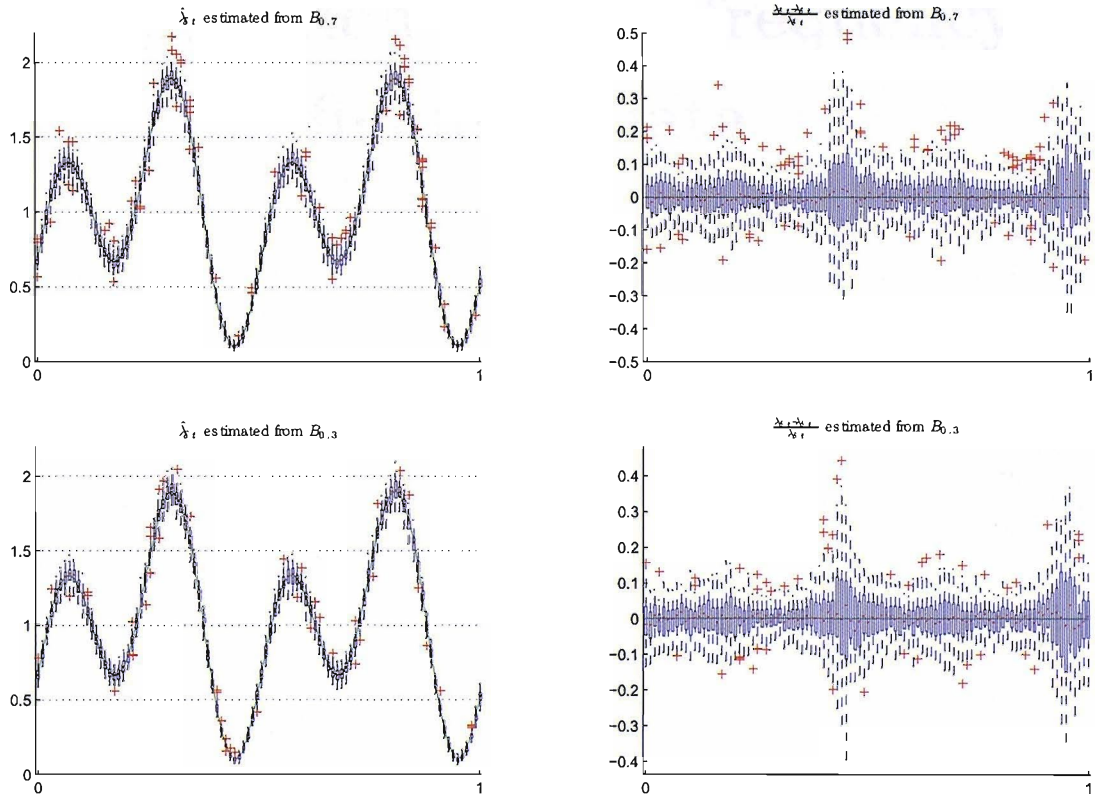

Figure 5.7: Time change estimate from fBm with $H = 0.7$ (top panels) and $H = 0.3$ (bottom panels).

Note that for high volatile period, the estimator has a better estimate than low volatile period for both case. we note a smaller error for $H = 0.3$ compare to $H = 0.7$ when θ' is high and bigger when θ is low.

Part III

Application to high-frequency financial data

Chapter 6

High frequency financial data : Introduction and data cleaning

The analysis of high-frequency financial data is becoming an issue in understanding the market behaviour. Millions of quotes are traded everyday and the advances in technology such as data storage and electronic trading systems implementation have made high-frequency financial data accessible to academics and commercial users. High frequency data (HFD) in finance refers to prices that are recorded several times a day. More prices we have per day, the higher the frequency of the observations. The prices are generally given with one minute interval. However, complete datasets contains all the quotes for each transaction made during the day and therefore the dataset has tick-by-tick data, also referred by Engle [76] as ultra-high-frequency data (UHFD). For UHFD, the time stamps are irregularly spaced in time and may contain several transactions for a same time stamp and no quote during week-end time. Various studies on the market microstructure exist in the literature, we refer to Andersen [77], Goodhart and O'Hara [78], and Wood [79]. Also in some books such as Gouriroux and Jasiak [80], Tsay [81] and Taylor [82], a chapter on high-frequency data is presented. A complete survey on the subject of high frequency finance is presented in the book of Dacarogna et al. [83]. High-frequency financial data present an advantage compared to daily data. The additional intraday prices gives more details on how the price reacts to information and so a better examination of the source and the volatility of the return can be made. However, it presents also a great challenge to econometric modeling and statistical analysis and those for various reasons due to the characteristics of high-frequency financial data. For example: irregular spaced time observations, difficulty in deciding what is the efficient price for a given time, existence of daily and weekly periodic patterns, erroneous quotes. We introduce the main characteristic of High frequency financial data and we describe the difficulties we may cross once one studies them. We then investigate the issues on cleaning high frequency data. We present a simple way to conduct basic cleaning, after describing the different data type error that datasets may contain. We then apply our data cleaner to some foreign exchange rate

indexes, in particular: AUD/USD, GBP/USD, JPY/USD and EUR/USD rate indexes.

6.1 UHFD: Characteristics and problems

Unlike daily and weekly data, ultra-high frequency data are characterised by its complex microstructure, such as the huge number of observations, the irregularly spaced time stamp, the price types (discreteness) and the temporal dependence of the asset price return. These microstructure characteristics complicate the UHFD analysis. Adding to the market microstructure, the market activity effects also introduce some complication on data analysis. The price volatility exhibits a periodic pattern. Some of these characteristics can be seen in Figure 6.1. In this Section, we explain the origins and the problems that microstructure elements, as described previously, may cause to financial data analysis.

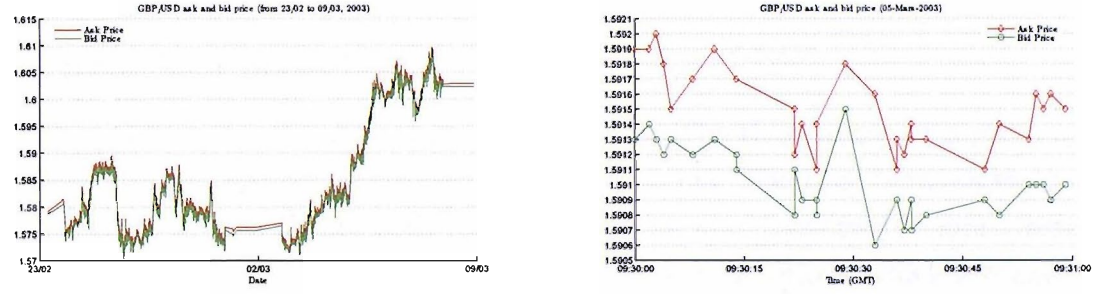

Figure 6.1: Quotes of the GBP/USD rate index of two weeks period(left panel) and one minute period(right panel)

6.1.1 The number of observations

Assets are quoted several times a day by market makers and so they generate a large dataset of quotes. A dataset of high frequency financial data may contain tens of thousands of transactions or quotes per day. Figure 6.2 represents the half-hour averaged number of quotes during the day. The averaged number of quotes per half-hour may exceed 700. Note the high number of quotes between 0h and 6h in GMT for the JPY/USD rate index compared to the GBP/USD. This is due to the trading hours of the Asian markets, and so the transactions are higher due to high trading activity for the JPY/USD currency. Recording a huge amount of data in a short period of time may induce some erroneous quotes. The dataset may then contain erroneous observations such as outliers, data gaps and even disordered sequences. By looking at the data, one may see the obvious bad quotes, however it becomes problematic when the dataset contains millions of quotes, such as the GBP/USD rate index dataset which contains around 4 million and JPY/USD rate index dataset with around 6 million. Indeed, the bigger the size of the dataset, the more bad quotes we expect to have. Thus, the risk

of detecting false a outlier or true price as an outlier, may become important.

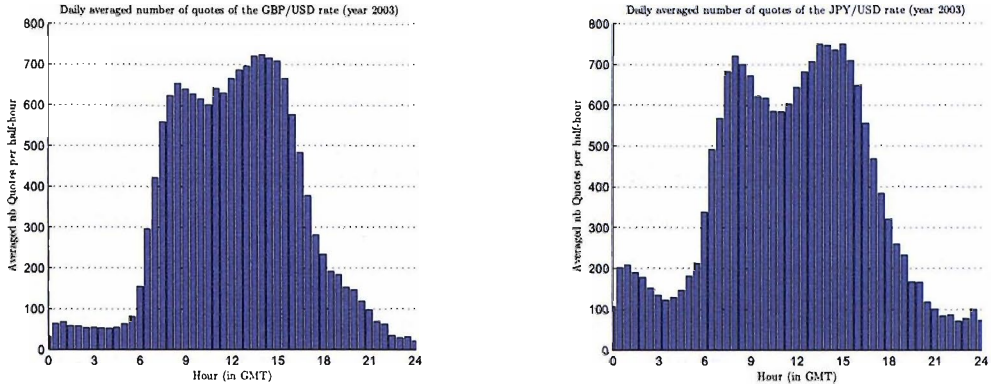

Figure 6.2: Half-hour averaged number of quotes during the day for the GBP/USD rate index (left panel (a)) and the JPY/USD rate index (right panel (b))

6.1.2 Irregularly Spaced time

Quotes or prices are given at irregular spaced time. The time stamp of a given quote corresponds to the moment where the transaction occurs. The time between two transactions may differ (see Figure 6.1). For example, we may have two or more different quotes having the same time stamp. This is due to transactions which are executed simultaneously (see Figure 6.1 (b) at 09:30:25). Also two successive quotes may have an interval, which correspond to more than a day, during the week-end for example (see Figure 6.1 (a) week end of the 2^{nd} of Mars). A cause for irregular spaced time quotes is the market active. This last will be discussed later. It is even harder to model such a time series since the interval time between two ticks are considered random. We refer to N.Hautsh [84] for a more detailed work on Modelling irregularly spaced financial data. To make data regularly spaced in time, one of the techniques is to sample the data to a certain frequency, using for example a linear interpolation of the data. The sampling frequencies may vary depending on the quantity we want to study. Generally, data are sampled at one minute intervals. Some studies are made with 5 minutes sampled data, for example Andersen and Bollerslev in [85]. In sampling data, they are two inconveniences: first, losing information and so breaking the high frequency system, and second, the statistical characteristic of the high frequency data, such as creating dependence of the time series increments may change. For example, during low activity market, the number of quotes are low and so a high frequency interpolation in that case may create a monotonic series in the data. This last may cause a strong dependence in the time series, and introduce a bias in the volatility estimate, see Oya [86]. However, if we sample at low frequency, we then loose the extra information of the market microstructure and the statistical properties may also differ from the UHFD.

Some works exist on the optimal[1] value of the sampling, see Aït-Sahalia [87] and Bandi and Russell [88].

6.1.3 Price types

Price formation of a given asset is based on its demand and supply. A more detailed explanation on price formation are presented by Madhavan [89] and O'Hara [90]. Prices are quoted in fixed increments or ticks. In particular this increments is a multiple of a minimum price change allowance, so prices fall on a pre-specified set of values; which implies the discreteness of the price. See for example, Figure 6.1 (b), the GBP/USD index rate has a minimum size increment, also called Pip[2], of 0.0001.
The price data are often available as quotes, set up by the market makers[3]. A market maker is a firm that stands ready and able to buy and/or sell a particular share on a regular and continuous basis at a publicly quoted price. These firms display a price at which they are willing to buy (bid price) and a price at which they are willing to sell (ask price) for specific securities, and they are willing to do so from their own accounts. For example, assuming you want to sell 100 shares of Alcatel for example, you will need to find a willing buyer, who is able to buy from you 100 shares at the moment you place your order, and here comes market makers. They will buy from you these 100 shares even if they do not have a seller lined up. Therefore market makers are important to keep the liquidity [4] and the efficiency [5] of the market. Since each market maker can either buy or sell a stock at any given time, market makers make profit on bid-ask price difference called the ask-bid spread.
Generally for UHFD, the dataset contains the date and the time of the transaction, and the quotes' bid and ask prices see Table 6.1. High frequency financial dataset when the data are given with a regular time interval, contains the date and the time of the transaction, followed by the open, the low, the high and the close price.

DATE	TIME	BID	ASK
03/07/2003	9:44:58	1.6613	1.6618
03/07/2003	9:45:00	1.6613	1.6618
03/07/2003	9:45:00	1.6612	1.6618
03/07/2003	9:45:03	1.6612	1.6618
03/07/2003	9:45:06	1.6614	1.6616

Table 6.1: Table representing a dataset of the UHFD GBP/USD quotes.

[1] In Chapter 8, we present a way to find the optimum value of the sampling, assuming the financial data are H-sssi

[2] Spread Terminology used in currency market to represent the smallest incremental move an exchange rate can make.

[3] Note the market is said to be quotes-driven, when quotes are set up by market makers.

[4] Converting securities into cash without affecting the asset's price. Liquidity is characterized by a high level of trading activity. One can rapidly buy and/or sell an asset. This is due to the large amount of market makers and so the availability of buyer and seller at anytime.

[5] How easily and inexpensively transaction can occur.

DATE	TIME	OPEN	HIGH	LOW	CLOSE
03/07/2003	11:44:00	1.6615	1.6615	1.6611	1.6611
03/07/2003	11:45:00	1.6612	1.6615	1.6612	1.6614
03/07/2003	11:46:00	1.6613	1.6613	1.6612	1.6612
03/07/2003	11:47:00	1.6611	1.6614	1.6611	1.6611
03/07/2003	11:48:00	1.6612	1.6613	1.6611	1.6613

Table 6.2: Table representing a dataset of GBP/USD rate index quotes, given at 1 minute interval

For UHFD, the bid and ask price respectively noted P_t^{bid} and P_t^{ask} do not reflect the true price P_t^* for a given asset. One has always [6]

$$P_t^{bid} \leq P_t^* \leq P_t^{ask}$$

One may be surprised to see through this inequality that the price P_t may be equal to the bid or the ask price. Indeed, the price may also exhibit a bid-ask bounce, which is the process of the transaction price moving between the ask and bid prices for a period of time. As the data do not contain the transaction price, we tend to use the mid point price

$$P_t = \frac{P_t^{ask} + P_t^{bid}}{2}$$

6.1.4 Periodic pattern in the price volatility

Several studies show that high-frequency financial data exhibit daily and weekly periodic patterns, also known as intraday and intraweek seasonality, in the volatility (see Figure 6.3). We refer to Dacorogna et al. [83] for more detailed discussion on daily and weekly seasonality of foreign exchange and Owain ap Gwilym [91] for a more general framework. The origin of this seasonality is due to the market activities, which depend on the time of day. Indeed, this activity is higher when trade opens and just before it closes, lower during lunch time and completely absent at nights and weekends, as noted by Owain ap Gwilym [91]; who also showed for most stock markets, that volatility exhibits a U-shaped pattern over the course of the day.
However for Foreign Exchange trading, the activity depends on the trading hours of the main markets, situated geographically at different places. When the trading hours of two main markets overlap, the activity tends to be higher. Also some markets have higher activity than others. We refer to Dacorogna et al. [3] for more details.
The heteroskedasticity [7] of the times series create strong temporal dependencies on the path of the process. One way to proceed is to remove week-ends and so that we create a new trading time. The trading time omitting Saturday and Sundays, the two inactive days of the week, is called the business time. It allows the time series to be

[6] Of course if the quote is not considered as an erroneous observation

[7] A sequence of random variables is heteroskedastic, if the random variables in the sequence may have different variances.

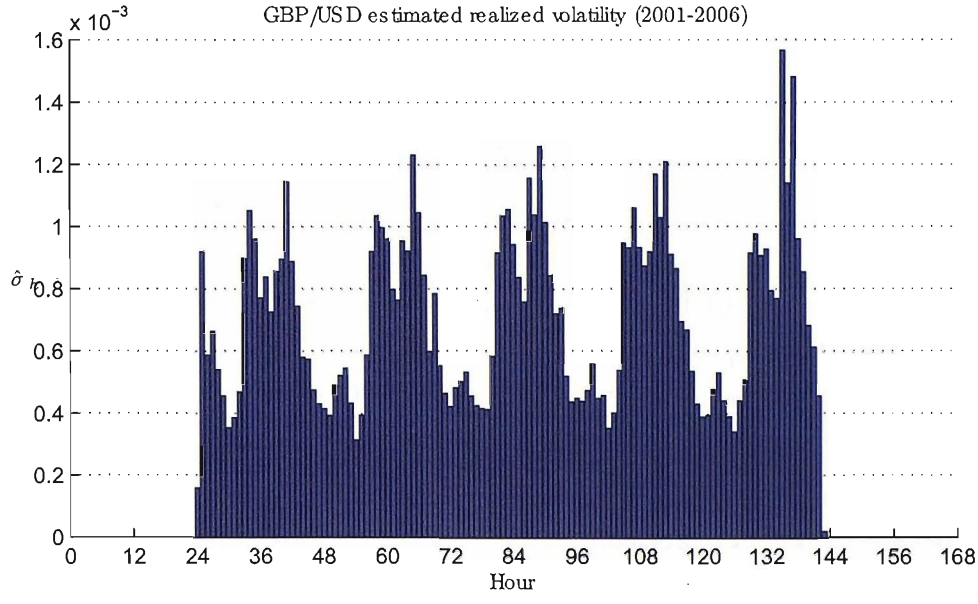

Figure 6.3: The averaged realized volatility of the GBP/USD rate index from 2001 to 2006

less affected by the weekly seasonality. However, the daily periodicity still remains in that case. Dacorogna et al. [3] present a new time-clock called the θ-time [8]. This new scaling time ensures that "the behavior of the market at any particular moment is interpreted in the context of the number of participants present" [9]. Even if the number of participants present is the same at any particular moment, the activity in the sense of having constant volatility over the trading time may remain different. In fact mathematically, price under θ-time has constant volatility over the time. θ-time is obtained by stretching the calendar time during high activity period and contracting it during low activity period. We refer also to Ghysels et al. [92] and Zhou [93] for more discussion on time deformation and deseasonalisation.
Note also the number of quotes per unit of time exhibits a seasonal pattern. In Figure 6.2, we see the daily seasonal pattern of number of quotes per half-hour for the GBP/USD and JPY/USD rate indexes.

6.2 Cleaning high-frequency financial data

Large data base record of ultra-high frequency data has often some aberrant observations. The data cleaning is then primordial in the high-frequency data analysis. Uncleaned data may introduce some error in their statistical analysis results. In this Section, we try to understand the meaning of an erroneous observation in UHFD and present the different data type errors that datasets may contain, as well as their origins. We finish by describing the common cleaning process presented in the literature.

[8] In the next chapter, we present methods improving the θ-time estimation
[9] Sentence taken from the Olsen group website: http://www.olsen.ch/research/approach.html.

6.2.1 What is an aberrant observation?

It is not really obvious how to detect whether or not an observation is erroneous. For example, assuming the following sequence of prices

$$\{1.56, 1.67, 1.97, 1.45, 289.78, 1.89\}$$

it is obvious to say that 289.78 is a bad tick, should we really ask ourselves if we have to remove it or not? The answer is no doubt, we remove the quote. In reality that may occur on one or two ticks over millions of ticks, but what about having 2.89 instead of 289.78, should we consider it as an erroneous price? This illustrate the case where the filtering becomes complicated. In fact the difficulty of cleaning high-frequency data is the inability to define an erroneous observation, as mentioned in [94]. Müller introduces a definition of the so called "erroneous observation" in [83]:

"*A data error is present if a piece of quoted data does not conform to the real situation of the market.*"

This may provide an answer to the question presented in the title as long as one understand what the real situation of the market conformity is, which may depend on personal viewpoints. Indeed, what may be considered as a bad tick for some can in fact be a real transaction, and what may appear as a good tick may be a spurious observation. Cleaning data is risky in the sense that we may not remove only spurious ticks but also true observations. The frequency of bad quotes may then vary from one filter to another. For examples, Lundin et al. has reported in [95] that their data filters identified 2% to 3% of all data points as false outliers, whereas the expected frequency of bad quotes in a high frequency dataset is estimated between 0.11% and 0.81% by Dacorogna et al. [96] for foreign exchange rates.

6.2.2 Data type error and its sources

Several papers in the literature describe the data type error and their origins, we refer to Müller et al. [97], Falkenberry [94] and Morgan [98]. There are two sources of error: human error, which is produced either unintentionally such as typing error and may cause decimal errors, or missing the fractional portion of the price value; or intentional error such as producing dummy quotes just for technical test and this may cause repeated or monotonic series of prices. This last error can occur especially just before the markets opening time, to test if the connection for data transmission is working for example. The second type of error is the result of from computer system failures such as time delay in transmission or computer system breakdown, which may create missing prices in the dataset. Another explanation on the origins of bad ticks is the simultaneous activity of the markets. Each local market have their own trading habits, for example the maximum bid-ask spread allowed are different from one market

to another and that is even if the market activities overlap. In fact this overlapped trading creates simultaneous trading with different quotes. As in [98], these sources produce errors that can be summarised and classified into 4 categories:

- No values have been input (Missing quotes)
- Non plausible values have been input (Negative prices, Quotes during inactive period, bid price > ask price)
- Inconsistent values have been input (Repeated quotes, Monotonic series)
- Incorrect values have been input (outliers, decimal error)

In the next section, we introduce the common high frequency financial data filtering process that has been examinated in the literature.

6.2.3 Common cleaning process

A common filtering process can be extracted from the filters presented in the literature. The extracted common cleaning process comes from filters presented by Dacorogna et al. in [83], Falkenberry [94], Green et al. [99] and Morgan [98]. In all filters, there are two main processes:

- Identifying potential bad quotes
- Correcting or removing bad quotes
- Filling gaps of missing quotes

The process starts first by identifying whether or not the quote is considered as a potential erroneous observation. The dataset of dirty data gets into the process of identifying potential bad quotes. Each filter has different algorithms [10] on identifying bad quotes. However, a common filter can be extracted which corresponds to detecting non-possible values. Quotes are marked if it seems to be a potential erroneous observation.
Once the Quotes is detected as a "bad Quote", we check to determine if we are able to correct the price values of the bid and/or the ask. If so, then the price is corrected. If a correction cannot be performed, then we remove the quote from the dataset. However, there are quotes that just need to be removed. These quotes are generally during inactive period, such as holidays and week end. For data that are removed during trading times, they will be considered as missing values [11].
Gaps created by missing values are usually filled by values provided by a model which fit the data. We shall not describe this last here, instead we refer to [98] for more discussion on using a model to predict missing values.

[10] see the corresponding papers to find out more. Ours will be presented in the next section
[11] Note that in our filter we will not consider this type of error. Our filter will just remove ticks which are considered bad.

For more detail on each particular high-frequency financial data filters presented in the literature, we refer to their appropriate article [83], [94] and [99].

6.3 Our basic data vacuum cleaner

For our study, we do not need a complicated filter, as described by Dacorogna et al in [83]. All that is needed, is to just remove non-plausible quotes and outliers, which may affect our time-changed estimators [12]. We present the different features of our data cleaner.

1. Check if data are sorted
2. Check if the bid and ask prices are strictly positive
3. Check if ask price greater than the bid price
4. Remove repeated lines in the dataset
5. Remove quotes during inactive periods
6. Detecting spikes

6.3.1 Filtering non plausible quotes

6.3.1.1 Date and time sequence

We check if the date and time are sorted. Taking the stamp time T_i of the i^{th} quote and setting n the number of quotes in the dataset, one checks for all $i \in \{1, ..., n-1\}$,

$$T_i \leq T_{i+1}$$

If the data are not sorted, we just sort the dataset.

6.3.1.2 Negative bid and ask price

The bid and ask price of the i^{th} quote respectively noted P_i^{bid} and P_i^{ask} must be strictly positive. We set

$$N_i^1 = (P_i^{bid} \leq 0) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases} \quad \text{and } N_i^2 = (P_i^{ask} \leq 0) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

[12] This will be described in the next chapter

6.3.1.3 Spread consistency

Here, we make sure that the bid price are strictly smaller than the ask price, which is equivalent to say that the bid/ask spread are strictly positive. We set $S_i = P_i^{ask} - P_i^{bid}$ the bid/ask spread of the i^{th} quote and we define

$$N_i^3 = (S_i < 0) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases} \quad \text{and } N_i^4 = (S_i = 0) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

6.3.1.4 Repeated quotes

We set $L_Q^i = \{T_i, P_i^{bid}, P_i^{ask}\}$ the i^th quotes line includes the time stamp. Dataset may contain successive repeated quotes with the same time stamp. In that case, we have $L_Q^{i+1} = L_Q^i$. We define

$$N_i^5 = (L_Q^i = L_Q^{i-1}) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

6.3.2 Filtering quotes in inactive period

During inactive periods, such as holidays and week-ends, quotes are removed. Using previous notation, we set

$$N_i^6 = (T_i \in \{ \text{ inactive periods } \}) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

6.3.3 Filtering spikes

We use a median filter of the price rate path to detect spikes. Indeed, unlike the moving average filter, the median filter is robust to outliers and jumps. For example, assuming X_i is the price or rate of the i^{th} quote, we set

$$\begin{cases} X_{i,k}^{med} = \text{median}(\{X_{i-k}, \ldots, X_{i+k}\}) \\ X_{i,k}^{m} = \dfrac{1}{2k+1} \displaystyle\sum_{j=-k}^{j=k} X_{i-k} \end{cases} \quad \text{and} \quad \begin{cases} Z_{i,k}^{med} = |X_i - X_{i,k}^{med}| \\ Z_{i,k}^{m} = |X_i - X_{i,k}^{m}| \end{cases}$$

We consider the price or rate of the i^{th} quote as a spike, when $Z_{i,k}^{med}$ and $Z_{i,k}^{m}$ are above a certain value Z_{max} fixed by the user. In Figure 6.4, we compare the performances and the robustness of the median filter and the moving average filter on detecting spikes. Note that using a moving average filter, we may confuse spikes and jumps and we may also remove quotes that actually are note spikes (see left panels of Figure 6.4). However, these problems are no longer faced when using a median filter (see right panels of Figure 6.4).

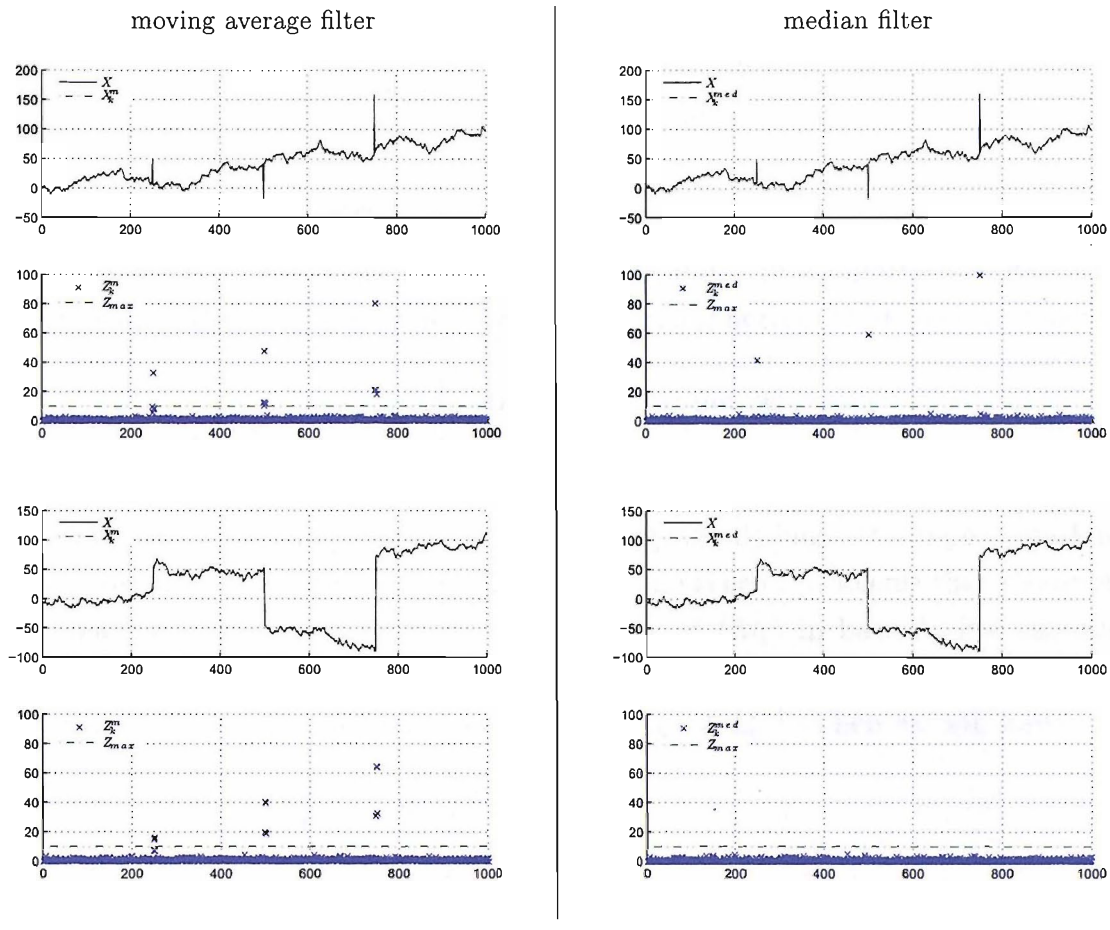

Figure 6.4: Comparison between moving average filter (left panels) and median filter (right panel) on detecting spikes, for k=2.

Let P^{tick} be the smallest tick price change (Pip), we set

$$Z_{i,k}^{ask} = \frac{|P_i^{ask} - P_{i,k,med}^{ask}|}{P^{tick}} \quad \text{and} \quad Z_{i,k}^{bid} = \frac{|P_i^{bid} - P_{i,k,med}^{bid}|}{P^{tick}}$$

so $Z_{i,k}^{ask}$ and $Z_{i,k}^{bid}$ fall into an integer set. We define the statistic of the spikes filter

$$A_i = (Z_{i,k}^{ask} > Z_{max}^{ask}) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases} \quad \text{and } B_i = (Z_{i,k}^{bid} > Z_{max}^{bid}) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

In the case where the bid and ask price are not both spikes, we look at the bid-ask spread value in order to decide whether or the quote has a spike. We define C_i as

$$C_i = (S_i > S_{max}) = \begin{cases} 1 & \text{if true} \\ 0 & \text{if false} \end{cases}$$

where S_{max} is the maximum bid-ask spread allowed. We finally set our statistic

$$N_i^7 = A_i.B_i + A_i.C_i + C_i.B_i = \begin{cases} 1 & \text{if quote has a spikes} \\ 0 & \text{otherwise} \end{cases}$$

6.3.4 Marking bad quotes

Once all the previous filters are performed, every quote will be marked by its corresponding errors. The i^{th} quotes will have the a marking noting M_i defined as follows

$$M_i = \sum_{j=1}^{7} 2^{j-1} N_i^j \begin{cases} = 0 & \text{if quote is not erroneous} \\ \geq 1 & \text{otherwise} \end{cases}$$

Since the dataset of high frequency financial data is generally huge, this type of marking will avoid creating huge database size. To identify the type of problems that the quote contain, one will just need to convert the sequence of $\{M_i\}$ in base 2. For example assuming that the i^{th} quote has a spike during inactive period with a bid price (> 0) greater than the ask price (> 0) and such that the $L_0^i \neq L_0^{i-1}$. Then we will have

$$\begin{cases} N_1 = 0 & P_i^{bid} > 0 \\ N_2 = 0 & P_i^{ask} > 0 \\ N_3 = 1 & S_i = P_i^{ask} - P_i^{bid} < 0 \\ N_4 = 0 & S_i = P_i^{ask} - P_i^{bid} \neq 0 \\ N_5 = 0 & L_0^i \neq L_0^{i-1} \\ N_6 = 1 & T_i \in \{ \text{ inactive periods } \} \\ N_7 = 1 & \text{Quotes has a spike} \end{cases}$$

So $M_i = 100$, which in base 2 corresponds to 1100100, is equivalent to

$$\{N_1, N_2, N_3, N_4, N_5, N_6, N_7\} = \{0, 0, 1, 0, 0, 1, 1\}$$

This way of defining the error type allows us to avoid creating a huge dataset, which will save both time computing and RAM[13] spaces. This simple algorithms is implemented using MATLAB®7 see Figure F.2. In the next Section we perform our filter on some tick by tick data foreign exchange rate indexes.

6.4 Filtering foreign exchange rates

6.4.1 Presentation of the data

We apply our data cleaner to the following foreign exchange rate indexes {AUD/USD, GBP/USD, JPY/USD, EUR/USD}. All these dataset are dated from the 1st January

[13] Random Access Memory

2003 to the 31st December 2003 and are provided by the Securities Industry Research Center of the Asia Pacific (SIRCA). The dataset contains 4 columns as shown in Table 6.1 (date, time, bid price, ask price). They are given as tick by tick data. Table 6.3 gives the number of ticks, the file size of the ultra-high frequency dataset and the Pip of the concerned dataset.

FX	AUD/USD	GBP/USD	JPY/GBP	EUR/USD
Number of Quotes	682 354	4517747	5359302	8348911
File Size (in MB)	25.9	171.3	224.7	316.8
Pip	0.0001	0.0001	0.01	0.0001

Table 6.3: Size of the FX dataset.

6.4.2 Foreign exchange trading hours

As the foreign exchange markets are open 24 hours, it is difficult to present an accurate opening and closing trading time of different markets. Therefore, we illustrate in Figures 6.5 an approximation of the opening trading times for the three main markets. This illustration takes into account the pre-opening market time, as some markets start to trade before the opening time. That practice is called the pre-market trading. It also includes the after-hours trading. Table F.5 shows the trading hours of main markets over the world. We note that the market is inactive between Fridays at 22:00 and Sundays 20:00 for the non Daylight Saving time (DST) period, and between Fridays at 21:00 and Sundays 21:00 for the DST period[14].

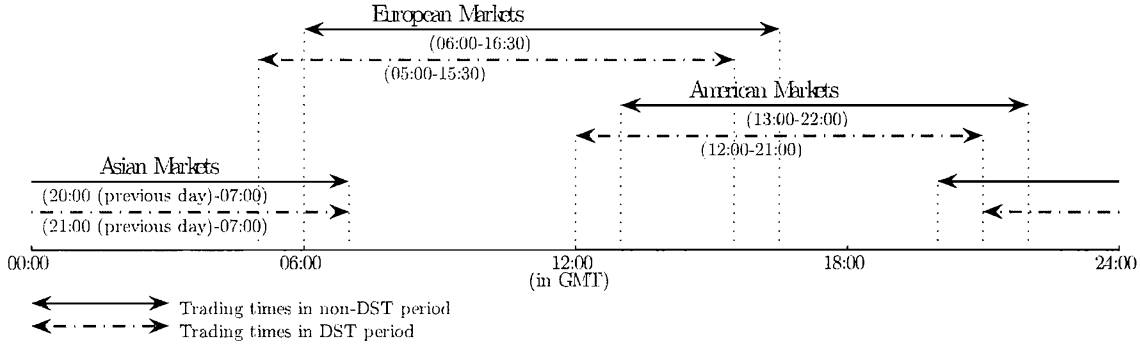

Figure 6.5: Geographical trading times in the foreign markets

Assuming the n^{th} day D_n is a bank holiday, then the inactive period is considered from 22:00GMT of day D_{n-1} to 20:00GMT of day D_n, for non DST period. Similarly for the DST period, the inactive period is from 21:00GMT of day D_{n-1} to 21:00GMT of day D_n. Bank Holidays date may differ from a market to another. While the market

[14]Note that the DST period in Europe. is from the last Sunday of March to the last Sunday of October. Some Asian Markets do not practice the DST such as Hon Kong, Tokyo, Singapore. Concerning The American markets. the DST start the first Sunday of April. For simplicity, we consider for our study the DST of the American markets starting as the European ones.

is active in Europe, for example, we may have bank holidays in Asia, and therefore some markets in Asia are closed. However, during such periods, the activity market over the world may differ from normal days trading, but we will still have trading running during that period. For this reason, we decided not to include bank holidays to detect quotes during inactive period of the markets in our filter. We present in the next section the filtering results of our foreign exchange rate index data.

6.4.3 Filtering results and analysis

For our filter, we use the following settings; $Z_{max}^{ask} = Z_{max}^{bid} = S_{max} = 10^{15}$. Note also as the data may contain up to four spikes in a row, we set $k = 5$, the number of prices taken in each side of a given price P_{t_i}, which makes a window of eleven data to compute the median of P_{t_i}. In our foreign exchange rate data, we did not detect bad quotes of the type: bid price greater than the ask prices and negative bid and ask prices. However, as shown in Table 6.4, our dataset contains between 1% and 2% of bad quotes. Most of the bad quotes arise from the fact that they are repeated quotes lines in the dataset files. Note also, that 0.1% to 0.3% of data are during inactive periods. As was explained earlier, this type of bad quotes is a result of the trading system is being tested, so aberrant quotes are sent during inactive quotes. Table 6.5 shows a higher rate of potential bad quotes during inactive periods. We also note that the presence of spikes in the data is little; only between 0.005% and 0.07%.

	AUD/USD		GBP/USD		JPY/USD		EUR/USD	
Quotes	#	%	#	%	#	%	#	%
$P^{ask} = P^{bid}$	17	0.0025	320	0.007	925	0.017	1087	0.013
Identical quotes	5832	0.85	50965	1.13	74842	1.39	123939	1.49
Inactive quotes	1897	0.28	5750	0.13	6578	0.12	18266	0.22
Spikes	185	0.062	1934	0.043	1144	0.021	432	0.0052
Total bad ticks	8032	1.18	58706	129945	83188	1.55	143413	1.717

Table 6.4: Filtering results of some foreign exchange rates data. The table gives the number and the percentage of potential bad quotes for the AUD/USD, GBP/USD, JPY/USD and EUR/USD rate indexes. Note when reading this table, that in fact one line of quote may contain different types of error.

	AUD/USD		GBP/USD		JPY/USD		EUR/USD	
Quotes	Active	Inac.	Active	Inac.	Active	Inac.	Active	Inac.
$P^{ask} = P^{bid}$	0	0.0023	0	0.007	0	0.017	0.013	0
Identical quotes	0.820	3.585	1.148	3.513	1.389	3.439	1.478	1.411
Spikes	0.0526	2.629	0.0417	0.696	0.020	0.411	0.0047	0.152

Table 6.5: Filtering results of some foreign exchange rates data. The table gives the rates different type error for the AUD/USD, GBP/USD, JPY/USD and EUR/USD rate indexes during active and inactive periods.

[15]This values seemed to be reasonable for us after, FX data studied have almost all their absolute returns less of equal to 10 Pip.

6.4.4 Suggestion for improving the filter

The foreign exchange rate filter can be improved. We can remove more data during low activity period, and less during high activity time. This will suggest, as shown by Dacorogna et al. [83], an estimate of the rate index activity, which means in our case that we estimate the time-change function of the data. Assuming that the logarithmic middle prices $\{X(t)\}_{t\geq 0}$, given by (6.1) , is of the form $\tilde{X}(\theta(t))$, where $\tilde{X}$ is an H-sssi process and θ is a strictly increasing periodic deterministic function, of period $\tau = 1$ week.

$$X(t) = \frac{\log(P_t^{bid}) + \log(P_t^{ask})}{2} \tag{6.1}$$

As $X(t) = \tilde{X}(\theta(t))$, we estimate the function θ and so one gets an estimate $\hat{\tilde{X}}$ of the process $\tilde{X}$. An empirical study on time-change estimation of foreign exchange rates indexes will be presented in the next Chapter. From the process $\hat{\tilde{X}}$, one detects susceptible outliers by using existing techniques such that the Grubbs's test if $\hat{\tilde{X}}$ has its increments normal distributed, see [100]. Note if the increments $\hat{\tilde{X}}$ are α-stable distributed, then we use the outlier statistics presented by Mittnik et al. in [101], which is adapted to heavy tailed processes with infinite variance. As the foreign exchange rate return may not satisfy the Gaussian or the α-stable distribution, this technique is not implemented in our filtering algorithm. Moreover, the filter should not remove a lot of potential outliers, because the more we detect potential outliers, the more we are susceptible to remove true price detected as an outlier. For this reason, we decided to keep our filtering weak, so we do not break the high frequency system of the financial data.

Chapter 7

Empirical time-changed estimation of some UHFD

High frequency financial data are well known to exhibit periodic seasonality in the volatility, see for example [3]. In this Chapter, we perform an empirical study on estimating time change function from high frequency financial data. The Chapter is divided into five Sections. In the first Section, we present the activity estimator introduced by Dacorogna et al. in [3]. We recall our time change estimators described in Chapters 4 and 5. We apply the time change estimators on high frequency financial data (HFD) in Section 7.2. The notion of HFD was introduced in the previous Chapter as well as the definition of UHFD. A detailed study on time-changed data will be conducted on HFD. We also study our estimators in the case of ultra-high frequency financial data (UHFD) in Section 7.3, and on the FTSE100 futures tick by tick data in Section 7.4. In each case, we estimate the activity of the corresponding data and compute the quality of the estimator. We check how well the periodicity is removed and check whether the estimated process time-changed data, also called de-seasonalised data, is self-similar, or at least has some scaling law property. For foreign exchange rates and the FTSE100 futures data, we model the activity to extract the intraday and the weekly seasonality. In the last Section, we show the advantages on using our estimator.

7.1 Preliminaries

7.1.1 Activity market in terms of time change function θ

Assuming a given financial time series $X = \{X(t)\}_{t\geq 0}$, generally defined as the logarithm middle price, is of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ has stationary increments and $\theta \in \Theta([0, +\infty))$, so that its derivative θ' is a periodic function of period $\tau = 1$ week. We define the activity of X, the derivative θ' of the time change function θ. Let

Θ_τ the set of functions defined by

$$\Theta_\tau = \{\theta \in \Theta([0,+\infty)) \text{ s.t } \forall t \geq 0,\ \theta'(t+\tau) = \theta'(t) \text{ and } \forall k \in \mathbb{N},\ \theta(k\tau) = k\tau\}$$

Assuming that the time change function θ exists, so that the time series $X(\theta^{-1}(t))$ has stationary increments, then the function θ is unique. Indeed, setting θ_1 and θ_2 two functions of the set Θ_τ, such that processes $\tilde{X}^1(t) = X(\theta_1^{-1}(t))$ and $\tilde{X}^2(t) = X(\theta_2^{-1}(t))$ have stationary increments, then one has for all positive s

$$\begin{aligned}\left\{\tilde{X}^1_{\theta_2^{-1}\circ\theta_1(t+s)} - \tilde{X}^1_{\theta_2^{-1}\circ\theta_1(s)}\right\}_{t\geq 0} &= \{\tilde{X}^2_{t+s} - \tilde{X}^2_s\}_{t\geq 0} \\ &\overset{(fdd)}{=} \{\tilde{X}^2_t - \tilde{X}^2_0\}_{t\geq 0} \\ &\overset{(fdd)}{=} \left\{\tilde{X}^1_{\theta_2^{-1}\circ\theta_1(t)} - \tilde{X}^1_{\theta_2^{-1}\circ\theta_1(0)}\right\}_{t\geq 0}\end{aligned}$$

Finally, one gets

$$\left\{\tilde{X}^1_{\theta_2^{-1}\circ\theta_1(t+s)-\theta_2^{-1}\circ\theta_1(s)} - \tilde{X}^1_0\right\}_{t\geq 0} \overset{(fdd)}{=} \left\{\tilde{X}^1_{\theta_2^{-1}\circ\theta_1(t)-\theta_2^{-1}\circ\theta_1(0)} - \tilde{X}^1_0\right\}_{t\geq 0}$$

Setting $g = \theta_2^{-1} \circ \theta_1$, like $\tilde{X}^1(t)$, the process $\tilde{X}^1(g(t))$ also has stationary increments, one must have

$$g(t+s) - g(s) = g(t) - g(0)$$

This last implies that the function g is linear. Moreover functions θ_1 and θ_2 belong to the set Θ_τ, so for all $k \in \mathbb{N}$, $\theta_1(k\tau) = \theta_2(k\tau) = k\tau$ which implies $g(k\tau) = k\tau$ for every k. Whence for all $t \geq 0$, $g(t) = t$, which is equivalent to say $\theta_1 = \theta_2$. This last equality ensures that the activity estimated from financial data are unique in the sense that we get the stationarity of the time series increments under the new time scale $\theta(t) = \int_0^t \theta'(s)ds$. This new time scale is called θ-time by Dacorogna. We introduce the scaling law estimator described in [3] in the following section.

7.1.2 Time change estimation using the scaling law

For $t = j\delta t$, $j = 0, \ldots, M-1$, put

$$\overline{|\Delta X_{t,\delta t}|} = \frac{1}{N}\sum_{k=0}^{N-1} |\Delta X_{t+k\tau,\delta t}|, \quad \overline{|\Delta X_{\cdot,\delta t}|} = \frac{1}{MN}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} |\Delta X_{j\delta t+k\tau,\delta t}|$$

where $\Delta X_{t,\delta t} = X(t+\delta t) - X(t)$. Dacorogna et al [3] gave the following estimate of the discrete activity between time t and $t+\delta t$ noted $a_{[t,t+\delta t)}$. For $t = j\delta t$, $j = 0, \ldots, M-1$,

$$\hat{a}^0_{[t,t+\delta t)} = \frac{c_0}{\delta t}\left(\overline{|\Delta X_{t,\delta t}|}\right)^{1/\hat{H}}, \tag{7.1}$$

where $\hat{H}$ is an estimate of H and c_0 is a scaling constant, chosen so that

$$\frac{1}{\tau}\sum_{j=0}^{M-1}\hat{a}^0_{[j\delta t,(j+1)\delta t)}\delta t = 1$$

Dacorogna et al assumed that $\log \mathbb{E}|\Delta X_{t,\delta t}| \sim H \log \delta t$. This is unsatisfactory, as our starting point is to assume that X exhibits periodic fluctuations in volatility, which necessarily interfere with its scaling behaviour. A more reasonable assumption is that $\tilde{X}$ exhibits scaling behaviour.
Given $X(t) = \tilde{X}(\theta(t))$ where $\tilde{X}$ is H-sssi, we have

$$\begin{aligned}
\mathbb{E}|\Delta X_{t,\delta t}| &= \mathbb{E}|X(t+\delta t) - X(t)| \\
&= \mathbb{E}|\tilde{X}(\theta(t+\delta t)) - \tilde{X}(\theta(t))| \\
&= \mathbb{E}|\tilde{X}(\theta(t+\delta t) - \theta(t))| \text{ by stationarity increments} \\
&= (\theta(t+\delta t) - \theta(t))^H \mathbb{E}|\tilde{X}(1)| \text{ by self-similarity.}
\end{aligned}$$

Rearranging with $a_{[t,t+\delta t)} = \dfrac{\theta(t+\delta t) - \theta(t)}{\delta t}$, we get

$$a_{[t,t+\delta t)} = \frac{1}{\delta t}\left(\frac{\mathbb{E}|\Delta X_{t,\delta t}|}{\mathbb{E}|\tilde{X}(1)|}\right)^{1/H} \tag{7.2}$$

Using $\overline{|\Delta X_{t,\delta t}|}$ to estimate $\mathbb{E}|\Delta X_{t,\delta t}|$ gives us the estimator (7.1), provided we can estimate H. Note that the term $(\mathbb{E}|\tilde{X}(1)|)^{1/H}$ is absorbed into the constant c_0.
Dacorogna et al estimated H from the slope of a regression of $\log \overline{|\Delta X_{\cdot,\delta t}|}$ against $\log \delta t$. However, if we have $X(t) = \tilde{X}(\theta(t))$ where $\tilde{X}$ is H-sssi, then (7.2) gives

$$\log \mathbb{E}\overline{|\Delta X_{\cdot,\delta t}|} = H \log \delta t + \log\left(a_{\sum}(H,\delta t)\right) + \log \mathbb{E}|\tilde{X}(1)|$$

where the function $a_{\sum} : (0,1) \times (0,\tau) \to [0,\infty)$ is defined by

$$a_{\sum}(H,\delta t) = \frac{1}{MN}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} a^H_{[j\delta t + k\tau,(j+1)\delta t + k\tau)}$$

Thus we see that the H estimator of Dacorogna et al fails to allow for the effect of the time-change, namely the $a_{\sum}(H,\delta t)$ terms. The solution to this problem is provided by the Embedded-Branching-Process (EBP) estimator for H, recently introduced by Jones and Shen [7]. The significance of the EBP estimator is that it is unaffected by a time-change, so that the self-similarity index of $\tilde{X}$ can be estimated using the observed process X. This last is described in Chapter 3 Section 3.3.

7.1.3 Estimation of periodic activity

We define by $\hat{a}^1(t)$, $\hat{a}^2(t)$ and $\hat{a}^3(t)$ respectively the $1/H$-Variation, log-variation and the crossing estimators. Assuming that the activity is weekly periodic, one can either consider the averaged or the median of the activities over the periods of the data sample. However, there is an advantage using the median instead of the mean in some cases. First, one does not need to consider bank holidays that makes computation harder for the mean. Second, the median is not sensible to outliers. Indeed, the data may contain some jumps; this creates spikes in the activity estimate. Thus, taking the median avoids taking the spikes into consideration. However, we will see later that the choice between median and mean can be discussed.
Let us recall our estimators assuming X of the form $X(t) = \tilde{X}(\theta(t))$, where $\tilde{X}$ is H-sssi and $\theta \in \Theta_\tau$. An estimate of the activity of the time series X between time t and $t+\delta t$ are given by

- The $1/H$-variation

$$\hat{a}^1_{[t,t+\delta t)} = \frac{C_1}{\delta t} \sum_{k=0}^{2^n-1} \left| X\left(t + \frac{k+1}{2^n}\delta t\right) - X\left(t + \frac{k}{2^n}\delta t\right)\right|^{1/\hat{H}}$$

- The log-variation

$$\hat{a}^2_{[t,t+\delta t)} = \frac{C_2}{\delta t} \exp\left(\frac{1}{2^n}\sum_{k=0}^{2^n-1} \log\left(\frac{|X(t+\frac{k+1}{2^n}\delta t) - X(t+\frac{k}{2^n}\delta t)|^{1/\hat{H}}}{2^{-n}\delta t}\right)\right)$$

- The crossing number

$$\hat{a}^3_{[t,t+\delta t)} = \frac{C_3}{\delta t}\mathcal{N}^h_{(t,t+\delta t)}(X)$$

where for $i = \{0,1,2,3\}$, C_i is the norming constant so that $\sum a^i_{[t,t+\delta t)} = 1$. Obviously the estimators depend on the choice of n. We now define $\hat{a}^i_m$ and $\hat{a}^i_{med}$ as the activity estimator by taking the mean and the median over weeks respectively defined by

$$\hat{a}^i_m(t) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{a}^i_{[t+k\tau,t+\delta t+k\tau)} \quad \text{and} \quad \hat{a}^i_{med}(t) = \text{Median}\left(\hat{a}^i_{[t+k\tau,t+\delta t+k\tau)}\right)$$

where $k \in \{0,1,\ldots,N\}$ and N is the number of weeks, τ represents the one week period time. As means for comparison, we will also use the median instead of the mean in the expression of the scaling law method without any theoretical justification.

7.1.4 Quality of the estimated activity

To check how well the activity is estimated, one needs to see how good it is when the seasonality is removed. To proceed, one can, for example, compute the autocorrelation of the process before and after removing the daily and weekly seasonality of the data.

The autocorrelation function estimate is often used in the literature to detect seasonality. Indeed, it appears a periodic pattern in the function. A similar way to check how well the strong seasonality is removed from the data, is to calculate the power spectrum of time series before and after time changing it. However, these two procedures are graphical tests and we will require a quantitative test for a better comparison. The following "relative volatility" measure of how effective a time-change θ is at removing periodic fluctuations in the activity, has been proposed by Dacorogna et al. in [83]. For $\tilde{X}(t) = X(\theta^{-1}(t))$, we define

$$Q(\theta) = \sqrt{\frac{\sum_{j=0}^{M-1}\left(\overline{|\Delta\tilde{X}_{j\delta t,\delta t}|} - \overline{|\Delta\tilde{X}_{.,\delta t}|}\right)^2}{\sum_{j=0}^{M-1}\left(\overline{|\Delta X_{j\delta t,\delta t}|} - \overline{|\Delta X_{.,\delta t}|}\right)^2}} \tag{7.3}$$

where for $t = j\delta t, j = \{0, \ldots, M-1\}$

$$\overline{|\Delta X_{t,\delta t}|} = \frac{1}{N}\sum_{k=0}^{N-1}|\Delta X_{t+k\tau,\delta t}| \text{ and } \overline{|\Delta X_{.,\delta t}|} = \frac{1}{NM}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}|\Delta X_{j\delta t+k\tau,\delta t}|$$

Smaller values of Q are considered better. We use Q to compare the activity estimators for our foreign exchange rate index. Note that δt here can be different from the one used to estimate the activity. We recall it δt_Q in order to make the difference with δt, notation already used in our estimator.

7.1.5 Investigating for self-similarity

We need to test whether our de-seasonalised high frequency data is self-similar[1]. In the literature several tests exist. In case of self-similar processes with gaussian and stationary increments, we refer to Bardet [102] for a theoretical work and [103] for application in finance. In the general case, we refer to Jones and Shen in [7]. This last was used to detect self-similarity at different crossing levels for a better estimate of the self-similarity index. To test self-similarity, we use the relation (7.4) for $\tilde{X}$ H-sssi process. Equality (7.4) can be found in [103].

$$\mathbb{V}ar\left(\tilde{X}(t+\delta t) - \tilde{X}(t)\right) = \delta t^{2H}\mathbb{V}ar\left(\tilde{X}(1)\right) \tag{7.4}$$

Setting $K = \log\left(\mathbb{V}ar\left(\tilde{X}(1)\right)\right)$, one has

$$\log\left(\mathbb{V}ar\left(\tilde{X}(t+\delta t) - \tilde{X}(t)\right)\right) = 2H\log(\delta t) + K \tag{7.5}$$

The log prices return do not need to be normal distributed, we just limit ourselves to the fact that for δt in a given range, say $\delta t \in \{\delta t_1, \ldots, \delta t_n\}$, the process $\tilde{X}$ is self-similar

[1] or has at least some scaling property

if $\log\left(\mathbb{V}ar\left(\tilde{X}(t+\delta t_i)-\tilde{X}(t)\right)\right)$ is a linear function of $\log(\delta t_i)$ for $i \in \{1,\dots,n\}$. A linear regression using a least squares method allows us to find the values H and K. In the case of a Gaussian process, Bardet's test can be applied by measuring the linear regression error; we refer to the corresponding article [102]. Unfortunately, this test is not sufficient. It does not tell us if the equality in distribution (7.6) is satisfied when $\tilde{X}$ has no Gaussian distributed increments.

$$\tilde{X}(1) \overset{(d)}{=} \frac{1}{\delta t^H}\tilde{X}(\delta t) \tag{7.6}$$

For this reason, we also need to study the distribution of the financial data return of a given range of δt time scales. We set $\tilde{Y}_{\delta t} = \frac{1}{\delta t^{\hat{H}}}\tilde{X}(\delta t)$, where $\hat{H}$ is the self-similar index estimated by the EBP estimator. The family of $\tilde{Y}_{\delta t_i}$ for $i \in \{1,\dots,n\}$, must have the same distribution. One way to check this hypothesis is to use a Kolmogorov-Smirnov test (see D.1 for details). For $i \neq j$, a Kolmogorov-Smirnov test will confirm whether $\tilde{Y}_{\delta t_i}$ and $\tilde{Y}_{\delta t_j}$ have the same distribution.

7.2 Application to HFD

Throughout this section, we will consider the logarithm middle prices of the foreign exchange rate index, on which all the study will be based on

$$X(t) = \frac{\log(P_t^{bid}) + \log(P_t^{ask})}{2}$$

7.2.1 Estimating the activity

We apply the time change estimator on the EUR/USD and GBP/USD exchange rate index using a one minute interval. These data are dated from January 2001 to December 2005. The graphs in Figure 7.1 give an estimate of the global self-similarity index of the data, using the EBP estimator. The little circles represent the levels at which the self-similarity index was estimated. The choice of these levels was decided by applying a self-similarity test on all the levels. This procedure is described in [22]. The estimated global indexes of self-similarity $\hat{H}$ for the EUR/USD and GBP/USD rate index are respectively 0.534 and 0.526. Once having the estimated global index, we estimate the activity using the estimators $\hat{a}^0$, $\hat{a}^1$, $\hat{a}^2$ and $\hat{a}^3$ for a step time of 30 minutes. Considering the time measure is in hour, we set $\tau = 168$ and $\delta t = 1/2$. For the $1/H$-variation and the log-variation estimators, we set $n = 7$; this corresponds to taking a subinterval of time on the interval $[t, t+\delta t)$ of approximatively 14 seconds. We will then have $2^7 - 1 = 127$ extra data compare to the scaling law method. For the crossing number, we set the size of the crossings as the standard deviation of the observed jumps increments of the data. For each rate index, we estimate the averaged

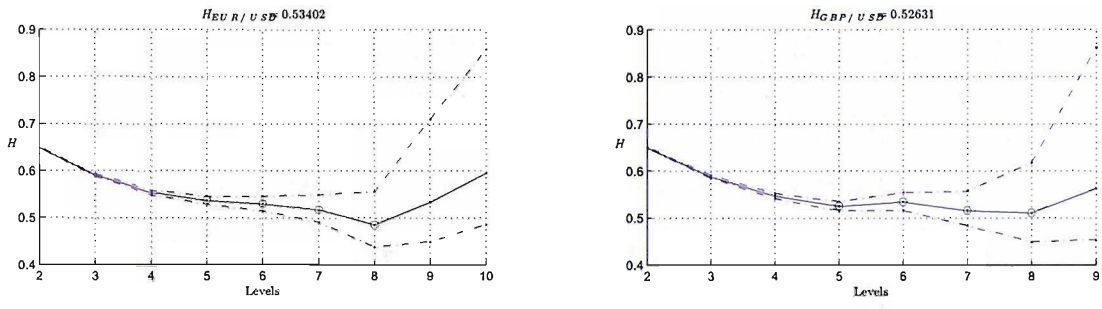

Figure 7.1: Global self-similar index estimation using the EBP estimator to the EUR/USD (left panel) and to the GBP/USD (right panel) .

and the median activities noted respectively $\hat{a}_m^i$ and $\hat{a}_{med}^i$, for $i \in \{0, 1, 2, 3\}$. The estimated activities are represented in Figures F.3 and F.4 for the EUR/USD rate index, and in Figures F.5 and F.6 for the GBP/USD rate index. In Figures 7.2 and 7.3 the data before and after the time change are represented (only 2 months (January-February 2003) of time-changed foreign exchange data using the crossing method are represented).

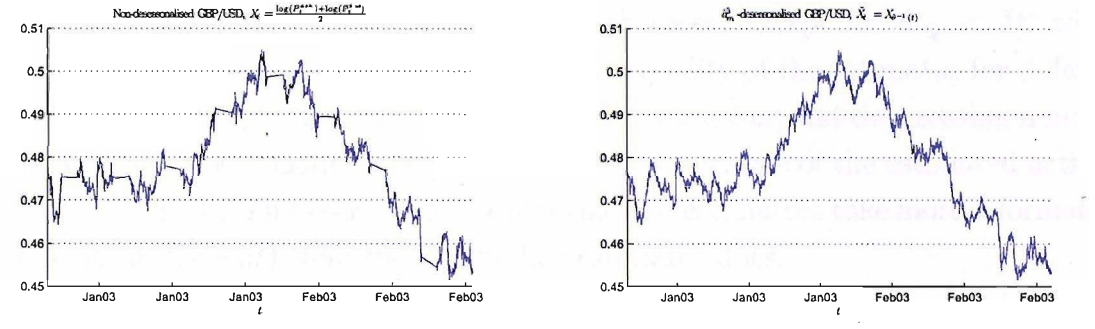

Figure 7.2: GBP/USD rate index before (left panel) and after (right panel) time change

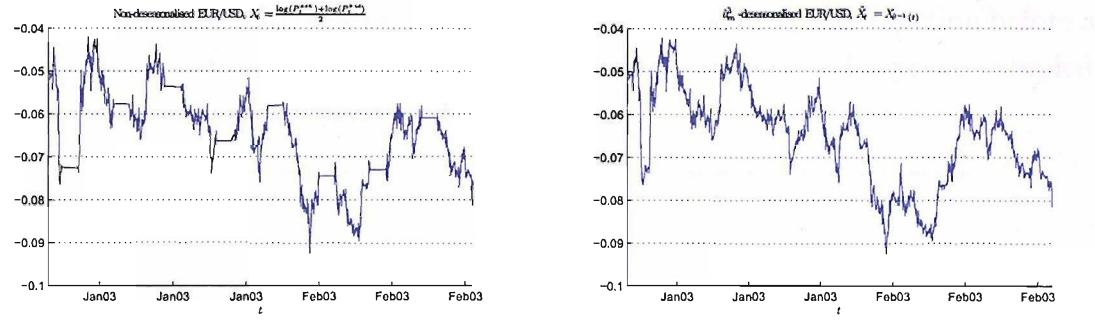

Figure 7.3: EUR/USD rate index before (left panel) and after (right panel) time change

In order to test their performances, we estimated the quality (defined by (7.3)) of each estimator. Tables 7.1 and 7.2 present respectively the quality values for the GBP/USD rate index and the EUR/USD rate index. We compute the quality of the estimators for $\delta t_Q = \{5, 30, 60\}$ in minutes.

Estimator	$\hat{a}^0_m/\hat{a}^0_{med}$	$\hat{a}^1_m/\hat{a}^1_{med}$	$\hat{a}^2_m/\hat{a}^2_{med}$	$\hat{a}^3_m/\hat{a}^3_{med}$
$\delta t_Q = 5$ min	0.2727/0.2438	0.2658/0.1990	0.2338/0.2628	0.2160/0.1916
$\delta t_Q = 30$ min	0.1389/0.1606	0.1767/0.1539	0.1709/0.2198	0.1509/0.1483
$\delta t_Q = 60$ min	0.1217/0.1357	0.1616/0.1505	0.1545/0.1868	0.1526/0.1556

Table 7.1: Quality of the estimators for the GBP/USD rate index .

Estimator	$\hat{a}^0_m/\hat{a}^0_{med}$	$\hat{a}^1_m/\hat{a}^1_{med}$	$\hat{a}^2_m/\hat{a}^2_{med}$	$\hat{a}^3_m/\hat{a}^3_{med}$
$\delta t_Q = 5$ min	0.3216/0.2328	0.2939/0.1951	0.2155/0.2296	0.2277/0.2107
$\delta t_Q = 30$ min	0.1695/0.1624	0.2120/0.1709	0.1571/0.1763	0.2066/0.1983
$\delta t_Q = 60$ min	0.1402/0.1424	0.2057/0.1635	0.1512/0.1715	0.2089/0.1961

Table 7.2: Quality of the estimators for the EUR/USD rate index .

Two main comments can be extracted from these results. The first concerns the choice between median and mean of the activities. The second concerns the effect of the choice of δt_Q. First we note that the median of the activities as estimator performs better than the mean of the activities for the $1/H$-variation estimator. However, it is the opposite for the log-variation. Note also, that like the log-variation estimate, the crossing number performs slightly better when using the median. The scaling law method gives a similar performance in both cases, except for $\delta t_Q < \delta t$, where the median gives better performance. Second, the quality of the estimator for different δt_Q shows that generally the $1/H$-variation, the log-variation, and the crossing number, capture better the seasonality at small scales, even if the scale of the estimated activity δt is bigger than δt_Q. This come from the fact that our estimators take more information in the interval $[t, t + \delta t)$ than the scaling law estimator does.

7.2.2 Investigating for stationarity of the return

To check whether the seasonality exhibited by the absolute log-middle prices return of the foreign exchange data was removed, we draw its autocorrelation function before and after deseasonalisation[2]. In Figure 7.4, autocorrelation functions for prices sampled at one minute intervals are represented.

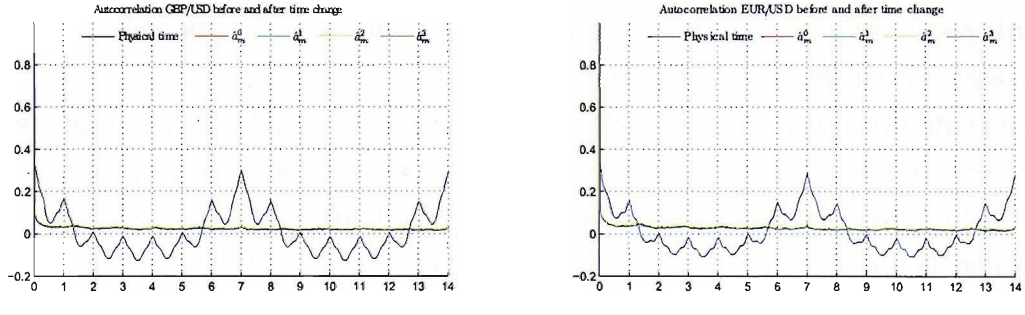

Figure 7.4: Autocorrelation of the absolute log-price return of GBP/USD (left panel) and EUR/USD (right panel) before and after time change.

[2]We consider the case where the activities were measured taking the means over weeks.

Figure 7.5 shows the autocorrelation function for 10 weeks data and thus for each method.

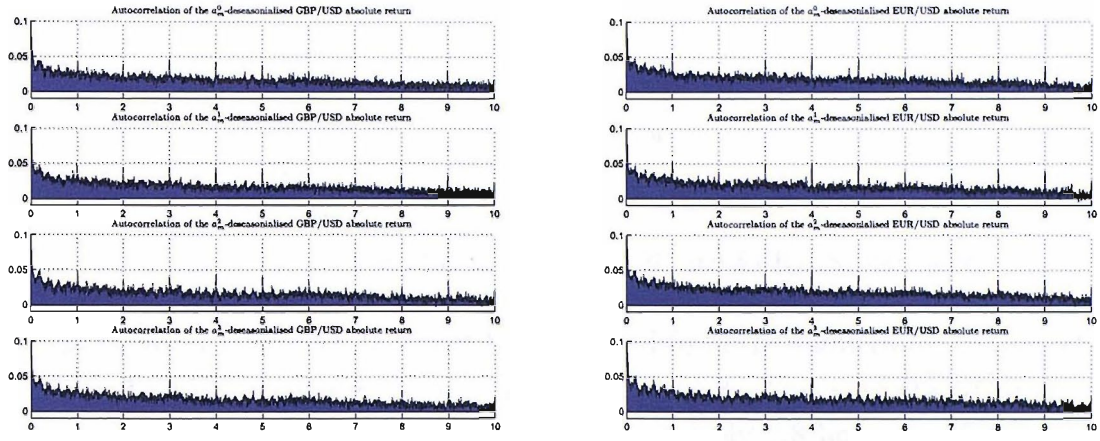

Figure 7.5: Autocorrelation of the absolute log-price return of GBP/USD (left panel) and EUR/USD (right panel) for each de-seasonalisation method.

For each method, the seasonal pattern exhibited by the absolute log-middle price returns of our foreign exchange rate was removed. However, a small weekly periodicity appears in Figure 7.5. This is due to the fact that the activity is measured at 30 minutes interval. For example, assuming that our estimator detects activity between 8:00 and 8:30 but in fact the market starts at 8:05, after time-changing our data, we will have a 5 minute period every week where the price does not change. This will imply a small periodic pattern in the autocorrelation function as seen in Figure 7.5. This can be fixed by measuring the activity at different intervals of time over the week. For example, we use small time interval during high activity periods and long interval during low activity. This fact was already described by Dacorogna et al. in [83].

7.2.3 Investigating for self-similarity

Since our estimators assume the form $X(t) = \tilde{X}(\theta(t))$ for our financial data where $\tilde{X}$ is H-sssi , one must check if this assumption is satisfied by the deseasonalised data. We could have limited ourselves by looking at Figures 7.1, and say that we have self-similarity through levels 5 to 8. This test does not tell us for which time scales we have self-similarity. For this study, let us consider the estimated deseasonalised data $\hat{\tilde{X}}$ using the crossing time change estimator. We analyse the scaling properties of our data by estimating the self-similarity index for a range of δt values $\{5, 10, 15, 20, 30, 40, 60, 120, 240\}$ in minutes using (7.5). The results are shown in Figure 7.6. The scaling exponent H is given by fitting a linear regression line through the points over the range of δt using a least squares method. The self-similarity indexes found after time change for the GBP/USD and the EUR/USD rate index are respectively 0.506 and 0.507. These values differ from the values found before time changing

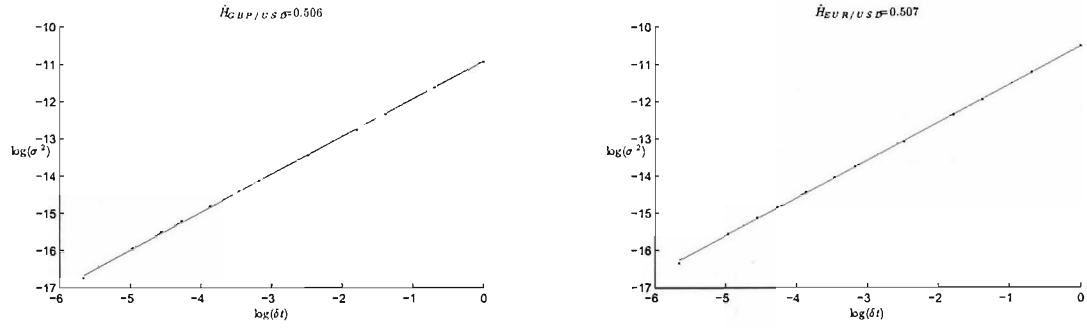

Figure 7.6: Estimated H of the GBP/USD and EUR/USD rate index.

the data. We may expect H to not be estimated from the same range of time scales δt. In fact, the EBP estimates H from level 5 to level 8. This correspond to an average of time between two hitting points of approximatively 7.8 hours to 17.8 days for the GBP/USD and 8.5 hours to 19.5 days for the EUR/USD. Using (7.5) for time scales δt between 12 hours and 16 days, we obtain an estimate of H still different for the one estimated from the EBP estimator. One gets an estimate of 0.519 for the GBP/USD and 0.49 for the EUR/USD. For smallest time scales, the EBP estimator is not able to give us a good estimate of H, as the crossings at small levels are badly approximated. As we have a scaling property of $\tilde{X}$ for the range of scale $\delta t \in \{5, 10, 15, 20, 30, 40, 60, 120, 240\}$ in minutes, one computes H and K for each week of the data $\hat{\tilde{X}}$ by using (7.5). If we have self-similarity, we must have H and K constant over the weeks.

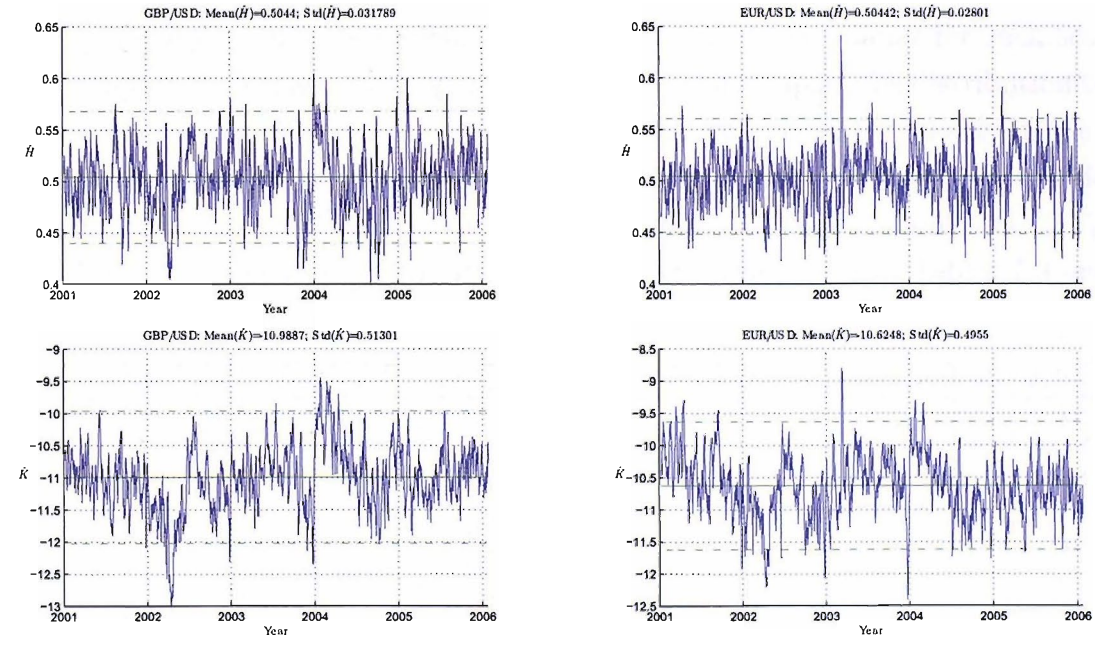

Figure 7.7: Estimated H and K of the GBP/USD and EUR/USD rate index. The green line represent the mean of the corresponding estimates and the dashed line represent twice the standard deviation distanced from the mean.

Figure 7.7 illustrates an estimate of H and K of the GBP/USD and EUR/USD rate indexes over weeks. We found respectively an average of 0.504 with a standard devia-

tion of 0.032 and 0.504 with a standard deviation of 0.028 for the self-similarity index. This seems to be reasonable to consider that our time changed foreign exchange data satisfies the scaling law for time scales δt between 5 minutes and 4 hours. Keeping the given range of time scales δt, we check whether the distribution of our log middle prices return is similar at different time scales increments. Let

$$\tilde{Y}_{\delta t} = \left\{ \frac{\tilde{X}(j\delta t) - \tilde{X}((j+1)\delta t)}{\delta t^{\hat{H}}} \right\}_{\{j=0,1,\ldots\}}$$

where $\hat{H}$ is given by Figure 7.6. The frequency distribution of the family of $\tilde{Y}_{\delta t}$ are represented in Figure 7.8.

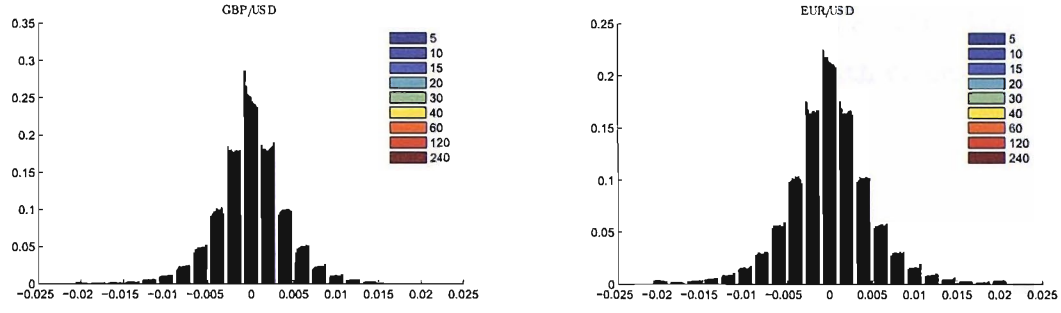

Figure 7.8: Distribution of GBP/USD (left panel) EUR/USD (right panel) return for different time scale

Note that the rate indexes returns distribution seems to be similar for each scale for some range of δt. Using a Kolmogorov-Smirnov test to compare the distribution two by two, one can confirm for which range of time scales we have self-similarity in the data. Taking $(\delta t, \delta t') \in \{(5,10),(5,20),\ldots,(5,240),(10,30),\ldots,(120,240)\}$, we consider the following hypothesis : H_0: "$\tilde{Y}_{\delta t_i}$ and $\tilde{Y}_{\delta t_j}$ have the same distribution". H_1: "$\tilde{Y}_{\delta t}$ and $\tilde{Y}_{\delta t'}$ do not have the same distribution". We present the results in Tables 7.3 and 7.4, where 0 stands for H_0 and 1 for H_1.

-	5	10	20	30	40	60	120	240
5	-	1	1	1	1	1	1	1
10	-	-	1	1	1	1	1	1
20	-	-	-	1	1	1	1	1
30	-	-	-	-	0	1	1	1
40	-	-	-	-	-	0	0	0
60	-	-	-	-	-	-	0	0
120	-	-	-	-	-	-	-	0
240	-	-	-	-	-	-	-	-

Table 7.3: Hypothesis test (GBP/USD).

-	5	10	20	30	40	60	120	240
5	-	1	1	1	1	1	1	1
10	-	-	1	1	1	1	1	1
20	-	-	-	0	0	0	0	0
30	-	-	-	-	0	0	0	0
40	-	-	-	-	-	0	0	0
60	-	-	-	-	-	-	0	0
120	-	-	-	-	-	-	-	0
240	-	-	-	-	-	-	-	-

Table 7.4: Hypothesis test (EUR/USD).

The self-similarity is satisfied for $\delta t \in \{20,\ldots 240\}$ in minutes for the EUR/USD rate index and $\delta t \in \{40,\ldots 240\}$ for the GBP/USD rate index.

7.2.4 Modelling the activity: Case of the crossing estimator

We model the activity using a sum of kernel density functions. We assume that the intraday activity is similar for each day of the week with a weighed factor depending on the day of the week. This is summarised by the following model

$$a_M^i(t) = \phi_N(t - 24\lfloor t/24 \rfloor) \sum_{j=1}^{7} w_j \mathbb{I}_{\{t \in I_{t,24}\}}(t) \tag{7.7}$$

where $\phi_N(t) = \sum_{k=1}^{N} p_k \exp\left(-\frac{(t-m_k)^2}{s_k}\right)$ and $I_{t,24} = [24\lfloor t/24 \rfloor, 24(\lfloor t/24 \rfloor + 1))$, w_j represent the attributed weight for each day of the week. $\mathbb{I}_{\{t \in I_{t,24}\}}(t) = 1$ if $t \in I_{t,24}$, 0 otherwise. The intra-day activity is given by the left term $\phi_N(t)$ of $a_M^i(t)$. We estimate the parameter w_j, p_k, m_k and s_k, using a least squares method with constraint

$$\int_0^{24} \phi_N(t) = 1 \text{ and } \sum_{i=1}^{7} w_i = 1$$

We separated winter time and summer time. We chose $N = 6$. The results are given in Figure 7.9 GBP/USD rate index and Figure 7.10 for the EUR/USD rate index. Note that the time is given in Central European Time (CET).

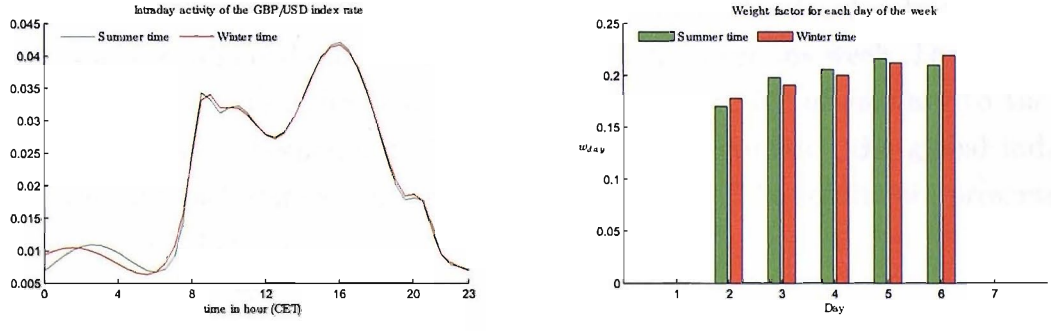

Figure 7.9: Intra-day day activity (left panel) and the day factor (right panel) of the GBP/USD rate index

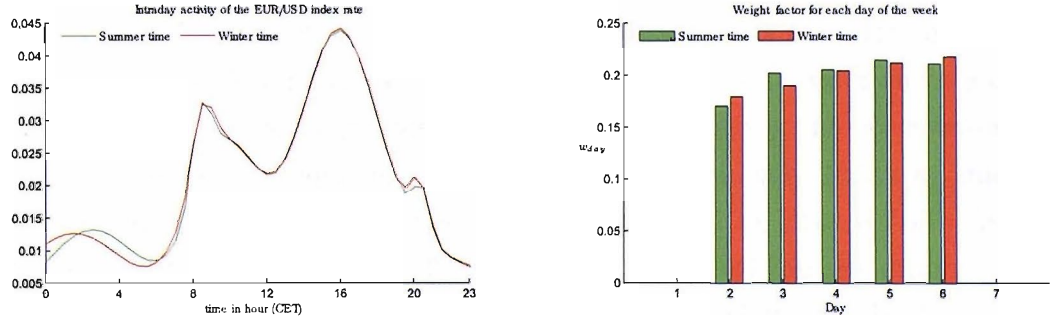

Figure 7.10: Intra-day day activity (left panel) and the day factor (right panel) of the EUR/USD rate index

Note that the activity is higher in the middle of the week in summer time compared to the winter period, and lower on Monday and Friday compared to the winter period. Note also that the weekly activity of the two rate indexes is almost identical.

7.3 Application to UHFD

We take the logarithmic middle price $X_{\alpha_{FX}}(t)$ between the bid and the ask price at time t respectively $P^{bid}_{\alpha_{FX},t}$ and $P^{ask}_{\alpha_{FX},t}$, where

$$\alpha_{FX} = \{\text{AUD/USD, GBP/USD, JPY/USD, EUR/USD}\}$$

We define the logarithmic middle price $X_{\alpha_{FX}}(t) = \dfrac{\log(P^{bid}_{\alpha_{FX},t}) + \log(P^{ask}_{\alpha_{FX},t})}{2}$.
Before manipulating the data, we were required to clean them. The cleaning process used is described in Chapter 6.

7.3.1 Time-changes UHFD

We apply our estimators to UHFD and we compare their performances to the scaling law method used by Dacorogna to estimate the activiy. The data taken into consideration are the AUD/USD, GBP/USD, JPY/USD and EUR/USD rate indexes, dated from 1^{st} of January 2003 to the 31^{st} of December 2003. We decided to take Sunday 00:00:00 as a start and end date to estimate the activity over one week. For this reason the activity of each rate index will be estimated from the 5^{th} of January to the 28^{th} of December. This corresponds to 51 weeks. We first estimated the global index of self-similarity for each data set using the EBP estimator. The results are presented in Figure 7.11 and Table 7.5.

$\hat{H}_{\text{AUD/USD}}$	$\hat{H}_{\text{GBP/USD}}$	$\hat{H}_{\text{JPY/USD}}$	$\hat{H}_{\text{EUR/USD}}$
0.499	0.461	0.483	0.498

Table 7.5: Global Hurst index estimation using the EBP estimators.

Note that the estimation of the global index of self-similarity for ultra high frequency financial data gives a lower value of H compared to the 1 minute interval data of section 7.2. This is due to the fact that our data are noisy[3], so taking the observation at very small scales looks very rough, and therefore the self-similarity index at small scales tends to be lower than at large scales. This is confirmed by Figure 7.11. As one has no real crossings, we must see a tail[4] in the three first levels on the graphs. The first values starting from this tail should be above the estimated self-similarity index of the data. For the AUD/USD rate index this is the case, as the dataset contains around 6.10^5 quotes and so has less noise than the GBP/USD, JPY/USD and EUR/USD

[3]The variation at very small scale are considered as noise.
[4]due to the bad crossing points approximation at low level size.

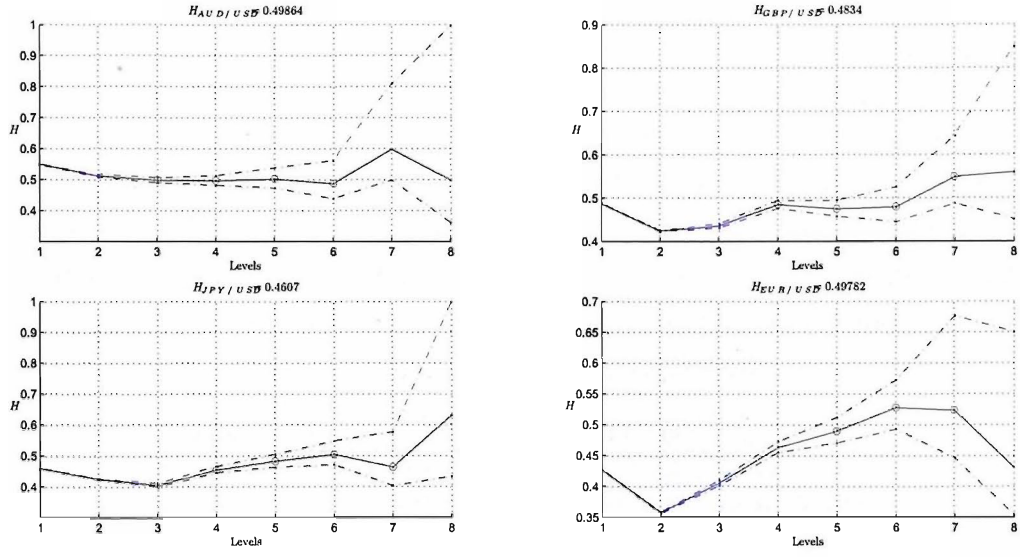

Figure 7.11: Self-similar index estimate using the EBP estimator to the AUD/USD (top-left panel), GBP/USD (top-right panel), JPY/USD (bottom-left panel), EUR/USD (bottom-right panel).

rate indexes, for which the dataset contains respectively around $4,5.10^6$, $5,3.10^6$ and $8,3.10^6$. Further, we may also expect that later on, estimating the activity using the crossing may have a bad performance for low crossing levels. As the crossing method uses the standard deviation of the jumps of the data as crossing size, we will have to increase the crossing size to avoid the prices variation (noise) at very small scales, which disturbs the crossing estimator. We estimate the activities of the UHFD using our time change estimators. The setting remains similar to those used in Section 7.2. We illustrate the results in Figures F.7, F.8, F.9, F.10, F.11, F.12, F.13 and F.14 in the Appendix. The quality of the estimators is presented in Table 7.6. As for the 1 minute

X/USD	Estimator	$\hat{a}^0_m/\hat{a}^0_{med}$	$\hat{a}^1_m/\hat{a}^1_{med}$	$\hat{a}^2_m/\hat{a}^2_{med}$	$\hat{a}^3_m/\hat{a}^3_{med}$
AUD	$\delta t_Q = 5$ min	0.3298/0.3621	0.3185/0.2826	0.291/0.3492	0.275/0.2846
	$\delta t_Q = 30$ min	0.2523/0.3002	0.2667/0.2643	0.2827/0.3689	0.2543/0.28
	$\delta t_Q = 60$ min	0.2129/0.2637	0.251/0.2411	0.2529/0.3465	0.2256/0.2631
GBP	$\delta t_Q = 5$ min	0.32/0.3535	0.3614/0.2742	0.3119/0.3899	0.2845/0.3193
	$\delta t_Q = 30$ min	0.2627/0.2935	0.3249/0.2688	0.2989/0.3829	0.2979/0.3131
	$\delta t_Q = 60$ min	0.2193/0.2484	0.2922/0.2368	0.2595/0.3442	0.2582/0.2743
JPY	$\delta t_Q = 5$ min	0.3855/0.3785	0.3492/0.3068	0.3006/0.3442	0.3291/0.3431
	$\delta t_Q = 30$ min	0.3078/0.3045	0.3305/0.2991	0.2766/0.3335	0.3377/0.3472
	$\delta t_Q = 60$ min	0.2798/0.2957	0.3046/0.2848	0.2837/0.3271	0.3132/0.3318
EUR	$\delta t_Q = 5$ min	0.3101/0.3407	0.3258/0.2982	0.2758/0.3149	0.3784/0.403
	$\delta t_Q = 30$ min	0.2551/0.3017	0.3213/0.3033	0.287/0.3023	0.3938/0.3942
	$\delta t_Q = 60$ min	0.21/0.2467	0.2911/0.283	0.262/0.2791	0.3597/0.4049

Table 7.6: Quality of the estimators for on tick by tick data.

interval data, the scaling law does not capture enough information on the interval of length δt, and so the quality given for $\delta t_Q < \delta t$ is not satisfactory compared to the quality that is given for $\delta t_Q \geq \delta t$. The other estimators are less affected. However, as

expected, the crossing method performs badly for the JPY/USD and EUR/USD rate indexes. We recalculate the quality for crossing sizes in $\{2h, 3h, 4h, 5h, 6h, 7h\}$ where h is the standard deviation of the jumps of EUR/USD rate index. The results are shown in Table 7.7. Clearly, the quality of the crossing estimator has improved by increasing the crossing size. However, we note from a certain size (here $5h$), the median of the activities does not perform well. On the other hand, the mean gives a better result as the crossing size increases. Note that we cannot increase the size as much as we would like, but just enough to avoid the noisy part of the time series.

Crossing size	$\delta t_Q = 5$ min	$\delta t_Q = 30$ min	$\delta t_Q = 60$ min
$2h$	0.3419/0.3675	0.3481/0.3653	0.3347/0.3295
$3h$	0.3145/0.3289	0.3196/0.3405	0.317/0.2998
$4h$	0.2827/0.3019	0.279/0.2846	0.2852/0.2721
$5h$	0.2573/0.3228	0.2459/0.3123	0.2368/0.2717
$6h$	0.2586/0.3877	0.256/0.3293	0.2393/0.3116
$7h$	0.2541/0.4649	0.254/0.3951	0.2323/0.3526

Table 7.7: Crossing estimator quality for different crossing size applied to the EUR/USD rate index.

7.3.2 Volatility of the time-changed data

We estimate the volatility of the data before and after time change. The volatility $v_{[t,t+\delta t)}$ of the time series at time between time t and $t + \delta t$, is defined by

$$v_{[t,t+\delta t)} = \frac{1}{N} \sum_{k=0}^{N-1} |\Delta X_{t+k\tau,\delta t}| \tag{7.8}$$

where $\tau = 1$ week and N is the number of weeks. The results are show in Figures 7.12 and 7.13. Clearly we see that the volatility looks homogenous for the deseasonialised data in each case.

7.3.3 Modelling the activity: Case of the crossing method

Here we carry out the same study as the one presented in Section 7.2.4. We use equation (7.7) to model the activity. The results are shown in Figures 7.14, 7.15, 7.16 and 7.17. Note from the bar Figures that the foreign exchange markets are less active on Mondays. We also note a strong similarity between the GBP/USD and EUR/USD rates index weekly activity. The weekly activity are almost identical. Both currency belong to the European markets. Concerning the daily activity, we note a flat and low activity between time 0 and 4 in GMT+3[5] hour, for the GBP/USD and EUR/USD. In comparison the activity is curvy and high for the AUD/USD and JPY/USD in this period. The Asian market is open at that time, which explains why Asian currencies are traded more than European currencies between time 0 and 4 in GMT+3.

[5]We decided to use GMT+3 time so one can have a null weighted factor on Sunday and Saturday for the activity modal given by (7.7).

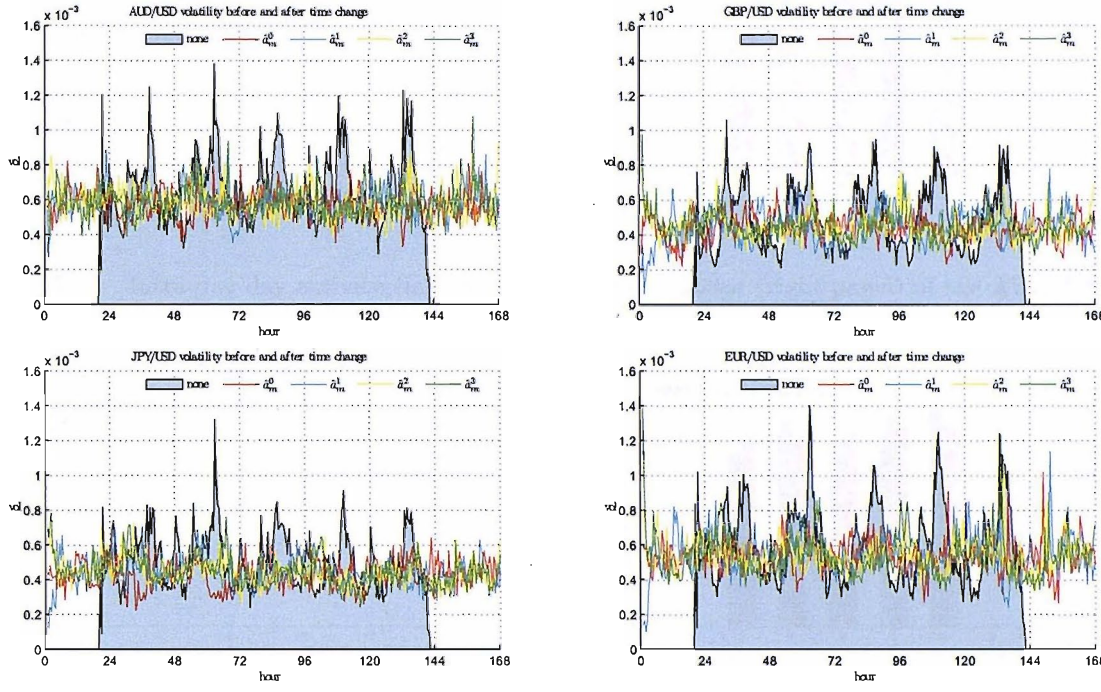

Figure 7.12: Volatility before and after time change of the AUD/USD (top left),GBP/USD (top right), JPY/USD (bottom left) and EUR/USD (bottom right), by taking the mean of the activities.

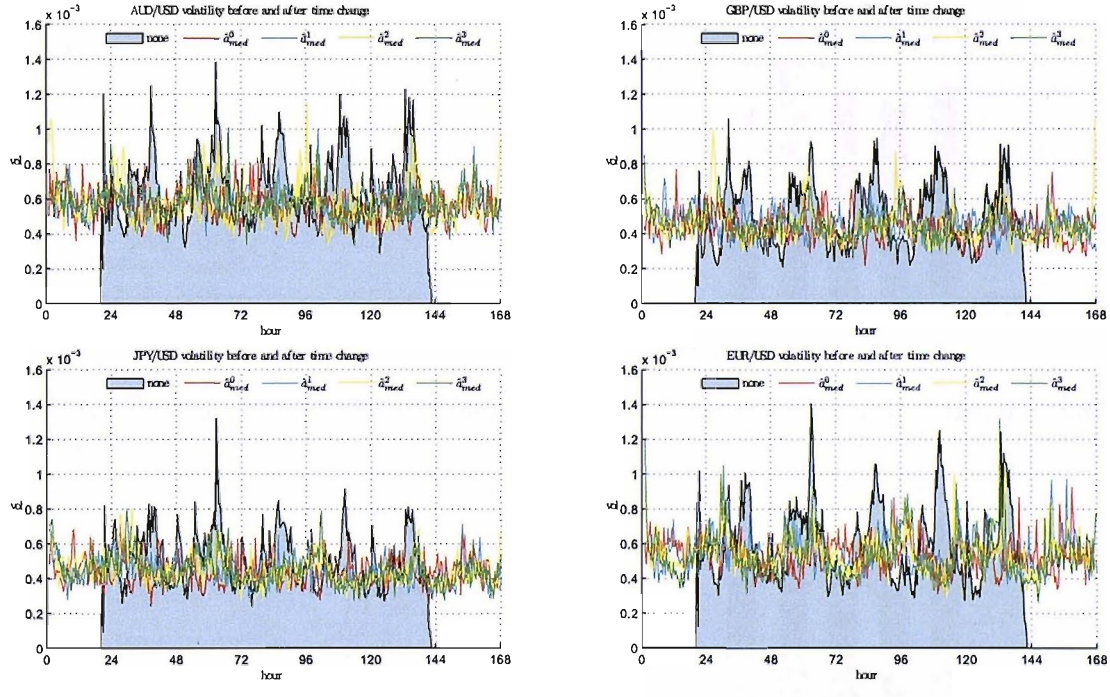

Figure 7.13: Volatility before and after time change of the AUD/USD (top left),GBP/USD (top right), JPY/USD (bottom left) and EUR/USD (bottom right), by taking the median of the activities.

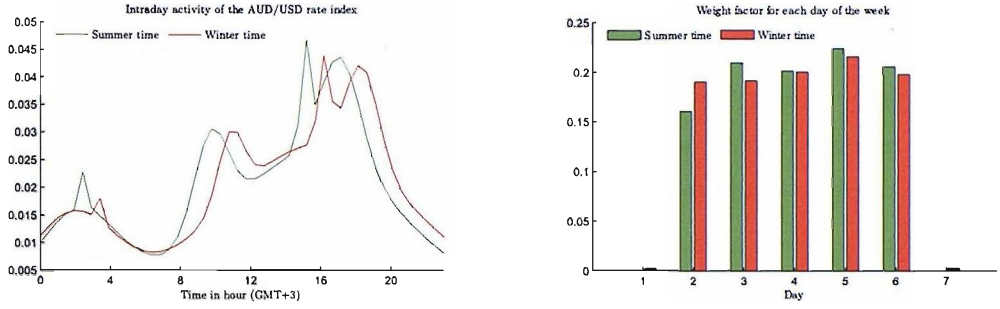

Figure 7.14: Intra-day day activity (left panel) and the day factor (right panel) of the AUD/USD rate index

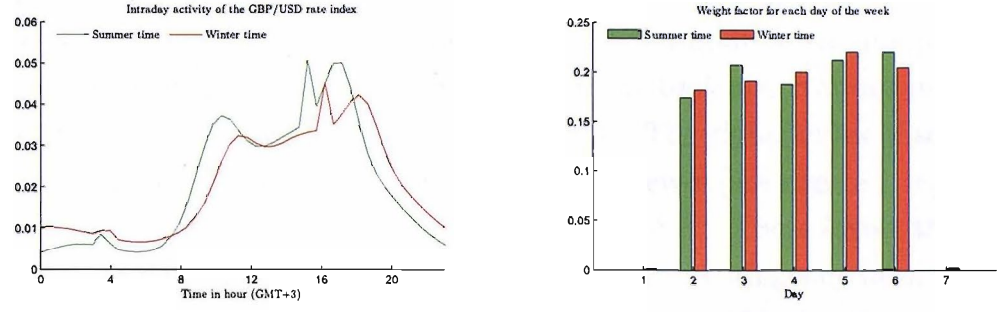

Figure 7.15: Intra-day day activity (left panel) and the day factor (right panel) of the GBP/USD rate index

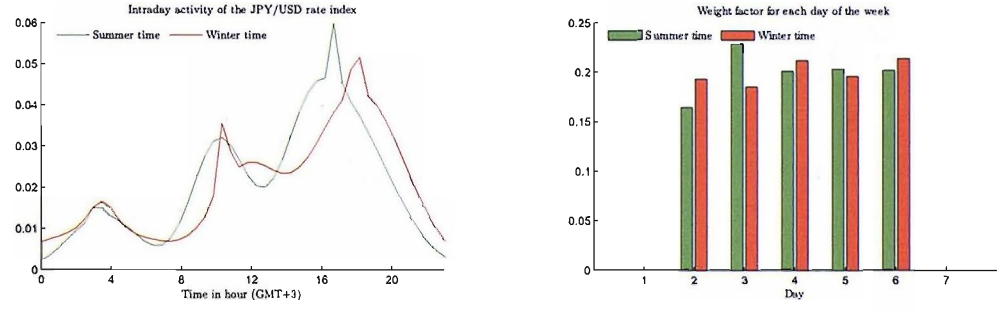

Figure 7.16: Intra-day day activity (left panel) and the day factor (right panel) of the JPY/USD rate index

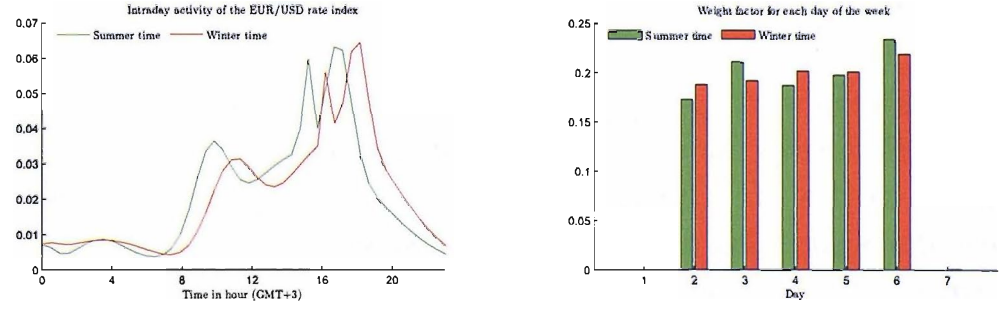

Figure 7.17: Intra-day day activity (left panel) and the day factor (right panel) of the EUR/USD rate index

7.4 Activity estimates of the FTSE100 futures

In this section, we study the FTSE100 futures prices (December 1998 contract) dating from the 10^{th} of October 1998 to the 12^{th} December 1998. In analysing this data set, we made no initial assumptions about periodicity of the activity. Instead we made local activity estimates for the entire series and then examined these estimates for evidence of daily and weekly periodicity.

7.4.1 Activity estimation

We measure time in hours and take $\delta t = 1$. We take as our observations $X(t) = \log P(t)$ where $P(t)$ is the price at time t of the FTSE100. The crossings method was used to estimate the activity, with h equal to the standard deviation of the jumps of X. The path variation were used with $n = 8$, which means we took observations in the interval δt separated by 14.0625 second (1 hour dived by 2^8). The global index of self-similarity H of X is estimated by the EBP method for different levels (see Figure 7.18). The green circles are the levels taken to compute an estimate of H. These levels are chosen by performing a test of self-similarity over the levels using a contingency table as described in [22]. To test for daily and weekly periodicity, we put $\tau = 168$, then for $j = 0, \ldots \tau - 1$,

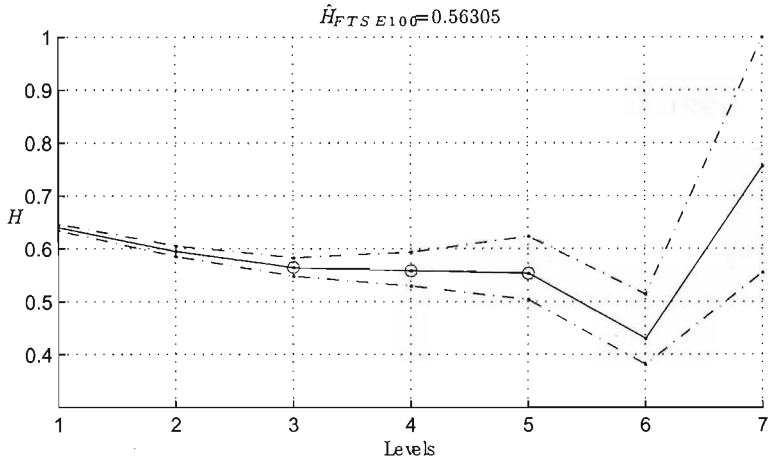

Figure 7.18: Global index of self-similarity estimation of the $X_t = \log(P_t)$ using the crossing tree method.

we grouped the estimates $\hat{a}^i_{[j,j+1)}$, $\hat{a}^i_{[j+\tau,j+1+\tau)}$, $\hat{a}^i_{[j+2\tau,j+1+2\tau)}$, ... where $i \in \{1, 2, 3\}$, and drew a box plot of these values against j. The results are given in Figure 7.19 and show strong evidence for daily and weekly seasonality in the activity for the three estimators. Because there are some outliers present, to get a single estimate of the activity at each hour j of the week, we took the median value of the corresponding group of activity estimates. Given our median weekly activity a, we time-changed X to get the deseasonalised series $\tilde{X}(t) = X(\theta^{-1}(t))$, where $\theta(t) = \int_0^t a(u)du$. The original and $\hat{a}^3$-deseasonalised series are plotted in Figure 7.20. To confirm that we had correctly

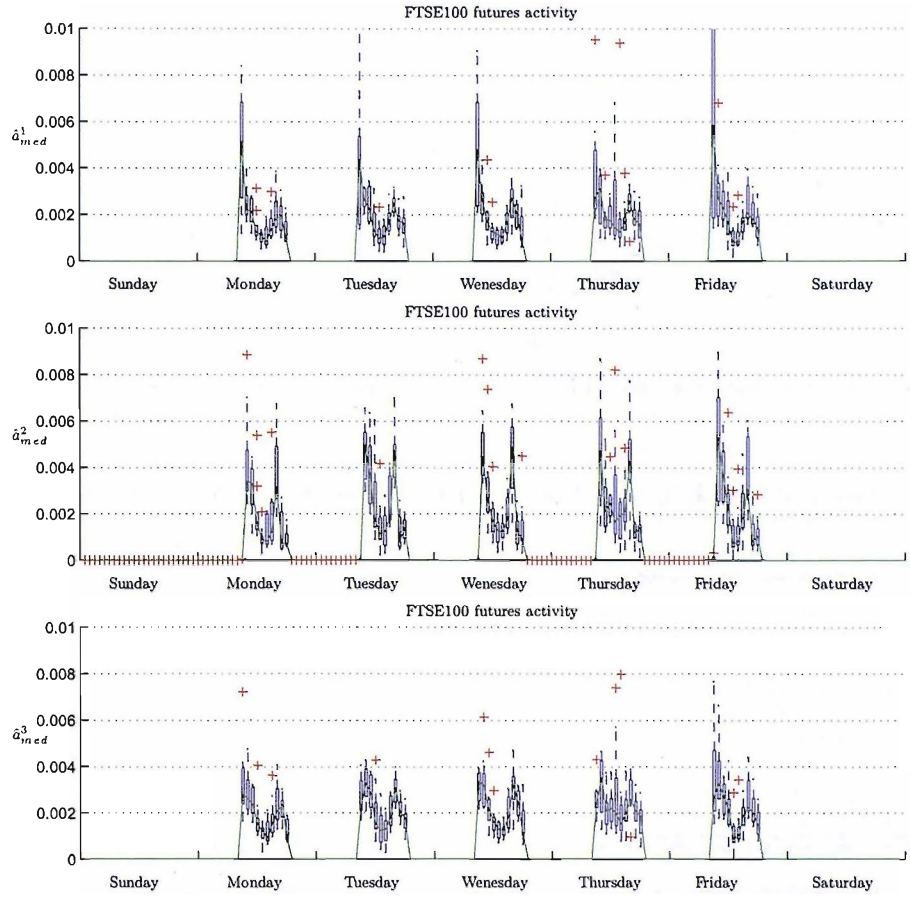

Figure 7.19: Evidence of daily and weakly seasonality in the activity of FTSE100 futures prices. For each half-hour of the week a boxplot of estimated activities is given (taken from different weeks). The median activity has been plotted on top.

captured the periodicity in the activity, we calculated the power spectrum of X and $\tilde{X}$. The results are given in Figure 7.21 and show no discernable remaining daily or weekly periodic effects, for all our estimators. We check if the time-changed series is self-similar. We use (7.5) across all the series. The results in the case of the 1/H-variation method and the crossing are given in Figure 7.22. We perform a least squares linear regression through the points $\left\{\log(\delta t), \log \mathbb{V}ar\left(\tilde{X}(\delta t)\right)\right\}$ (green line on Figure 7.22) and compute the slope of the line to get an estimate of the self-similarity index. The time scale used δt is in $\{1, 2, 5, 10, 15, 20, 30, 45, 60, 120, 240, 360, 720, 1440, 2880\}$ in minutes. The red line represents the polynomial of second order fitting points $\left\{\log(\delta t), \log \mathbb{V}ar\left(\tilde{X}(\delta t)\right)\right\}$. The red line is not straight. This shows that the self-similar index changes through scales. It is difficult to confirm that the time-changed series is self-similar for the given range of δt.

7.4.2 Daily and weekly seasonal analysis

To separate the daily and weekly seasonal effects, we fitted a multiplicative model a^M to our activity estimates $\hat{a}^i_{[j,j+1)}$. Write $k \in \{0, \ldots, 167\}$ as $24(d-1)+j-1$ where

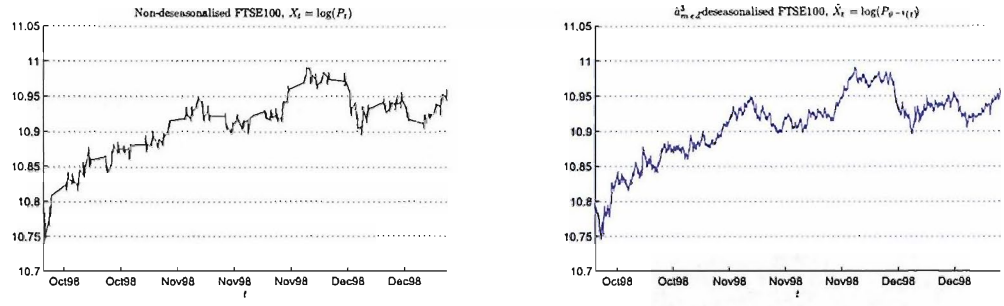

Figure 7.20: FTSE100 futures prices before (left panel) and after (right panel) removal of the periodic variation in activity.

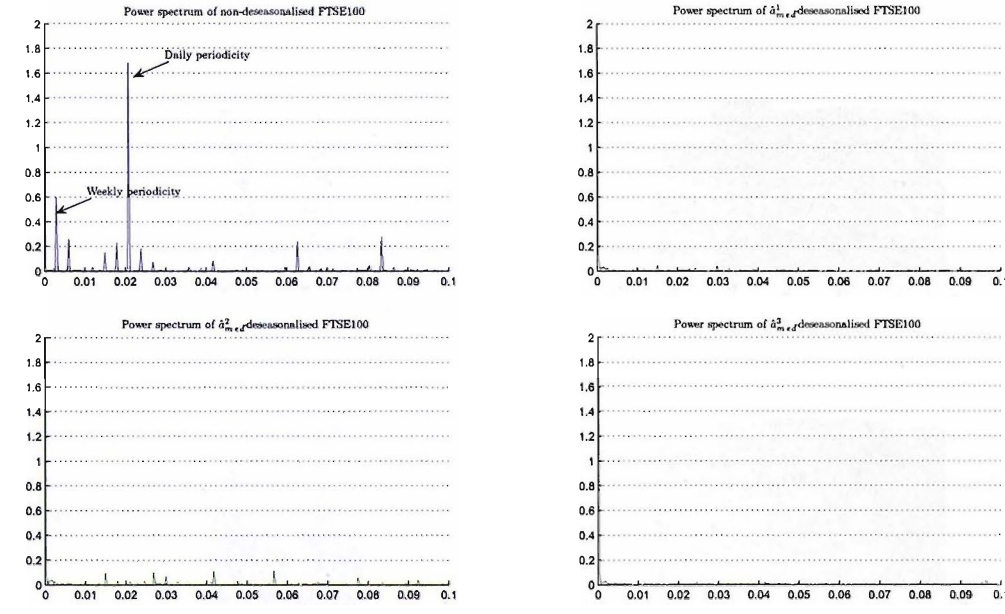

Figure 7.21: Power spectrum of the FTSE100 futures prices before (left) and after (right) removal of the periodic variation in activity.

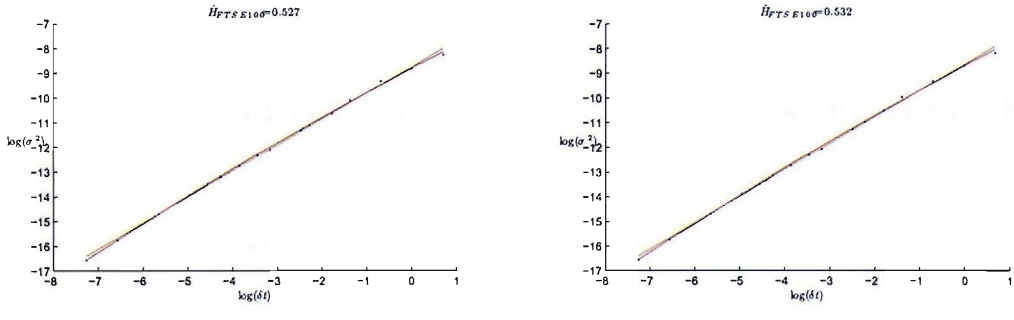

Figure 7.22: Test of self-similarity of the time-changed FTSE100 using the $1/H$-variation (left panel) and the crossing (right panel).

d is the day and j the hour, then suppose that for some α_j, $j = 1, \ldots, 24$, and β_d, $d = 1, \ldots, 7$, we have $a^M_{[k,k+1)} = a^M_{[24(d-1)+j-1,24(d-1)+j)} = \alpha_j \beta_d$.
For $k \geq 168$ let $k' = k \mod 168$ then we have $a^M_{[k,k+1)} = a^M_{[k',k'+1)}$. The parameters α_j and β_d were estimated using least squares and the results are given in Figure 7.23. The characteristic "U shape" of the intra-day activity has been previously observed by other authors, for example ap Gwilym and Sutcliffe [91].

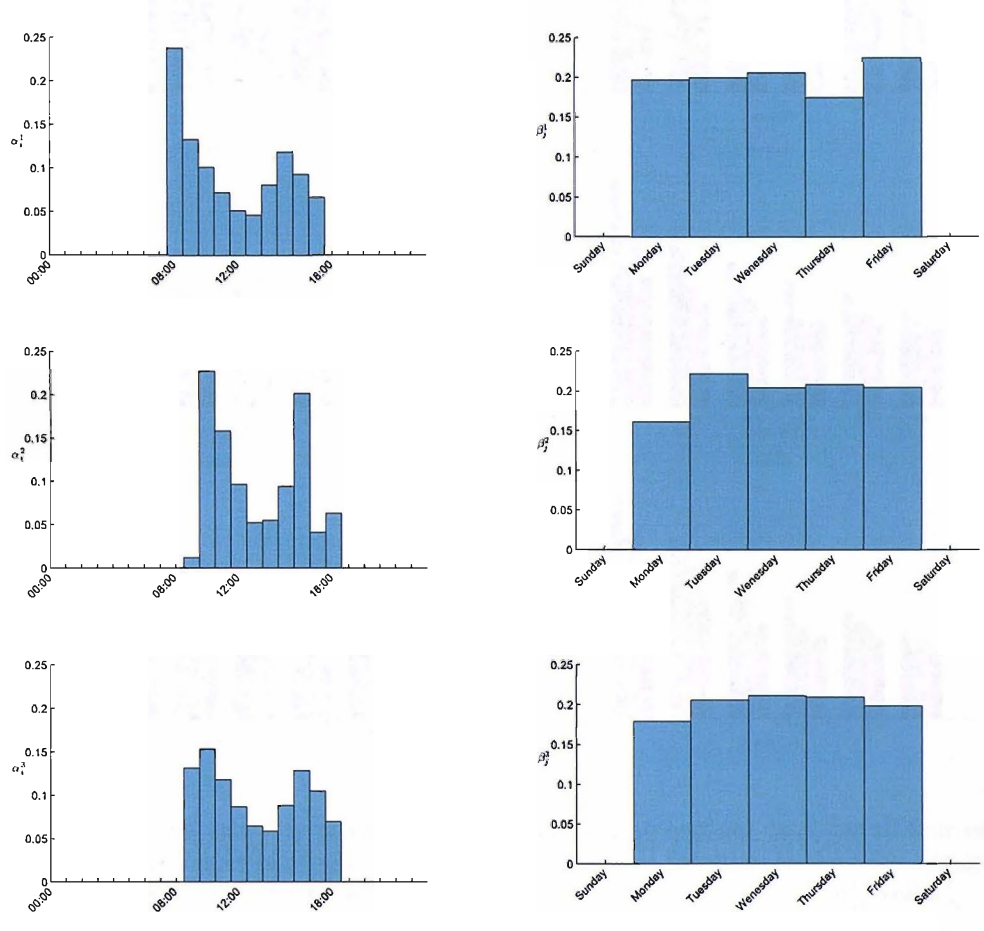

Figure 7.23: Daily and weekly periodic effects in the activity of FTSE100 futures prices.

7.5 Comment on the path variation and crossing time change estimators

7.5.1 The advantage of our estimators

The advantage of our time changed estimators in comparison to the scaling law method, is that we obtain a good activity estimate through using just a few weeks of data. In comparison, the scaling law needs several weeks of financial data in order to give a good estimate of the given asset activity. In this Section, we compute the quality of the estimator by taking a small number of weeks and applying it to time change the

whole dataset. This study will be carried on the AUD/USD rate index as an example.

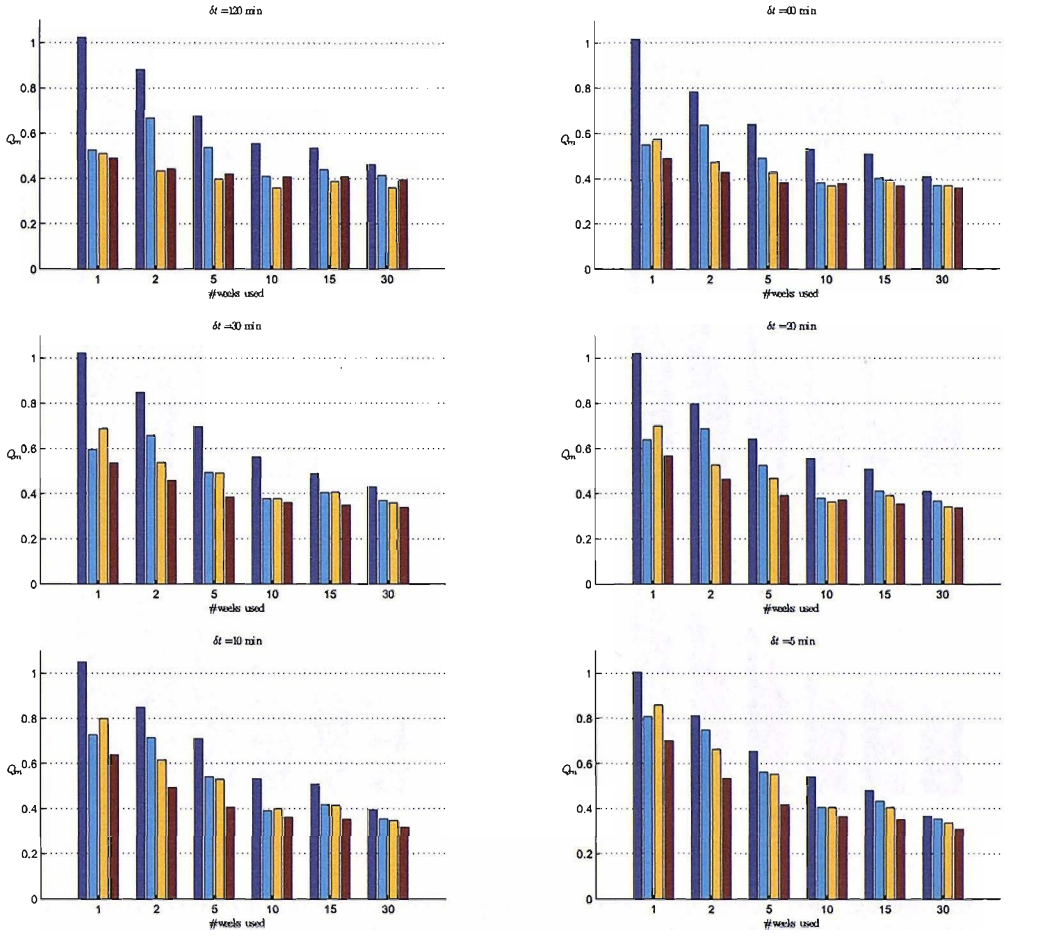

Figure 7.24: The relative volatility of time-changed compared to original data, for all four estimators and for various values of δt and the amount of data used to estimate the time change (in weeks). In each cluster of results the estimators are given in the order $\hat{a}^0_m$, $\hat{a}^1_m$, $\hat{a}^2_m$ and $\hat{a}^3_m$.

We used Q to compare the activity estimators $\hat{a}^i$ (where $i \in \{0, 1, 2, 3\}$ corresponding to the estimators described in 7.1). Thirty weeks of the AUD-USD exchange rate data was chosen from the summer time period. For each estimator, we varied the number of weeks of data used to fit it and the size of δt used when calculating $\hat{a}_{[t,t+\delta t)}$. The amount of data used was either 1, 2, 5, 10, 15 or 30 weeks; the size of δt was either 120, 60, 30, 20, 10 or 5 minutes. Each activity estimate calculated was used to time-change all 30 weeks of data and then calculate Q (using $\delta t = 5$ minutes). The results of these calculations are given in Figure 7.24 when taking the mean of the activities and Figure 7.25 when taking the median of the activities. Within the Figure each panel deals with a single value of δt. Within each panel, the results for the four estimators are grouped together for each of the different amounts of data used. The following conclusions can be drawn: for $\delta t = 60$, 30, 20, 10 or 5 minutes the crossings estimator performs the best; while in all cases except for $\delta t = 5$ minutes the estimator of Dacorogna et al has

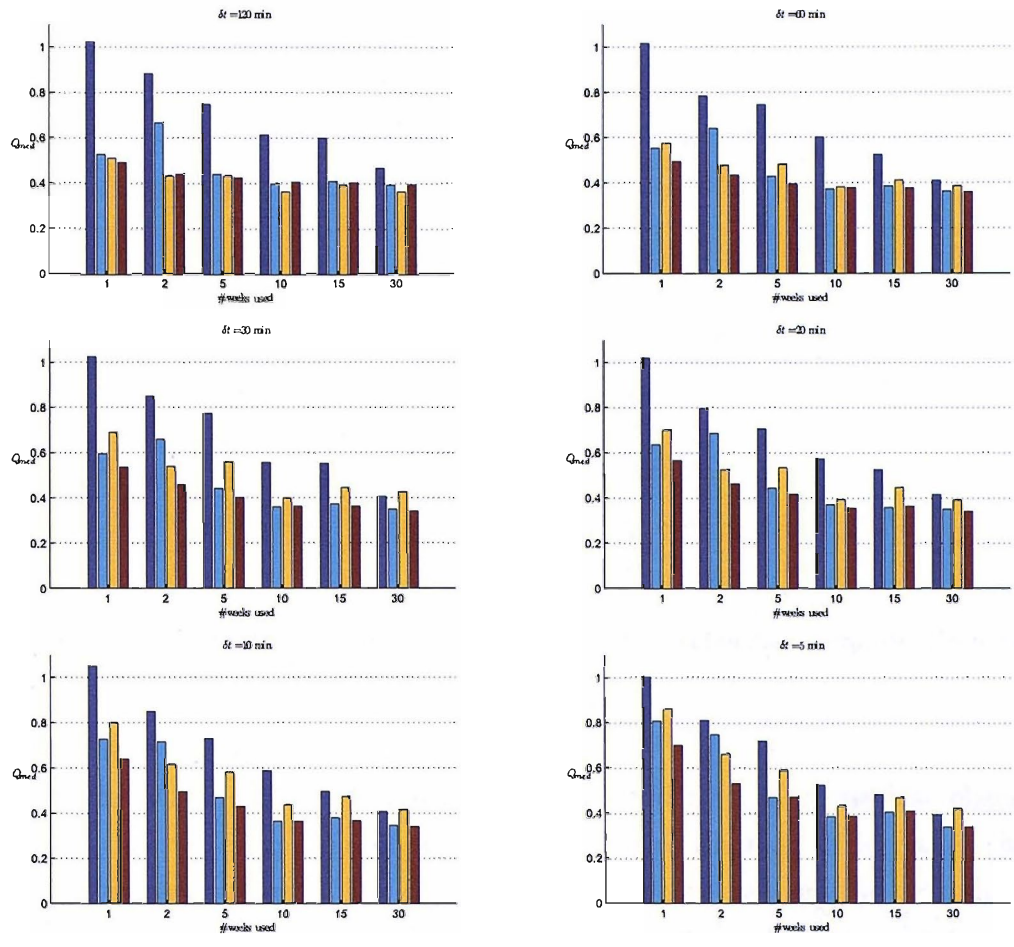

Figure 7.25: The relative volatility of time-changed compared to original data, for all four estimators and for various values of δt and the amount of data used to estimate the time change (in weeks). In each cluster of results the estimators are given in the order $\hat{a}^0_{med}$, $\hat{a}^1_{med}$, $\hat{a}^2_{med}$ and $\hat{a}^3_{med}$.

the worst performance. For $\delta t = 120$ minutes the log-variation estimator performs best; while the $1/H$-variation estimator provided a better performance than the log-variation when taking the median of the actvities.

7.5.2 How to use the estimators?

To conclude, we present a brief explanation on how to use the different time change estimators. In each case, we explain how to deal with real data as this last has irregularly spaced observations in time. First: how to interpolate the observed data in order to perform a good estimate of the time change function. Second: once the activity has been estimated, should we use the mean or the median of the weekly activity. Assuming we have observations of our financial data $\{X(t_j)\}_{\{j\in\{1,\dots,n\}\}}$, we define

$$X_l(t) = \sum_{j=0}^{n} X(t_j) + (t - t_j)\frac{X(t_{j+1}) - X(t_j)}{t_{j+1} - t_j}\mathbb{I}_{t\in[t_j,t_{j+1})} \tag{7.9}$$

$$X_p(t) = \sum_{j=0}^{n} X(t_j)\mathbb{I}_{t\in[t_j,t_{j+1})} \tag{7.10}$$

The process X_l and X_p are represented in Figure 7.26 and will facilitate to describe the type of interpolation needed to perform our estimators.

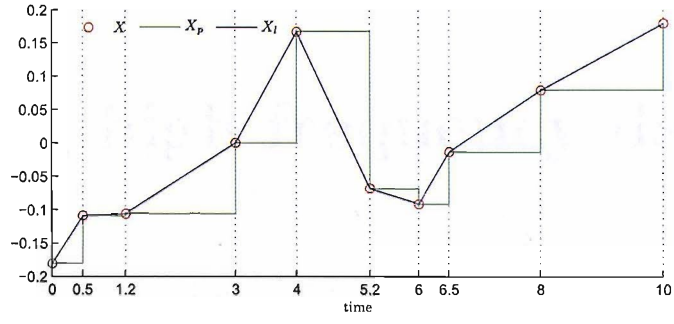

Figure 7.26: An example of the process X_p (green line) and X_l (blue line) using the observation X (red circle).

Let us comment on each estimator:

The $1/H$-variation: The data must be interpolated by taking the last observation. This means that we estimate θ of X using X_p. For continuous data, either the mean or the median can be used. For data containing jumps, we may have spikes in the activity at the corresponding time jump; this may affect the de-seasonalisation quality. In that case, the median of the activities over weeks are considered. This estimator performs well on foreign exchange data, as these last are considered not to having important changes in price when taking two consecutive quotes. This comes from the high liquidity of the foreign exchange market.

The log-variation: The observed data must be linearly interpolated. This means that we estimate the time-changed function θ of X using X_l. However, we have to make sure that there are no two consecutive identical prices in the data; this will generate increments of value 0. Assuming there is activity between time t and $t + \delta t$, an increment of value 0 will imply that the sum used for this estimator will be equal to $-\infty$, and so the exponential will be 0. This is similar to saying that the activity will be considered null during that period. It suffices to remove successive identical prices. The estimator can be applied for data whether it contains jumps or not. The mean or the median of the activities over weeks can be used.

The crossing estimator: This estimator is very sensitive to outliers; one must ensure that the dataset used does not contain any outliers. The crossings are taken from jumps of the observed data (from X_p) in order to avoid crossings during inactive periods, such as week-ends or bank holidays. The overall activity is estimated by taking the mean over weeks. As noted in this section, this last method seems to give a better quality of the activity estimate.

Chapter 8

Modelling high frequency data

Several techniques exist for modeling high frequency financial data. Generally these models can be grouped into four classes:

(1) Model based on price change

(2) Model based on duration between two ticks

(3) A bivariate model based on both price change and price duration

(4) Modelling price change using a time deformation, also know as subordinated process

For a survey on (1)-(3) type model, we refer to [81]. Models based on price duration are especially adapted to tick by tick data, where the waiting time between two transaction is random. We refer to [84] for a good study of such models. The type of model (4) is of our interest. This model is based on mapping the calendar time t to an operational time $\theta(t)$, with respect to which the data has constant volatility. This model was proposed by Mandelbrot [104] and Clark [105] for stochastic time change and then by Dacorogna et al. in [3] for deterministic case, where the operational time $\theta(t)$ is called the "θ-time". The function θ is determined from the scaling property of the high frequency financial data. This was described in the first Section of the previous Chapter. Later on, Chysels and Jasiek [106] proposed a time deformation determined by the trading volume, and Zhou [93] make use of the volatility to determinate θ.
In our case, the operational time θ is determined so that the data has stationary increments under θ (see previous Chapter). The function θ is estimated through one of the following methods: $1/H$-variation ideal for continuous time processes satisfying some scaling properties, log-variation adapted to processes with jumps satisfying also some scaling properties, and the level crossings method.
In this Chapter, we study the statistical properties of the $\hat{a}^3_m$-deseasonalised AUD/USD index rate, where $\hat{a}^3_m$ was estimated in the previous Chapter using $dt = 30$ minutes, and the $\hat{a}^2_{med}$-deseasonalised FTSE100 future contract where $\hat{a}^2_{med}$ was estimated using $dt = 60$. In particular the following will be analysed: the price duration before and

after removing seasonality, the autocorrelation of the log price return and the absolute return, the distribution of the log price return.
For clarity, we note $DS_{a_m^i}$-data for a_m^i-deseasonalised data, where $i = 0, \ldots, 3$ depending on the method used (see previous Chapter); e.g. $DS_{a_m^i}$-AUD/USD for a_m^i-deseasonalised AUD/USD. Similarly, we note $DS_{a_{med}^i}$-data for a_{med}^i-deseasonalised.

8.1 Analysis of the deseasonalised AUD/USD index rate

8.1.1 Time duration and data sampling analysis

We define the time duration as the time separating two consecutive ticks. As mentioned in Chapter 6, the time duration for ultra high frequency financial data are irregularly spaced. The deseasonalised data will also be irregularly spaced in time. Here we analyse the distribution of the time increments before and after time-change of the AUD/USD rates index.

8.1.1.1 Time duration analysis

We denote by τ the time duration between two ticks. The distribution of τ for the non deseasonalised and for deseasonalised data for each method ($\hat{a}_m^0$, $\hat{a}_{med}^1$, $\hat{a}_m^2$, $\hat{a}_m^3$) are given in Table 8.1. Note for deseasonalised AUD/USD rate index, some waiting times τ can still be greater than four hours. This is the minimum[1] time duration that can exist when one of the foreign exchange Markets[2] has bank holidays. For this reason, we omit in non-deseasonalised data time duration greater than 4 hours, and we remove the corresponding time duration from $DS_{\hat{a}_m^i}$-AUD/USD. The results are given in Table 8.2. This practice is needed before sampling the data to obtain regular time spaced data. Indeed, if we do not remove bank holidays, the deseasonalised data contains many returns with value 0 and so fails the statistical analysis of the data.
We investigate on the distribution of τ at smaller scales. These are given in Table 8.3.

The time deformation consists on stretching periods of high volatility and contracting periods of low volatility. This is observable in Table 8.3. The frequency of the time duration is lower for $DS_{\hat{a}_m^i}$-AUD/USD ($i = 0, \ldots, 3$) than the AUD/USD when $\tau <$ 1 minute. It is also observable in Table 8.4, the standard deviation of the $DS_{\hat{a}_m^i}$-AUD/USD price duration, where holidays were previously removed, are more than 19 times[3] less (except for the scaling law method, which is more than 16 times) than the one of the non-deseasonalised data where companies bank holidays and week-ends are included.

[1] in fact 3,5hours
[2] This correspond to the American market in non-DST period (see Figure F.5)
[3] The scaling error due to companies bank holidays omission was taken into consideration and is $\simeq 0.777/0.0754 \simeq 1.03$

(in Hour)	$\tau < 4$	$4 \leq \tau < 12$	$12 \leq \tau < 24$	$24 \leq \tau$
AUD/USD	674260	5	1	55
$DS_{\hat{a}_m^0}$-AUD/USD	674309	5	3	4
$DS_{\hat{a}_{med}^1}$-AUD/USD	674309	5	2	5
$DS_{\hat{a}_m^2}$-AUD/USD	674312	2	2	5
$DS_{\hat{a}_m^3}$-AUD/USD	674309	5	2	5

Table 8.1: Time duration (in hour) distribution of the AUD/USD rate index price change before removing susceptible possible bank holidays.

(in Hour)	$\tau < 4$	$4 \leq \tau < 12$	$12 \leq \tau < 24$	$24 \leq \tau$
AUD/USD	674321	0	0	0
$DS_{\hat{a}_m^0}$-AUD/USD	674319	2	0	0
$DS_{\hat{a}_{med}^1}$-AUD/USD	674320	1	0	0
$DS_{\hat{a}_m^2}$-AUD/USD	674320	1	0	0
$DS_{\hat{a}_m^3}$-AUD/USD	674320	1	0	0

Table 8.2: Time duration (in hour) distribution of the AUD/USD rate index price change after removing susceptible possible bank holidays.

(in minute)	$\tau < 1$	$1 \leq \tau < 5$	$5 \leq \tau < 10$	$10 \leq \tau < 60$	$60 \leq \tau$
AUD/USD	594333	74300	4433	1236	19
$DS_{\hat{a}_m^0}$-AUD/USD	528524	137292	6905	1572	28
$DS_{\hat{a}_{med}^1}$-AUD/USD	517201	151135	5046	919	20
$DS_{\hat{a}_m^2}$-AUD/USD	515320	153671	4707	604	19
$DS_{\hat{a}_m^3}$-AUD/USD	517231	151139	5045	887	19

Table 8.3: Time duration (in minute) distribution of the AUD/USD rate index price change after removing possible bank holidays.

	AUD/USD	$DS_{\hat{a}_m^0}$	$DS_{\hat{a}_{med}^1}$	$DS_{\hat{a}_m^2}$	$DS_{\hat{a}_m^3}$
Including Bank Holidays					
$\mathbb{E}[\tau]$	0.77758	0.77757	0.77753	0.77751	0.77756
$\sqrt{\mathbb{V}ar(\tau)}$	25.9047	7.4557	7.626	8.0017	7.7393
Bank Holidays not included					
$\mathbb{E}[\tau]$	-	0.75491	0.75365	0.75393	0.75347
$\sqrt{\mathbb{V}ar(\tau)}$	-	1.5342	1.3022	1.3238	1.2702

Table 8.4: Time duration (in minute) average and standard deviation of the AUD/USD rate index price change before and after removing possible bank holidays.

8.1.1.2 Data sampling

In order to analyse the data with observations taken at regular time intervals, we sample the data. Sampling a dataset may introduce some error in its statistical analysis, such as a bias in the autocorrelation function estimate, for example. This error comes from the approximation caused by the interpolation. From Table 8.3, we note that less than 1% of time duration are more than 5 minutes for the deseasonalised data, except for the $DS_{\hat{a}^0_m}$ for which the time duration above 5 minute represents 1.26%. We may expect taking observations at a minimum step time of 5 minutes, which can be considered reasonable. Assuming the data are continuous in time, one can use a linear interpolation of the data to get regularly spaced sample from the observed one. This choice was suggested by Müller in [97]. Wasserfallen and Zimmerman [107] used the last published price as interpolation. This approach produces many returns with value zero and jumps in the data. The error of interpolation is even more pronounced when the path of the data is extremely rough. To justify the choice of the linear interpolation for our study, we analyse the error of the log price return in both types of interpolation. Let $X(t) = \log(P(t))$ the logarithmic price of $DS_{\hat{a}^3_m}$-AUD/USD. Let $\breve{X}(t)$ and $\check{X}(t)$ the log price at time t respectively taken by linear and last price interpolation between two consecutive tick times t_i and t_{i+1}. If we assume X H-sssi with finite second moment and self-similarity index H, which can be seen as the coefficient of roughness of the data, the linear interpolation error is given by (see Proposition 3.2.2)

$$\mathbb{P}\left(\left|\breve{X}(t) - X(t)\right| \geq (t_{i+1} - t_i)^H \sqrt{\mathbb{E}\left|X(1)\right|^2}\right) \leq \Lambda^H\left(\frac{t - t_i}{t_{i+1} - t_i}\right) \tag{8.1}$$

Where

$$\forall \lambda \in [0,1], \quad \Lambda^H(\lambda) = \begin{cases} 0 & \text{if } \lambda \in \{0,1\} \\ \lambda(1-\lambda)\left(\lambda^{2H-1} + (1-\lambda)^{2H-1} - 1\right) & \text{if } \lambda \in (0,1) \end{cases} \tag{8.2}$$

and by the Markov inequality, the last price interpolation error is given by

$$\mathbb{P}\left(\left|\check{X}(t) - X(t)\right| \geq (t_{i+1} - t_i)^H \sqrt{\mathbb{E}\left|X(1)\right|^2}\right) \leq \left(\frac{t - t_i}{t_{i+1} - t_i}\right)^{2H} \tag{8.3}$$

We represent functions $\lambda \mapsto \Lambda^H(\lambda)$ and $\lambda \mapsto \lambda^{2H}$ in Figure 8.1. Clearly the linear interpolation is a better choice for sampling our data.

The question remaining is, what is the smallest discrete time we should use to sample the data so one can reduce the bias in statistical analysis?

Let $R(t, \delta t) = X(t + \delta t) - X(t)$ and $\breve{R}(t, \delta t) = \breve{X}(t + \delta t) - \breve{X}(t)$, for $\delta t > 0$ which represent the step time used for the interpolation. We define the mean square error of the return taken from the linear interpolated data by $\eta_{\delta t}(t)^2$. In Proposition 8.1.1, we compute $\eta_{\delta t}(t)^2$ assuming the deseasonalised data is self-similar. We consider the time

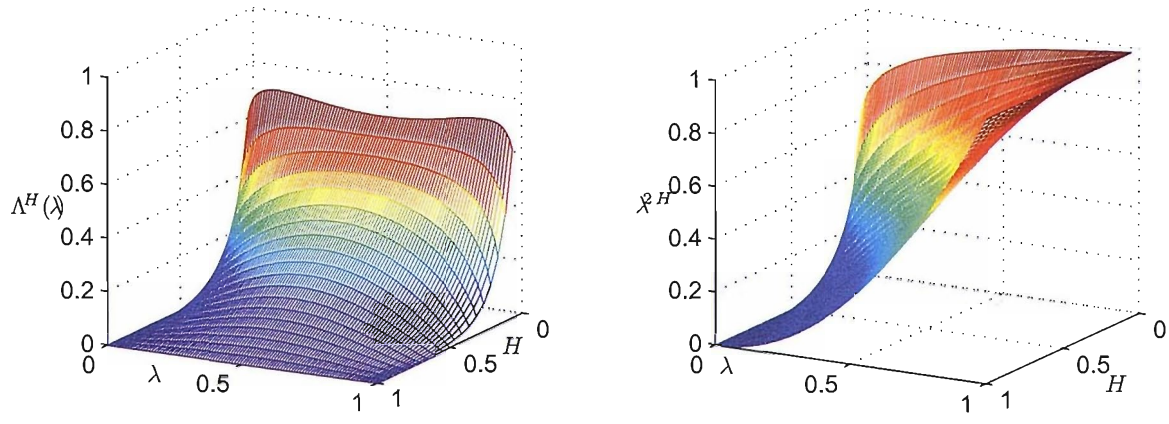

Figure 8.1: Functions $\lambda \mapsto \Lambda^H(\lambda)$ (left panel) and $\lambda \mapsto \lambda^{2H}$ (right panel)

duration between two consecutive prices as random.

Proposition 8.1.1. *Let $X = \{X(t)\}_{t\geq 0}$ be a H-sssi process such that $\mathbb{E}[X(1)]^2 < +\infty$ and $\mathbb{E}[X(1)] = 0$. We suppose having the observation of X at given random times $T = \{T_i\}_{i\in\mathbb{N}}$, so that the process T is strictly increasing and independent of X. We set*

$$\mathcal{F}_t = \sigma\left(\{T_u, u \leq i_t + 1\}\right)$$

Let $\breve{X}$ a process constructed from $\{X(T_i)\}_{i\in\mathbb{N}}$ by linear interpolation such that for $t \geq 0$, $T_{i_t} \leq t < T_{i_t+1}$, $\breve{X}(t) = (1-\lambda_t)X(T_{i_t+1}) + \lambda_t X(T_{i_t})$, where $\lambda_t = \dfrac{t - T_{i_t}}{T_{i_t+1} - T_{i_t}}$. For $t > s$, we set

$$R(t, t-s) = X(t) - X(s) \text{ and } \breve{R}(t, t-s) = \breve{X}(t) - \breve{X}(s)$$

and

$$\eta_t^2 = \frac{\mathbb{E}\left[\left(\breve{R}(t,t-s) - R(t,t-s)\right)^2 |\mathcal{F}_t\right]}{\mathbb{E}\left|R(t,t-s)\right|^2}$$

Then

$$\eta_t^2 = \frac{\Lambda^H(\lambda_t)(T_{i_t+1} - T_{i_t})^{2H} + \Lambda^H(\lambda_s)(T_{i_s+1} - T_{i_s})^{2H} - 2\Xi^H(t, t-s)}{(t-s)^{2H}} \tag{8.4}$$

Where Λ^H is defined as in (8.2) and $\Xi^H(t,s)$ is a function of s, t and H.

Proof. Note first that by self-similarity and stationarity of the increments of X,

$$\mathbb{E}\left|R(t,t-s)\right|^2 = (t-s)^{2H}\mathbb{E}\left|X(1))\right|^2$$

One has

$$\begin{aligned}\mathbb{E}\left[|\breve{R}(t,t-s)-R(t,t-s)|^2|\mathcal{F}_t\right] &= \mathbb{E}\left[|(\breve{X}(t)-\breve{X}(s))-(X(t)-X(s))|^2|\mathcal{F}_t\right]\\ &= \mathbb{E}\left[|(\breve{X}(t)-X(t))-(\breve{X}(s)-X(s))|^2|\mathcal{F}_t\right]\\ &= \mathbb{E}\left[|\breve{X}(t)-X(t)|^2|\mathcal{F}_t\right]+\mathbb{E}\left[|\breve{X}(s)-X(s)|^2|\mathcal{F}_t\right]\\ &\quad -2\mathbb{E}\left[(\breve{X}(t)-X(t))(\breve{X}(s)-X(s))|\mathcal{F}_t\right]\end{aligned}$$

The two first terms are given by Proposition 3.2.2:

$$\mathbb{E}\left[\left|\breve{X}(t)-X(t)\right|^2|\mathcal{F}_t\right] = \Lambda^H(\lambda_t)(T_{i_t+1}-T_{i_t})^{2H}\mathbb{E}\left|X(1)\right|^2$$

where the $\lambda_t = \dfrac{t-T_{i_t}}{T_{i_t+1}-T_{i_t}}$. Since we use a linear interpolation of the process X, one gets $\breve{X}(t)-X(t) = \lambda_t(X(T_{i_t+1})-X(t)) + (1-\lambda_t)(X(T_{i_t})-X(t))$ and

$$\begin{aligned}&\mathbb{E}\left[(\breve{X}(s)-X(s))(\breve{X}(t)-X(t))|\mathcal{F}_t\right]\\ &\quad= \lambda_t\lambda_s\mathbb{E}\left[(X(T_{i_s+1})-X(s))(X(T_{i_t+1})-X(t))|\mathcal{F}_t\right]\\ &\quad+ \lambda_s(1-\lambda_t)\mathbb{E}\left[(X(T_{i_s+1})-X(s))(X(T_{i_t})-X(t))|\mathcal{F}_t\right]\\ &\quad+ \lambda_t(1-\lambda_s)\mathbb{E}\left[(X(T_{i_s})-X(s))(X(T_{i_t+1})-X(t))|\mathcal{F}_t\right]\\ &\quad+ (1-\lambda_s)(1-\lambda_t)\mathbb{E}\left[(X(T_{i_s})-X(s))(X(T_{i_t})-X(t))|\mathcal{F}_t\right]\end{aligned} \tag{8.5}$$

As X is H-sssi with $\mathbb{E}[X(1)]=0$, we note the correlation function

$$r(t,s) = \frac{\mathbb{E}[X(t)X(s)]}{\mathbb{E}[X(1)^2]} = \frac{1}{2}(t^{2H}+s^{2H}-|t-s|^{2H})$$

Whence expanding (8.5), one gets

$$\begin{aligned}\Xi^H(t,t-s) &= \frac{\mathbb{E}\left[(\breve{X}(s)-X(s))(\breve{X}(t)-X(t))|\mathcal{F}_t\right]}{\mathbb{E}[X(1)^2]}\\ &= \lambda_t\lambda_s\left(r(T_{i_s+1},T_{i_t+1})-r(T_{i_s+1},t)-r(T_{i_t+1},s)+r(s,t)\right)\\ &+ \lambda_s(1-\lambda_t)\left(r(T_{i_s+1},T_{i_t})-r(T_{i_s+1},t)-r(s,T_{i_t})+r(s,t)\right)\\ &+ \lambda_t(1-\lambda_s)\left(r(T_{i_s},T_{i_t+1})-r(T_{i_s},t)-r(s,T_{i_t+1})+r(s,t)\right)\\ &+ (1-\lambda_s)(1-\lambda_t)\left(r(T_{i_s},T_{i_t})-r(T_{i_s},t)-r(s,T_{i_t})+r(s,t)\right)\end{aligned} \tag{8.6}$$

□

For $H=1/2$, the term $\Xi^H(t,t-s)$ can be simplified as follow

Lemma 8.1.2. *Assuming $t > s$, for $H = 1/2$,*

$$\begin{cases} \Xi^{1/2}(t, t-s) = 0 & \text{if } i_s \neq i_t \\ \Xi^{1/2}(t, t-s) = \dfrac{(T_{i+1} - t)(s - T_i)}{T_{i+1} - T_i} & \text{if } i_s = i_t = i \end{cases} \tag{8.7}$$

Proof. For $i_s \neq i_t$ this is obvious, since disjoint increments are uncorrelated the correlation function has values zeros. For $i_s = i_t = i$ and $t > s$, one has from (8.5)

$$\begin{aligned} \mathbb{E}\left[(X(T_{i+1}) - X(s))(X(T_{i+1}) - X(t))|\mathcal{F}_t\right] &= \mathbb{E}\left[|X(T_{i+1}) - X(t)|^2|\mathcal{F}_t\right] \\ &\quad + \mathbb{E}\left[(X(t) - X(s))(X(T_{i+1}) - X(t))|\mathcal{F}_t\right] \\ &= (T_{i+1} - t)\mathbb{E}|X(1)|^2 \end{aligned}$$

Similarly we find

$$\begin{aligned} \mathbb{E}\left[(X(T_{i+1}) - X(s))(X(T_i) - X(t))|\mathcal{F}_t\right] &= -(t-s)\mathbb{E}|X(1)|^2 \\ \mathbb{E}\left[(X(T_{i+1}) - X(t))(X(T_i) - X(s))|\mathcal{F}_t\right] &= 0 \\ \mathbb{E}\left[(X(T_i) - X(s))(X(T_i) - X(t))|\mathcal{F}_t\right] &= (s - T_i)\mathbb{E}|X(1)|^2 \end{aligned}$$

And so, one gets

$$\Xi^{1/2}(t, t-s) = \lambda_{t+\delta t}\lambda_t T_{i+1} - \lambda_s t + (1 - \lambda_t)s + (1 - \lambda_t)(1 - \lambda_{t+\delta t})T_i$$

Substituting $\lambda_t = \dfrac{t - T_i}{T_{i+1} - T_i}$ and $\lambda_s = \dfrac{s - T_i}{T_{i+1} - T_i}$, one has

$$\Xi^{1/2}(t, t-s) = \frac{(T_{i+1} - t)(s - T_i)}{T_{i+1} - T_i}$$

□

This last Lemma leads to the following Corollary.

Corollary 8.1.3. *For $H = 1/2$*

$$\eta_t^2 = \begin{cases} \dfrac{(T_{i_t+1} - t)(t - T_{i_t})}{(t-s)(T_{i_t+1} - T_{i_t})} + \dfrac{(T_{i_s+1} - s)(s - T_{i_s})}{(t-s)(T_{i_s+1} - T_{i_s})} & \text{if } i_s \neq i_t \\ 1 - \dfrac{t-s}{T_{i+1} - T_i} & \text{if } i_s = i_t = i \end{cases} \tag{8.8}$$

Proof. Using (8.2) for $H = 1/2$, then substituting $\lambda_t = \dfrac{t - T_{i_t}}{T_{i_t+1} - T_{i_t}}$ and $\lambda_s = \dfrac{s - T_{i_s}}{T_{i_s+1} - T_{i_s}}$ in expression (8.4) and using Lemma 8.1.2, one gets (8.8).

□

Having the observations of X at time $\{T_i\}_{i\geq 0}$, we note $i_t = \max_{i\geq 0}\{i, T_i < t\}$. We set

$$\eta_{\delta t}(t)^2 = \frac{\mathbb{E}\left[\left(\breve{R}(t,\delta t) - R(t,\delta t)\right)^2 |\mathcal{F}_t\right]}{\mathbb{E}\,|R(t,\delta t)|^2}$$

By Proposition 8.1.1, we note that as the δt decreases, $\eta_{\delta t}(t)^2$ increases. As we will see later, our financial data has uncorrelated (or a little correlated) log-prices return and has $H \approx 0.5$. For simplicity we determine the minimum discrete sampling time δt by computing $\eta_{\delta t}(t)^2$ for $H = 1/2$. In that case, one has by Corollary 8.1.3

$$\eta_{\delta t}(t)^2 = \begin{cases} \dfrac{(T_{i_{t+\delta t}+1} - t - \delta t)(t + \delta t - T_{i_{t+\delta t}})}{\delta t(T_{i_{t+\delta t}+1} - T_{i_{t+\delta t}})} + \dfrac{(T_{i_t+1} - t)(t - T_{i_t})}{\delta t(T_{i_t+1} - T_{i_t})} & \text{if } i_{t+\delta t} \neq i_t \\ 1 - \dfrac{\delta t}{T_{i+1} - T_i} & \text{if } i_{t+\delta t} = i_t = i \end{cases} \tag{8.9}$$

One can then choose the minimum δt by computing the mean of the square root error of the return $\eta_{\delta t}(t)$. We draw in Figure 8.2 $\eta_{\delta t}(t)$ for $\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$. The green line represents the average of $\eta_{\delta t}(t)$. Note that for $\delta t \geq 60$ minutes, the mean of the square root error is smaller or equal than 0.10. We expect having a good approximation of the linear interpolation. However, for our analysis, we also consider values of δt smaller than 60 minutes, knowing that for $\delta t < 60$ minutes, the statistical analysis may be meaningless.

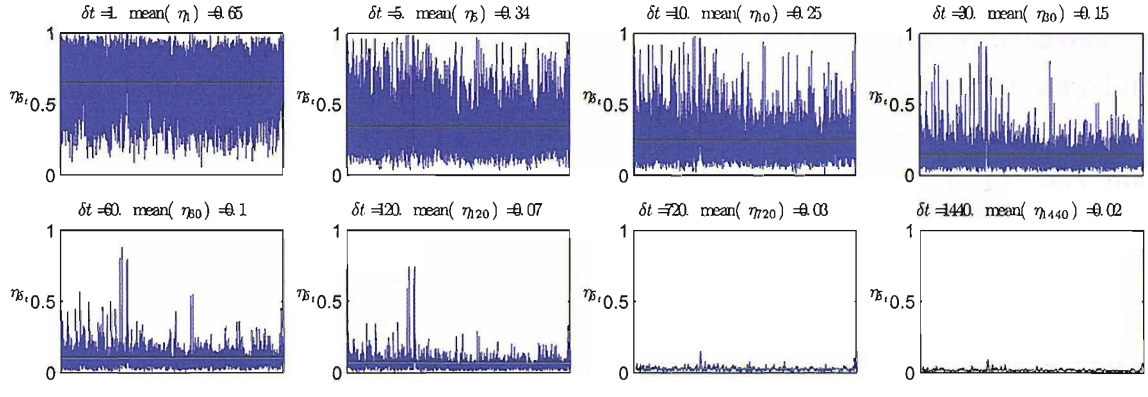

Figure 8.2: The root mean square error of the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns at different time scales ($\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$).

8.1.2 Statistical properties of the log-prices returns

Here, we concentrate on the $DS_{\hat{a}^3_m}$-AUD/USD, for which the deseasonalisation using the method of crossings gives a better estimate quality than the other methods. Further, the time duration have smaller a standard deviation. We study the autocorrelation, the distribution of the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns denoted by $\tilde{R}(t,\delta t) = \tilde{X}(t+\delta t) - \tilde{X}(t)$, where $\tilde{X}(t)$ is the deseasonalised log prices of the AUD/USD rate index. We also study the scaling behaviour of the data.

8.1.2.1 Autocorrelation

We draw the autocorrelation functions of the return (Figure 8.3), the absolute return (Figure 8.4) and the squared return (Figure 8.5), and thus at different time scales δt, in particular for $\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$ in minute. The red lines correspond to the 95% confidence interval bound, determined by using Bartletts formula for moving average processes. We fit the autocorrelation function of the return by the function

$$\rho_H(k) = \frac{1}{2}(|k+1|^{2H} + |k-1|^{2H} - 2|k|^{2H}) \tag{8.10}$$

where k is the k^{th} lag and $H \in (0,1)$. ρ_H is in fact the autocorrelation function of the increments of an H-sssi process and H represent the self-similarity index. We estimate H by fitting ρ_H to the autocorrelation function of the return using a least squares method. This is represented by the green line on Figure 8.3, H is also given for each δt. Note that the value of H is slightly smaller than 1/2, for which $\rho_{1/2}(k) = 0$ for all k and so the return would be uncorrelated. We see that there is little autocorrelation in returns. One can consider that the returns are uncorrelated except for very high frequency in particular for $\delta t \leq 10$ minutes, where there is a negative correlation for the first lag. This was already noticed by Goodhart and Figliuoli [108] for very high frequency data.

The autocorrelation function of the absolute and squared return are correlated, which becomes more pronounced as the sampling frequency increases. Note that for small sampling frequency (larger than half a day here), the data can be considered as having independent return. We observe a strong correlation for the absolute return. Such a characteristic has already been observed by Andersen and Bollerslev [109]. The autocorrelation function of the absolute return and the squared return seem to decay as a power law. We then choose to fit, using a least squares method, these autocorrelation functions by the following function

$$\varrho_{\alpha,\beta}(k) = \frac{1}{(1+|k|^{\alpha})^{\beta}} \tag{8.11}$$

where α and β are two positive real. The fitted line is represented in green in Figures 8.4 and 8.5, as well as their corresponding estimated α and β.

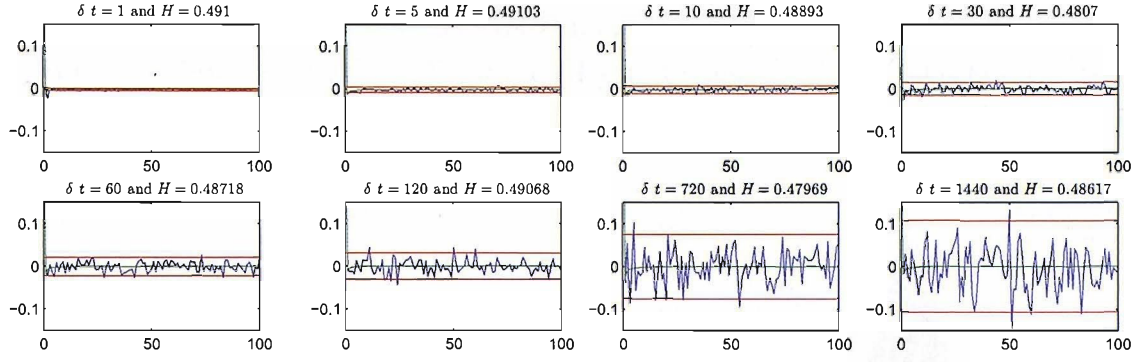

Figure 8.3: Autocorrelation function of the $DS_{\hat{a}_m^3}$-AUD/USD log return at different time scales.

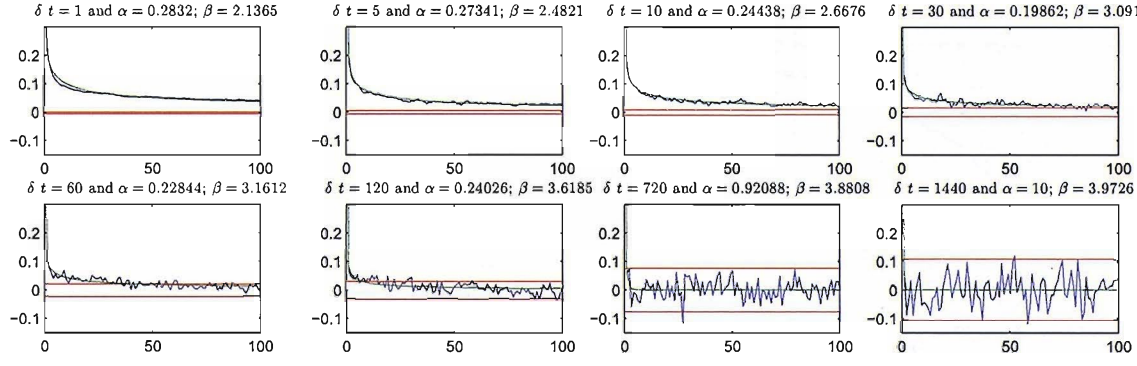

Figure 8.4: Autocorrelation function of the absolute $DS_{\hat{a}_m^3}$-AUD/USD log return at different time scales.

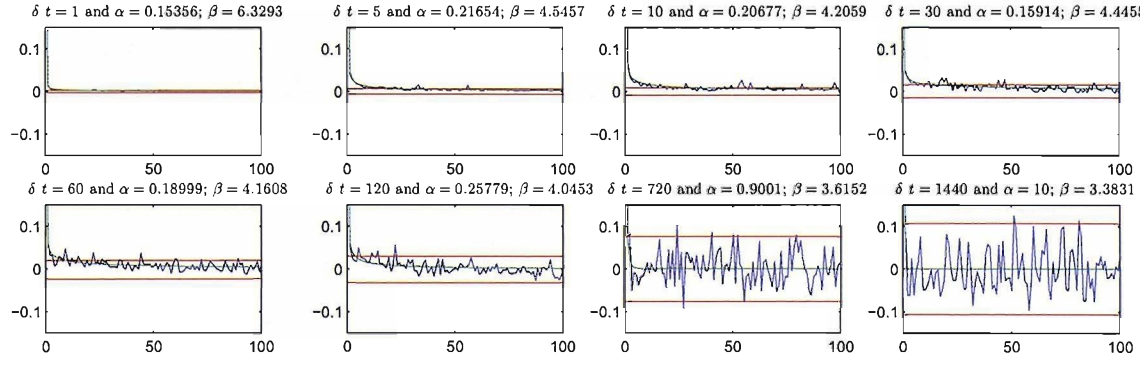

Figure 8.5: Autocorrelation function of the squared $DS_{\hat{a}_m^3}$-AUD/USD log return at different time scales.

8.1.2.2 Fitting the distribution of log-prices returns

We assume that log-prices returns of the $DS_{\hat{a}^3_m}$-AUD/USD are ergodic. We use a pseudo-maximum likelihood estimate to fit the log-prices returns to three different distributions: Normal, α-Stable and symmetric scaled t-distribution. For details on the t-distribution and the uses of pseudo-maximum likelihood see respectively Appendices B and C. We draw the PDF, CDF, log(CDF) and log(1-CDF) of the one hour $DS_{\hat{a}^3_m}$-AUD/USD log-prices return in Figure 8.6 with their fitted distributions.

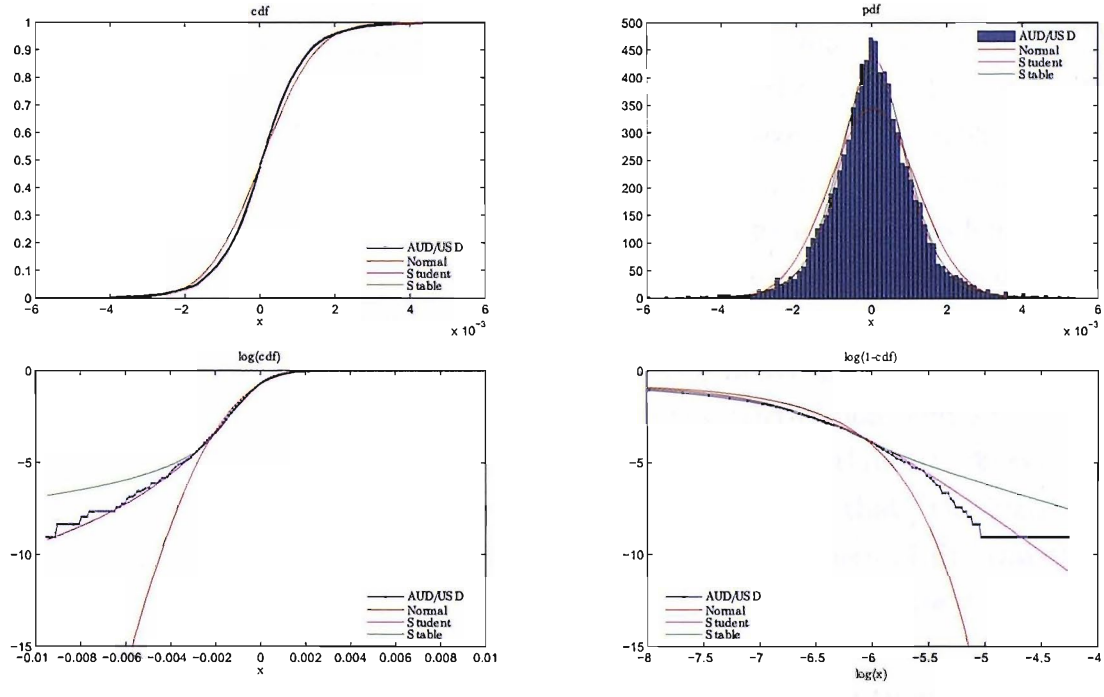

Figure 8.6: PDF (top right), CDF (top left), log(CDF) (bottom left) and log(1-CDF) (bottom right) of the one hour $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns, fitted to the normal (in red), α-stable (in green) and symmetric scaled t-distribution (in magenta).

The estimated parameters for each distribution is given in Figure 8.7. The quantile-quantile plots (QQ-plot) are presented graphically to illustrate the goodness of fit for each PDF to the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns distribution. Note through

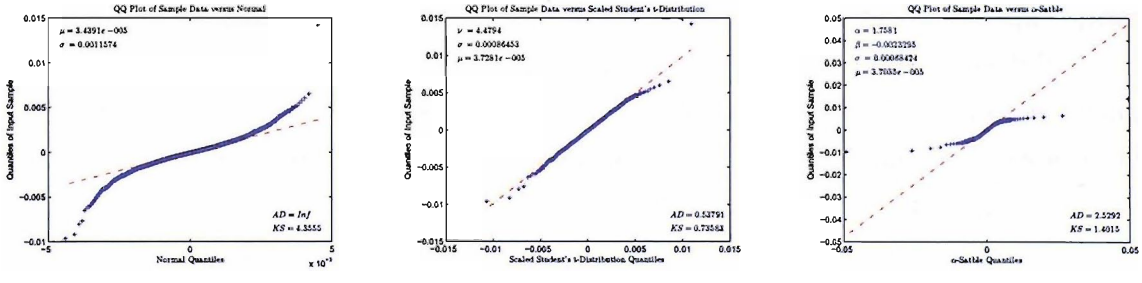

Figure 8.7: QQ-plot goodness of fit of the one hour $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns

these QQ-plots the tail of the marginal distribution of the $DS_{\hat{a}^3_m}$-AUD/USD log-prices

returns are too heavy compared to the Normal distribution and not heavy enough for the α-Stable distribution. The scaled t-distribution shows a better fit.

We fit our data with the symmetric scaled t-distribution $\mathcal{T}(\nu, \sigma, \mu)$, where σ and μ represent respectively the scale parameter and the shift parameter (see Appendix B), and thus for returns at different time scales δt. For formality, we use the Kolmogorov-Smirnov and the Andersen-Darling tests for goodness of fit. These tests are described in Appendix D. The critical values are computed using a bootstrap method. This technique is also described in Appendix D. While the the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns are uncorrelated but dependent, the sample used for bootstrapping are independent random variables. This is a good approximation as the size of the sample are large enough for $\delta t \leq 120$ minutes. In this instance, the size is from 4234 observations for $\delta t = 120$ minutes to 508080 observations for $\delta t = 1$ minutes. For $\delta t > 120$ minutes, we assume that $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns are independents. This assumption is satisfactory since for bigger scales, the absolute return looks uncorrelated in Figure 8.4. The estimated t-distribution parameters for each scales are shown in Tables 8.5. D and A^2 are respectively the Kolmogorov-Smirnov and Anderson-Darling test statistics of the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns distribution, and D_c and A^2_c their respective critical values estimated using the bootstrap method of 500 samples issued from the t-distribution. The hypothesis H_0 is defined so that "the log-prices returns follow a t-distribution". H_0 is accepted when the goodness of fit statistic is smaller than its corresponding critical values and marked by ✓, otherwise it is marked by a cross ✗.

We see in Table 8.5 that the degree of freedom ν increases as the time scale δt gets larger. This shows that the higher the frequency of the sample is, the more peaked the distribution becomes and the heavier its tails are.

Figures 8.8 shows the QQ-plot of the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns distribution to the symmetric scaled t-distribution.

In order that the $DS_{\hat{a}^3_m}$-AUD/USD log price index satisfies the self-similarity property, the degree of freedom ν must be constant for different time scale δt. To check for a possible self-similarity of the data, we fit the log-prices return to a symmetric scaled t-distribution with degree of freedom $\nu = 4$. This seems to be the most appropriate value for different scale. the results are presented in Table 8.6 and Figures 8.9.

The results in Tables 8.5 and 8.6 shows that the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns follow a t-distribution up to a minimum discrete time of 60 and 45 minutes respectively by the D and A^2 statistics test. Note also that the daily return (1440 minutes) does not follow a t-distribution with respect to goodness of fit tests.

δt	Data size	parameters			KS test			AD test		
		ν	σ	μ	D	D_c	H_0	A^2	A_c^2	H_0
1	508080	3.03	9.95e-5	5.43e-7	32.484	0.755	×	369.927	0.637	×
2	254040	3.89	1.57e-4	1.46e-6	10.995	0.766	×	42.275	0.631	×
5	101616	3.80	2.41e-4	3.39e-6	3.547	0.757	×	7.551	0.614	×
10	50808	3.78	3.39e-4	6.01e-6	2.012	0.750	×	3.602	0.661	×
15	33872	3.82	4.18e-4	9.77e-6	1.917	0.761	×	3.196	0.653	×
20	25404	3.88	4.85e-4	1.07e-5	1.485	0.750	×	1.585	0.581	×
30	16936	4.11	6.03e-4	1.71e-5	0.980	0.735	×	0.925	0.603	×
40	12702	4.12	6.93e-4	2.45e-5	1.055	0.762	×	0.724	0.589	×
45	11290	4.19	7.33e-4	2.71e-5	0.846	0.751	×	0.504	0.613	✓
60	8468	4.48	8.65e-4	3.73e-5	0.736	0.768	✓	0.538	0.602	✓
120	4234	4.93	1.24e-3	6.87e-5	0.668	0.761	✓	0.377	0.603	✓
240	2117	5.31	1.81e-3	1.62e-4	0.513	0.766	✓	0.259	0.630	✓
360	1411	5.33	2.19e-3	2.56e-4	0.681	0.754	✓	0.357	0.611	✓
720	705	6.03	3.21e-3	5.23e-4	0.453	0.753	✓	0.438	0.571	✓
1440	352	6.66	4.48e-3	1.01e-3	1.007	0.797	×	0.920	0.581	×
2880	176	6.62	6.21e-3	1.89e-3	0.788	0.792	✓	0.472	0.633	✓

Table 8.5: Fitting the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns to the symmetric scaled t-distribution for different time scale $\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes and their goodness of fit statistics.

δt	Data size	parameters			KS test			AD test		
		ν	σ	μ	D	D_c	H_0	A^2	A_c^2	H_0
1	508080	4.00	1.07e-4	5.75e-7	32.505	0.772	×	556.633	0.656	×
2	254040	4.00	1.58e-4	1.46e-6	10.985	0.777	×	43.352	0.678	×
5	101616	4.00	2.43e-4	3.39e-6	3.533	0.789	×	9.466	0.653	×
10	50808	4.00	3.43e-4	6.02e-6	2.278	0.762	×	4.859	0.623	×
15	33872	4.00	4.22e-4	9.82e-6	2.083	0.784	×	3.967	0.670	×
20	25404	4.00	4.88e-4	1.08e-5	1.493	0.779	×	1.833	0.680	×
30	16936	4.00	6.00e-4	1.70e-5	0.933	0.765	×	0.802	0.661	×
40	12702	4.00	6.89e-4	2.45e-5	1.016	0.778	×	0.657	0.690	✓
45	11290	4.00	7.26e-4	2.70e-5	0.847	0.781	×	0.470	0.703	✓
60	8468	4.00	8.46e-4	3.74e-5	0.593	0.771	✓	0.526	0.685	✓
120	4234	4.00	1.19e-3	6.72e-5	0.720	0.759	✓	0.556	0.711	✓
240	2117	4.00	1.73e-3	1.64e-4	0.507	0.782	✓	0.381	0.666	✓
360	1411	4.00	2.08e-3	2.58e-4	0.461	0.782	✓	0.324	0.668	✓
720	705	4.00	3.00e-3	5.50e-4	0.488	0.767	✓	0.478	0.673	✓
1440	352	4.00	4.13e-3	1.11e-3	0.861	0.737	×	0.927	0.621	×
2880	176	4.00	5.72e-3	2.01e-3	0.661	0.777	✓	0.463	0.654	✓

Table 8.6: Fitting the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns to the symmetric scaled t-distribution with degree of freedom $\nu = 4$, for different time scale $\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes and their goodness of fit statistics.

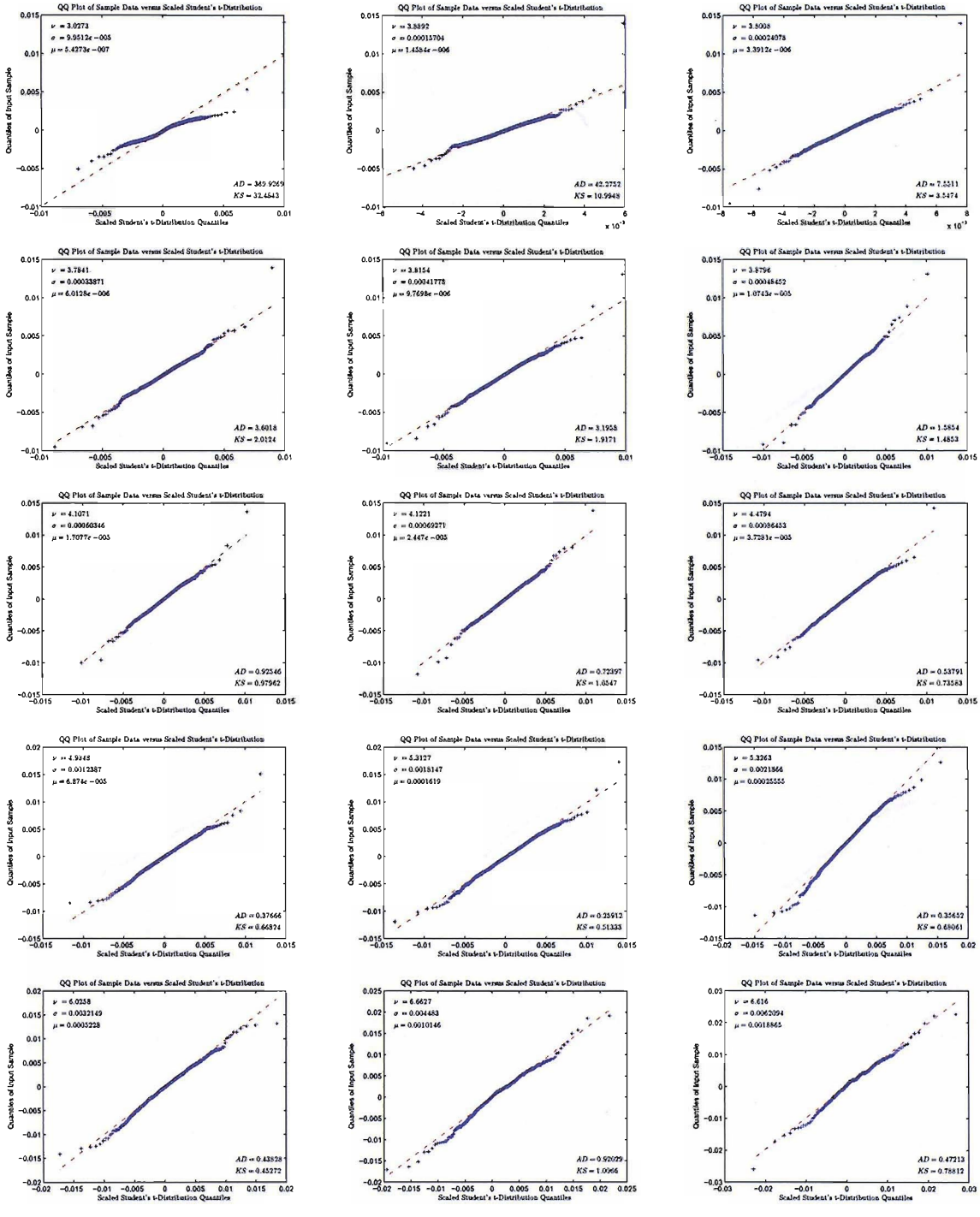

Figure 8.8: QQ-plots the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns distribution to the symmetric scaled t-distribution for different time scale $\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes.

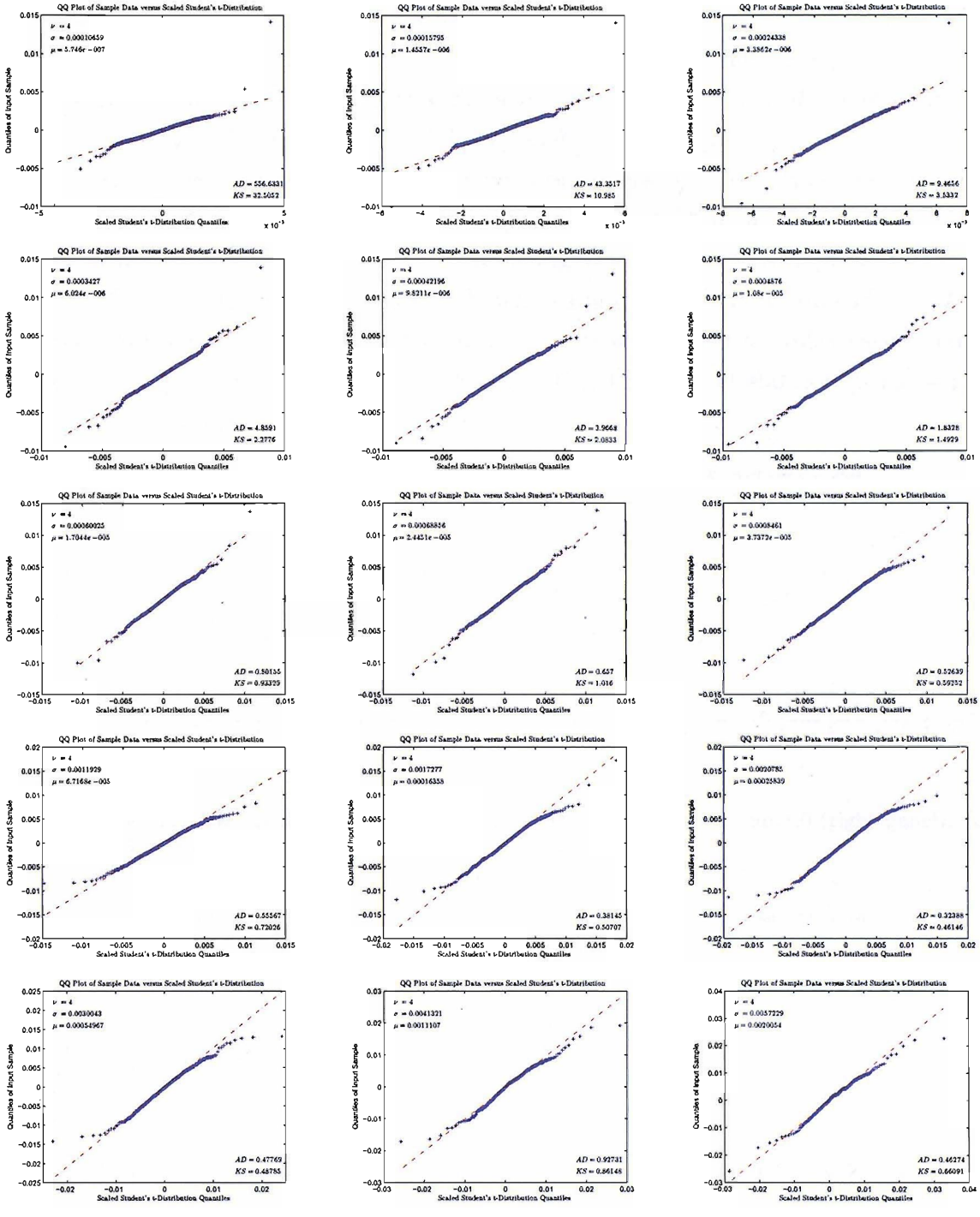

Figure 8.9: QQ-plots the $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns distribution to the symmetric scaled t-distribution with degree of freedom $\nu = 4$, for different time scale $\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes.

8.1.2.3 Scaling law behaviour and self-similarity

Under the self-similarity assumption of the $DS_{\hat{a}_m^3}$-AUD/USD log price, the scaling parameter $\sigma_{\delta t}$ of the t-distribution estimated from the data, must satisfy $\sigma_{\delta t} = \delta t^H \sigma_1$, where H is the scaling exponent. Moreover, it can be shown empirically that $Z(t) = \tilde{X}(t) - \mu_1 t$, where μ_1 is the location parameter for the 1 minute log-prices returns, given in Table 8.6, is self-similar. It suffices to check that $\mu_{\delta t} = \mu_1 \delta t$. Indeed, if Z is H-sssi , then $\mathbb{E}[Z(1)] = 0$ and $\mu_{\delta t} = \mathbb{E}[X(\delta t)] = \mathbb{E}[Z(\delta t)] + \mu_1 \delta t = \mu_1 \delta t$.
As Table 8.6 shows that the t-distribution with parameter $\nu = 4$ fits well the log-prices return for $\delta t \in I = \{30, 40, 45, 60, 120, 240, 360, 720, 1440\}$[4] in minutes, we compute $\hat{H}_I$ an estimate of H for $\delta t \in I$ by computing the slope of the regression line of $\log \sigma_{\delta t} = H \log \delta t + \log \sigma_1$. The result is shown in Figure 8.10 and we find $\hat{H}_I = 0.499$[5]. The value of the estimator is almost identical to the self-similarity index estimated on Figure 8.3 using the autocorrelation function (8.10). Figure 8.11 shows that $\mu_{\delta t} = \mu_1 \delta t$ is reasonably satisfied.

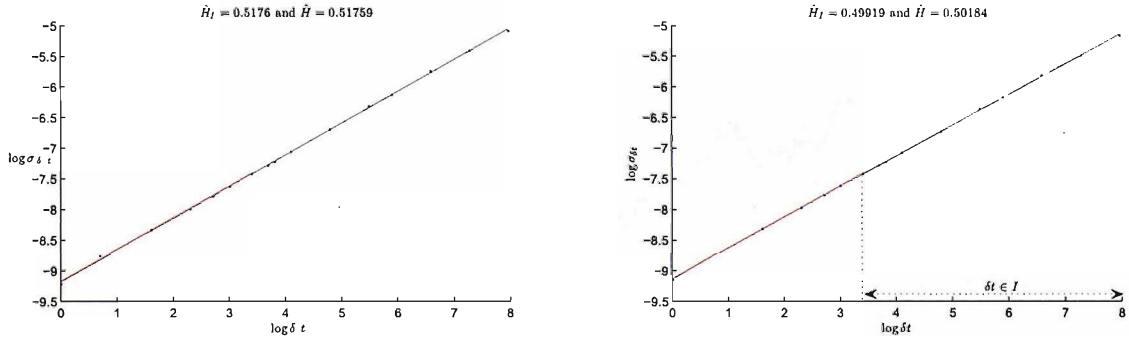

Figure 8.10: Scaling law exponent estimation using Tables 8.5 (left panel) and 8.6 (right panel). Regression over $\delta t \in I$ (green line) and regression over all δt (red line).

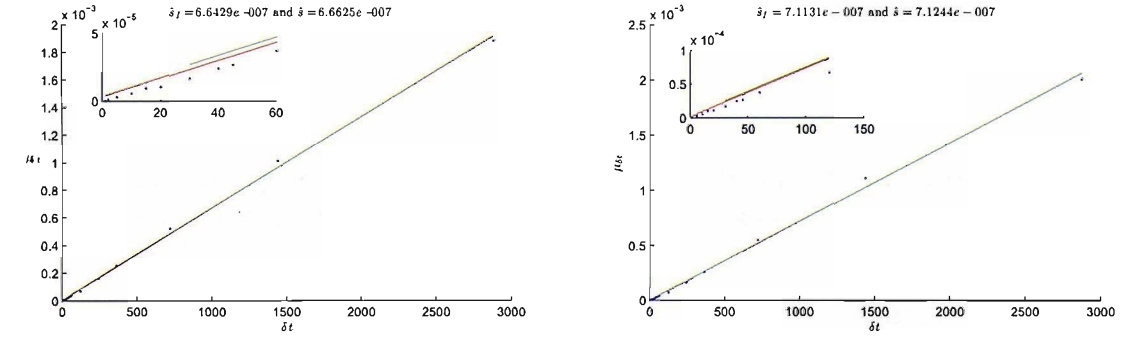

Figure 8.11: Location parameter estimation using Tables 8.5 (left panel) and 8.6 (right panel). Regression over $\delta t \in I$ (green line) and regression over all δt (red line).

[4]The AD and KS tests are quite small for $\delta t = 30, 40, 1440$, so we consider that the t-distribution fits well their corresponding returns.

[5]Note that this same values was found when estimating the self-similarity exponent of the non deseasonalised AUD/USD data using the EBP method in the previous Chapter

8.2 Statistical analysis of the deseasonalised FTSE100 future contract

The interesting side of the FTSE100 future contract is it contains jumps at the opening market time. This allows us to test the performance of our time change estimators on data with jumps and in that case, we expect the log-variation estimator to perform the best as this last is robust to jumps. This performance will be noticed once we study the time duration of the $DS_{\hat{a}^i_{med}}$-FTSE100, where $i \in \{0, 1, 2, 3\}$. In this section, we carry the same analysis as for the $DS_{\hat{a}^i_m}$-AUD/USD. Once the performance of the estimators on data containing jumps are studied, we remove the jumps at opening market times from the data so the statistical analysis does not take into consideration the jumps at opening market times. Jumps may be modeled separately. They can be added into a model as a jump process, where jumps occur at opening market times and should follow a certain distribution. Figure 8.12 shows the separation of the jumps from the $DS_{\hat{a}^2_{med}}$-FTSE100 future contract prices.

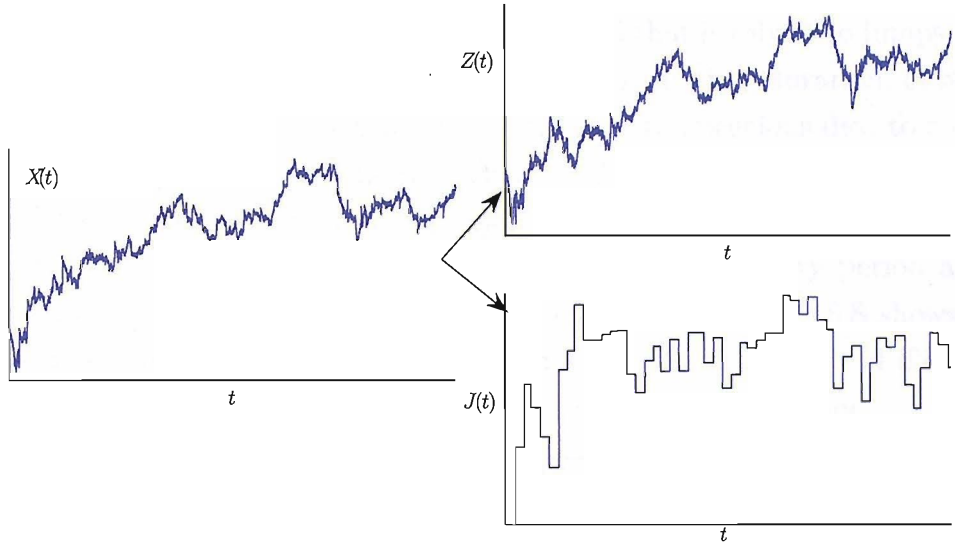

Figure 8.12: Removing opening times jumps from the $DS_{\hat{a}^2_{med}}$-FTSE100 future contract prices. $X(t)$ represent the logarithmic price process, $J(t)$ the jumps at opening market time and $Z(t)$ a copy of $X(t)$ without jumps.

The $DS_{\hat{a}^i_{med}}$-FTSE100 log-prices returns (with jumps to opening market time removed) analysis will be conducted in the same way as for the $DS_{\hat{a}^i_m}$-AUD/USD.

8.2.1 Time duration and data sampling analysis

8.2.1.1 Time duration analysis

Defining τ as the time duration between two price ticks, its distribution is presented in Table 8.7 for large scales time and Table 8.8 for smaller scales.

(in hour)	$\tau < 1$	$1 \leq \tau < 12$	$12 \leq \tau < 24$	$24 \leq \tau$
FTSE100	110942	0	41	10
$DS_{\hat{a}^0_m}$-FTSE100	110910	83	0	0
$DS_{\hat{a}^1_{med}}$-FTSE100	110942	51	0	0
$DS_{\hat{a}^2_{med}}$-FTSE100	110993	0	0	0
$DS_{\hat{a}^3_{med}}$-FTSE100	110922	71	0	0

Table 8.7: Time duration distribution (in hour) of the FTSE100 at large time scale.

(in minute)	$\tau < 1$	$1 \leq \tau < 5$	$5 \leq \tau < 10$	$10 \leq \tau < 60$	$60 \leq \tau$
FTSE100	107173	3691	25	53	51
$DS_{\hat{a}^0_m}$-FTSE100	84473	24588	1555	294	83
$DS_{\hat{a}^1_{med}}$-FTSE100	83509	26438	865	130	51
$DS_{\hat{a}^2_{med}}$-FTSE100	78326	30653	1733	281	0
$DS_{\hat{a}^3_{med}}$-FTSE100	81148	28638	1005	131	71

Table 8.8: Time duration (in minute) distribution of the FTSE100 at small time scale.

We note in Table 8.7 that the time duration does not exceed one hour when using the log-variation method for estimating the time change, which is not the case for the other methods. In fact, except for the log-variation method that is robust to jumps, the other methods are affected by jumps. They tend to enlarge the time duration, between ticks at opening market times and ticks at closing times of the previous day, to compensate for the size of the jump so that the time changed data looks continuous. In Table 8.8, we observe the phenomena of stretching small time duration and contracting large time duration. This is when the θ-time runs faster during low activity period and slower when the activity period is high. Once again the result in Table 8.8 shows that the log-variation performs better and this is confirmed by the low standard deviation of τ (see Table 8.9) compared to the ones resulting from the other method.

	FTSE100	$DS_{\hat{a}^0_m}$	$DS_{\hat{a}^1_{med}}$	$DS_{\hat{a}^2_{med}}$	$DS_{\hat{a}^3_{med}}$
$\mathbb{E}[\tau]$	0.92217	0.93302	0.93351	0.93486	0.93282
$\sqrt{\mathbb{V}ar(\tau)}$	39.4073	4.5542	6.2861	1.4161	3.6942

Table 8.9: Time duration (in minute) average and standard deviation of the FTSE100.

For good performance of the log-variation method, we decide to carry on with, the study on the $DS_{\hat{a}^2_{med}}$-FTSE100 log return.

8.2.1.2 Data sampling

As shown in the previous Section, a linear interpolation is a reasonable choice. We draw on Figure 8.13 $\eta_{\delta t}(t)$ given by (??) for $\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$. The green line represent the average of $\eta_{\delta t}(t)$. The linear interpolation will be a good approximation up to a certain time scale δt. We may expect a reasonable approximation of the linear interpolation for $\delta t \geq 30$ minutes.

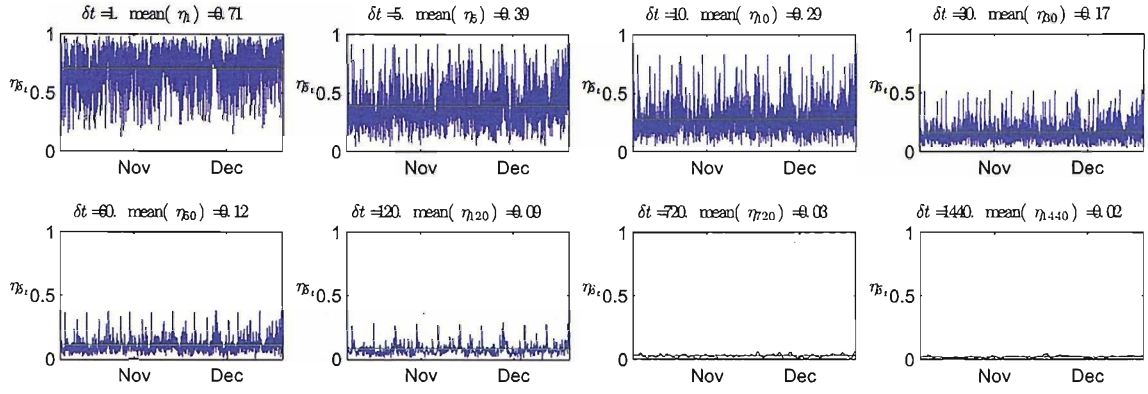

Figure 8.13: The root mean square error of the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices return at different time scales ($\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$).

8.2.2 Statistical properties of the log-prices returns

8.2.2.1 Autocorrelation

We draw the autocorrelation function of the return Figure 8.14, the absolute return Figure 8.15 and the squared return Figure 8.16, and thus at different time scales δt, in particular for $\delta t \in \{1, 5, 10, 30, 60, 120, 720, 1440\}$ in minute. The red lines corresponds to the 95% confidence interval bound for zero autocorrelation.
We fit the autocorrelation function of the log-prices returns by the function (8.10), and the autocorrelation function of the absolute return and the squared log-prices returns by function (8.11), both using the least square method. These are represented by the green line. Note that there is positive autocorrelation in the log-prices returns for small scales. One can consider that the returns are uncorrelated for $\delta t \geq 10$ minutes. As for the $DS_{\hat{a}^3_m}$-AUD/USD, the absolute and squared log-prices returns are positively correlated and decay as a power law function as the number of lag increases.

8.2.2.2 Fitting the distribution of log-prices returns

We assume that log-prices returns of the $DS_{\hat{a}^2_{med}}$-FTSE100 are ergodic. We use a pseudo-maximum likelihood estimate to fit the log-prices returns to three different distributions: Normal, α-Stable and symmetric scaled t-distribution. We draw the PDF, CDF, log(CDF) and log(1-CDF) of the one hour $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices returns in Figure 8.17 with their fitted distributions. The estimated parameters for each distribution is given in Figure 8.18. The quantile-quantile plots (QQ-plot) are presented to graphically illustrate the goodness of fit of each PDF to the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices returns distribution.
As for $DS_{\hat{a}^3_m}$-AUD/USD, the symmetric t-distribution seem to fit better the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices return. We fit our data with the symmetric scaled t-distribution $\mathcal{T}(\nu, \sigma, \mu)$, where σ and μ represent respectively the scale and the shift parameters (see Appendix B), and thus for returns at different time scales δt. The Kolmogorov-

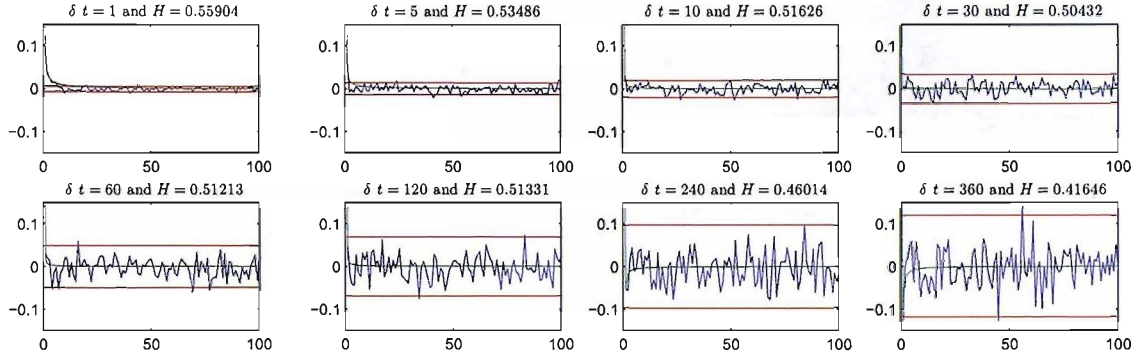

Figure 8.14: Autocorrelation function of $DS_{\hat{a}^2_{med}}$-FTSE100 log return at different time scales.

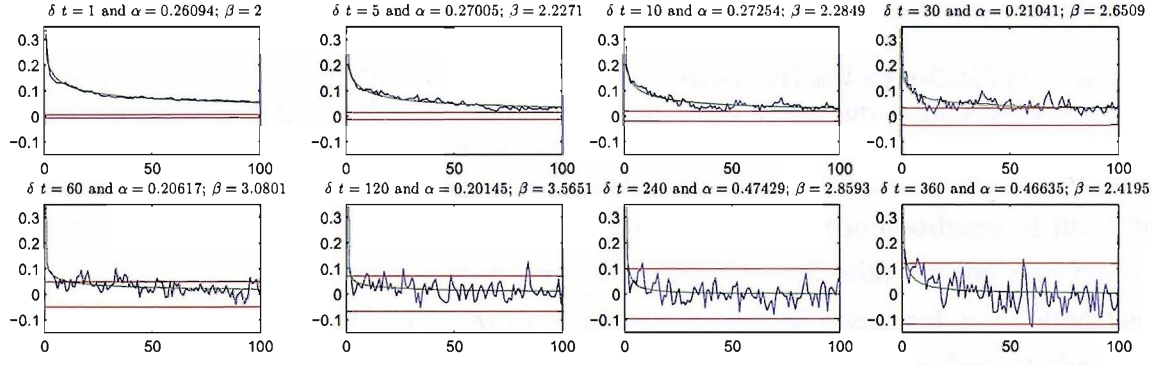

Figure 8.15: Autocorrelation function of the absolute $DS_{\hat{a}^2_{med}}$-FTSE100 log return at different time scales.

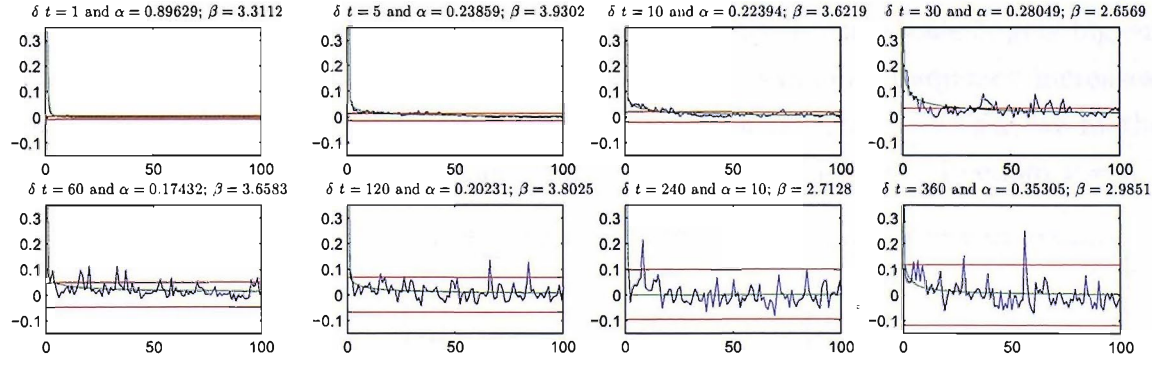

Figure 8.16: Autocorrelation function of the squared $DS_{\hat{a}^2_{med}}$-FTSE100 log return at different time scales.

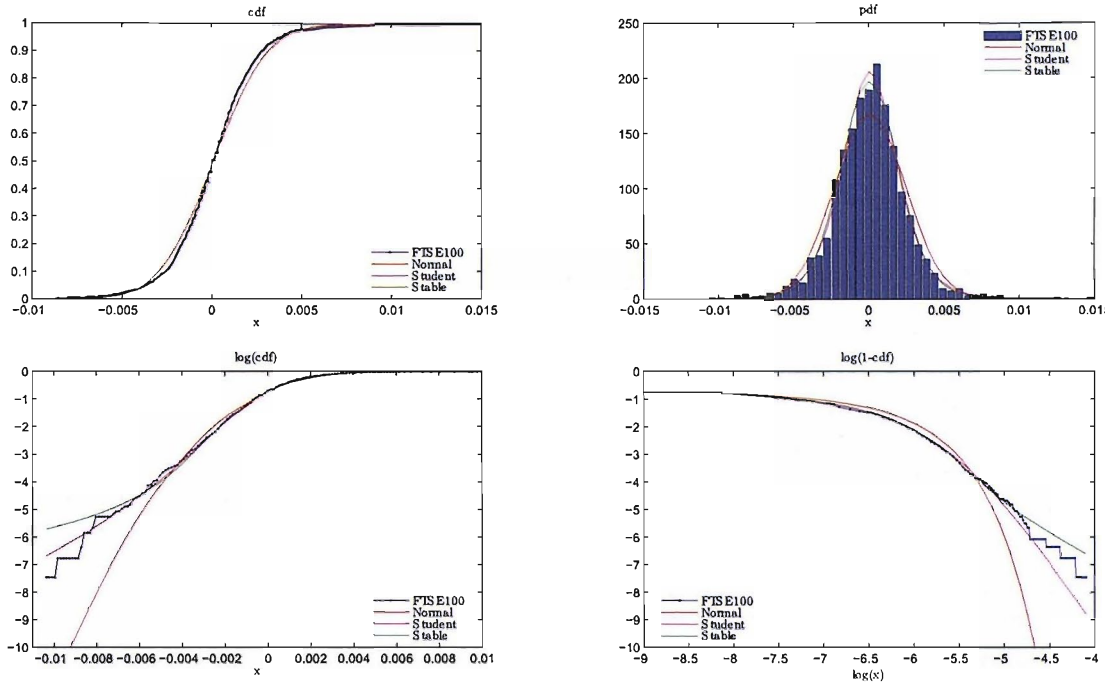

Figure 8.17: PDF (top right), CDF (top left), log(CDF) (bottom left) and log(1-CDF) (bottom right) of the one hour $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices returns, fitted to the normal (in red), α-stable (in green) and symmetric scaled t-distribution (in magenta).

Smirnov and Andersen-Darling tests are computed to check the goodness of fit. The critical values are computed using a bootstrap method. Similar to the parameters estimation of $DS_{\hat{a}^3_m}$-AUD/USD log-prices returns, a similar argument is applied here to justify the use of the pseudo-likelihood estimator and the critical values of the tests computation by bootstrapping. The hypothesis H_0 is defined so that "log-prices returns follow a t-distribution". H_0 is accepted when the goodness of fit statistic is smaller than its corresponding critical values and marked by ✓, otherwise it is marked by a cross ×. We note that the degree of freedom ν increases as the time scale δt gets bigger. The distribution of the sample gets more peaked as the sampling frequency increases. In order to check for a possible self-similar properties of the $DS_{\hat{a}^2_{med}}$-FTSE, we fit the log-prices returns to a symmetric scaled t-distribution with degree of freedom $\nu = 5$.

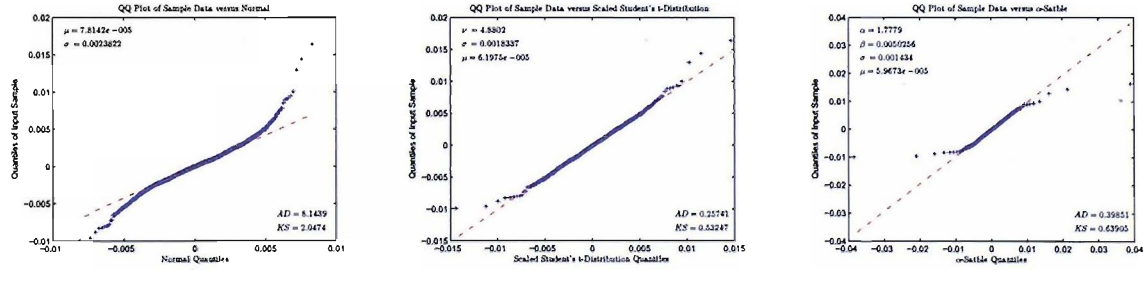

Figure 8.18: QQ-plot goodness of fit of the one hour $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices returns

The results in Tables 8.10 and 8.11 shows that the $DS_{\hat{a}^2_{med}}$-FTSE log-prices increments follow a t-distribution up to a minimum discrete time of 20 and 10 minutes respectively by the D and A^2 statistics test.

δt	Data size	parameters			KS test			AD test		
		ν	σ	μ	D	D_c	H_0	A^2	A^2_c	H_0
1	103763	2.10	1.23e-4	2.18e-6	25.793	0.788	×	336.130	0.771	×
2	51881	2.96	2.33e-4	3.74e-6	11.583	0.759	×	62.663	0.670	×
5	20752	3.85	4.45e-4	5.13e-6	3.045	0.771	×	2.878	0.639	×
10	10376	4.46	6.98e-4	9.79e-6	1.187	0.744	×	0.482	0.564	✓
15	6917	4.31	8.70e-4	1.16e-5	0.829	0.758	×	0.287	0.649	✓
20	5188	4.40	1.02e-3	2.23e-5	0.652	0.737	✓	0.326	0.595	✓
30	3458	4.66	1.29e-3	2.39e-5	0.378	0.750	✓	0.185	0.594	✓
45	2305	4.85	1.57e-3	5.46e-5	0.564	0.767	✓	0.167	0.596	✓
60	1729	4.88	1.83e-3	6.20e-5	0.532	0.741	✓	0.257	0.670	✓
120	864	5.13	2.61e-3	1.17e-4	0.457	0.751	✓	0.199	0.583	✓
240	432	7.53	4.29e-3	3.74e-4	0.679	0.791	✓	0.243	0.625	✓
360	288	5.84	5.13e-3	4.92e-4	0.530	0.789	✓	0.280	0.620	✓

Table 8.10: Fitting the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices return to the symmetric scaled t-distribution for different time scale $\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes and their goodness of fit statistics.

δt	Data size	parameters			KS test			AD test		
		ν	σ	μ	D	D_c	H_0	A^2	A^2_c	H_0
1	103763	5.00	1.62e-4	2.03e-6	26.280	0.788	×	690.962	0.625	×
2	51881	5.00	2.64e-4	3.40e-6	11.812	0.783	×	114.801	0.657	×
5	20752	5.00	4.69e-4	5.09e-6	3.073	0.784	×	5.845	0.644	×
10	10376	5.00	7.12e-4	9.90e-6	1.190	0.814	×	0.483	0.628	✓
15	6917	5.00	8.94e-4	1.18e-5	0.833	0.781	×	0.315	0.635	✓
20	5188	5.00	1.04e-3	2.28e-5	0.644	0.774	✓	0.347	0.700	✓
30	3458	5.00	1.30e-3	2.42e-5	0.457	0.795	✓	0.192	0.617	✓
45	2305	5.00	1.58e-3	5.46e-5	0.569	0.790	✓	0.166	0.667	✓
60	1729	5.00	1.84e-3	6.19e-5	0.515	0.788	✓	0.256	0.688	✓
120	864	5.00	2.60e-3	1.17e-4	0.447	0.801	✓	0.196	0.637	✓
240	432	5.00	4.06e-3	3.87e-4	0.533	0.751	✓	0.270	0.619	✓
360	288	5.00	5.01e-3	4.98e-4	0.485	0.765	✓	0.275	0.675	✓

Table 8.11: Fitting the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices return to the symmetric scaled t-distribution with degree of freedom $\nu = 5$, for different time scale$\delta t \in \{1, 2, 5, 10, 15, 20, 30, 40, 60, 120, 240, 360, 720, 1440, 2880\}$ minutes and their goodness of fit statistics.

8.2.2.3 Scaling law behaviour and self-similarity

Assuming that the $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices satisfies the self-similar property with index $\hat{H}_I$, the scaling parameter of the t-distribution $\sigma_{\delta t}$ at a given δt, must satisfy

$$\sigma_{\delta t} = \delta t^{\hat{H}_I} \sigma_1$$

where H is the scaling exponent computed from the slope of the line fitted to

$$\log \sigma_{\delta t} = \hat{H}_I \log \delta t + \log \sigma_1$$

We find $\hat{H}_I = 0.559$ and $\hat{H}_I = 0.540$ respectively when using Tables 8.10 and 8.11. These are represented in Figure 8.19. Concerning the relation between the location

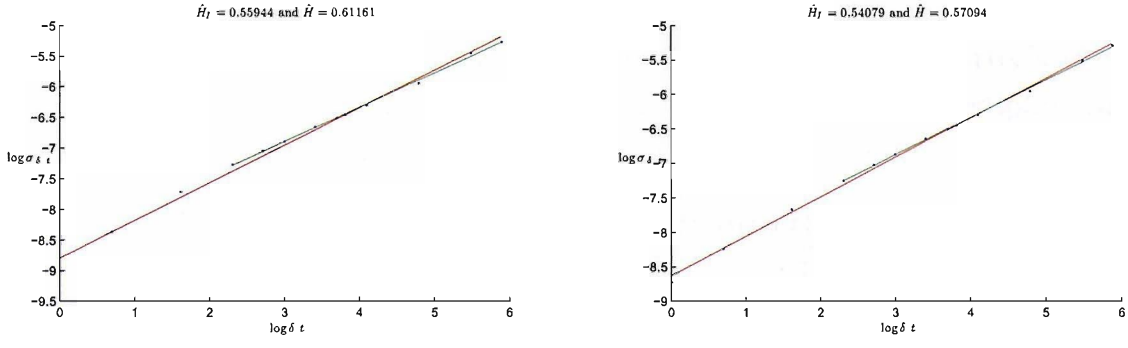

Figure 8.19: Scaling law exponent estimation using Tables 8.10 (left panel) and 8.11 (right panel). Regression over $\delta t \in I$ (green line) and regression over all δt (red line).

parameters, one should have $\mu_{\delta t} = \mu_1 \delta t$. This is shown in Figure 8.20.

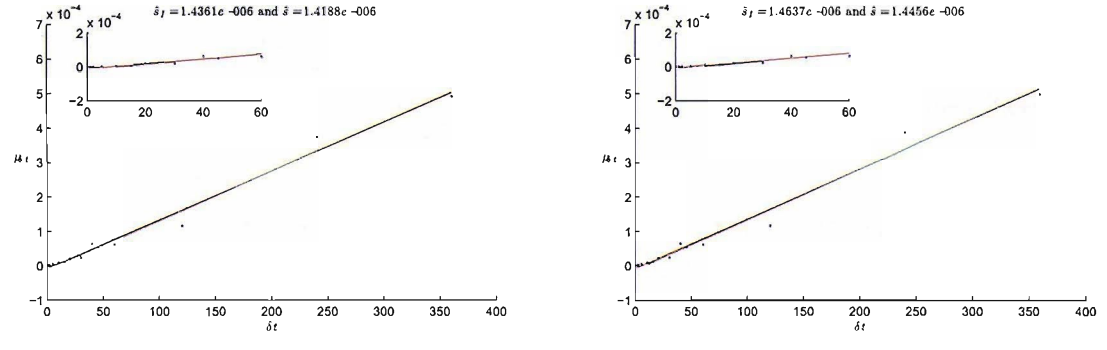

Figure 8.20: Location parameter estimation using Tables 8.10 (left panel) and 8.11 (right panel). Regression over $\delta t \in I$ (green line) and regression over all δt (red line).

In conclusion, it reasonable to argue that the $DS_{\hat{a}^3_{med}}$-FTSE100 has some scaling properties. It may be modelled by a scaling process. In the next Section, we show how the AUD/USD index rates and the FTSE100 future contract can be modelled using a time-changed process.

8.3 Modelling with a time-changed subordinated process

Throughout this Section, we denote by $S(t)$ and $\tilde{S}(t)$ the price of a given asset at time t respectively in calendar time and operational time, so that $S(t) = \tilde{S}(\theta(t))$, where $\theta \in \Theta_\tau$, such that $\tau = 1$ week and $\theta(0) = 0$. We set the log-prices $X(t) = \log(S(t)/S(0))$ (resp. $\tilde{X}(t) = \log(\tilde{S}(t)/S(0))$) and its increments $R(t, \delta t) = X(t + \delta t) - X(t)$ (resp. $\tilde{R}(t, \delta t) = \tilde{X}(t + \delta t) - \tilde{X}(t)$).

8.3.1 Statistical properties of the log-price increments

Let us look at the skewness and kurtosis of the log-prices increments $\tilde{R}(t, \delta t)$, given respectively by $s_{\delta t}$ and $\kappa_{\delta t}$ (8.12), where $\mu_{\delta t}$ and $\sigma_{\delta t}$ are respectively the mean and the standard deviation of $\tilde{R}(t, \delta t)$. For a Gaussian random variables $s_{\delta t} = 0$ and $\kappa_{\delta t} = 3$. In Table 8.12, we present an estimate of the skewness and kurtosis of the previously studied deseasonalised data distribution and thus at different sampling time frequencies. We see that $s_{\delta t}$ is non null and $\kappa_{\delta t} > 3$.

$$s_{\delta t} = \frac{\mathbb{E}[(\tilde{R}(t, \delta t) - \mu_{\delta t})^3]}{\sigma_{\delta t}^3} \text{ and } \kappa_{\delta t} = \frac{\mathbb{E}[(\tilde{R}(t, \delta t) - \mu_{\delta t})^4]}{\sigma_{\delta t}^4} \tag{8.12}$$

	$DS_{\hat{a}^3_m}$-AUD/USD		$DS_{\hat{a}^2_{med}}$-FTSE100	
δt (in min)	$s_{\delta t}$	$\kappa_{\delta t}$	$s_{\delta t}$	$\kappa_{\delta t}$
5	0.246	11.58	0.3675	7.85
15	0.0035	11.36	0.2980	5.17
30	0.0361	8.45	0.3319	4.87
60	-0.0035	6.86	0.3471	4.71
120	0.0226	6.11	0.2061	5.59
720	-0.2716	3.99	-0.1335	5.24
1440	-0.2512	4.52	-0.2803	3.32

Table 8.12: Kurtosis estimate of the $DS_{\hat{a}^3_m}$-AUD/USD and $DS_{\hat{a}^2_{med}}$-FTSE100 log-prices increments.

Let us summarise [6] the empirical log-prices returns $\tilde{R}(t, \delta t)$ statistical characteristics of the deseasonalised financial data previously studied as follows

(A_1)- $\tilde{R}(t, \delta t) \sim \mathcal{T}(\nu, \sigma, \mu)$. The log-prices increments follow a symmetric scaled t-distribution with degree of freedom $\nu \in (3, 6)$ depending on the sampling frequency. The higher the sampling frequency is, the smaller the degree of freedom ν is.

(A_2)- $\kappa_{\delta t} > 3$. The log-prices increments exhibit higher kurtosis than the normal distribution.

(A_3)- For an integer $k > 0$, $Cov\left(\tilde{R}(t + k\delta t, \delta t), \tilde{R}(t, \delta t)\right) \approx 0$. The log-prices increments are uncorrelated (or little correlated).

[6] We omit the skewness of the log prices return.

(A_4)- For some positive real α_1, $Cov\left(|\tilde{R}(t+k\delta t,\delta t)-\mu\delta t|,|\tilde{R}(t,\delta t)-\mu\delta t|\right)\sim k^{-\alpha_1}$, as $k\longrightarrow+\infty$. The log-prices increments are correlated and decay as a power law function.

(A_5)- For some positive real α_2, $Cov\left(|\tilde{R}(t+k\delta t,\delta t)-\mu\delta t|^2,|\tilde{R}(t,\delta t)-\mu\delta t|^2\right)\sim k^{-\alpha_2}$, as $k\longrightarrow+\infty$. The log-prices increments are little correlated and decay as a power law function.

(A_6)- $\mathbb{E}[|\tilde{R}(t,\delta t)|]=(\delta t)^H\mathbb{E}[|\tilde{R}(t,1)|]$. For large δt, the log-prices increments satisfies the scaling law property, where $H\approx 1/2$.

(A_7)- We assume the process $\tilde{R}(t,\delta t)$ stationary. Since the time-change function is estimated in order to obtain a process that has stationary increments, the deseasonalised log-prices should have stationary increments. Even if we do not have direct evidence of this last, we assume that our deseasonalised data has stationary increments.

8.3.2 Modelling the deseasonalised log-prices with Heyde's FATGBM

8.3.2.1 The log-prices returns as a subordinated Brownian motion

In this section, we describe a model[7] proposed by Heyde [13] for the log-prices returns, which satisfies the assumptions $(A_1),\ldots,(A_7)$. We assume the log-price of an asset $\tilde{X}(t)$ is such that

$$\tilde{X}(t)=\mu t+\sigma Z(t)$$

where μ and σ represent respectively the drift and the scale parameters, and $Z(t)$ is a stochastic process, with stationary increments ($Z(t)$ will be defined later). The log-price return is then given by

$$\tilde{R}(t,\delta t)=\mu\delta t+\sigma(Z(t+\delta t)-Z(t)) \tag{8.13}$$

We choose a process Z, so that the appropriate statistical characteristic of our empirical data is obtained. As mentioned by Heyde in [13], such a process can be obtained through a subordinated Brownian motion $B(T(t))$, where $B(t)$ is a Brownian motion and $T(t)$ is a strictly increasing process with stationary increments and independent of $B(t)$. This process was initially introduced by Clark in [105], where $T(t)$ represents the stochastic market clock. Clark shows that such a subordinated process has stationary increments [105]. This satisfies the assumption (A_7) of the log-prices returns. Since processes T and B are independent, setting $\tau_t=T(t+\delta t)-T(t)$, one has $B(T(t+\delta t))-B(T(t))\overset{(d)}{=}\sqrt{\tau_t}B(1)$, and substituting $Z(t)=B(T(t))$ in (8.14), one gets the

[7]The model for the stock prices belongs to Heyde [13] and developed by Heyde and Leonenko [110]. It is called FATGBM for Fractal Activity Time Geometric Brownian Motion.

following equality in distribution

$$\tilde{R}(t,\delta t) = \mu\delta t + \sigma(B(T(t+\delta t)) - B((t))) \overset{(d)}{=} \mu\delta t + \sigma\sqrt{\tau_t}B(1) \tag{8.14}$$

The kurtosis κ of $\tilde{R}(t,\delta t)$ is larger than 3, and so (A_2) is satisfied. Indeed one has

$$\kappa = \frac{\mathbb{E}[{\tau_t}^2(B(1))^4]}{\mathbb{V}ar\left(\sqrt{\tau_t}B(1)\right)^2} = \frac{\mathbb{E}[{\tau_t}^2]\mathbb{E}[(B(1))^4]}{(\mathbb{E}[\tau_t])^2(\mathbb{E}|B(1)|^2)^2} = 3\left(1 + \frac{\mathbb{V}ar\,(\tau_t)}{(\mathbb{E}[\tau_t])^2}\right) > 3$$

The following Proposition shows that the log-prices increments as defined in (8.13) satisfies the statistical properties (A_3),(A_4) and (A_5) if the increments of the process $T(t)$ are positively correlated. This is in fact given in [110].

Proposition 8.3.1. *Assuming that the process $T(t)$ has dependent increments, Then*

- $Cov\left(\tilde{R}(t+k\delta t,\delta t), \tilde{R}(t,\delta t)\right) = 0$
- $Cov\left(|\tilde{R}(t+k\delta t,\delta t) - \mu\delta t|, |\tilde{R}(t,\delta t) - \mu\delta t|\right) = \sigma^2 Cov\left(\sqrt{\tau_{t+k\delta t}}, \sqrt{\tau_t}\right)$
- $Cov\left(|\tilde{R}(t+k\delta t,\delta t) - \mu\delta t|^2, |\tilde{R}(t,\delta t) - \mu\delta t|^2\right) = \sigma^4 Cov\left(\tau_{t+k\delta t}, \tau_t\right)$

Proof. We have

$$\begin{aligned}
&Cov\left(\tilde{R}(t+k\delta t,\delta t), \tilde{R}(t,\delta t)\right) \\
&\quad = Cov\left(\mu\delta t + \sigma\sqrt{\tau_{t+k\delta t}}B_k(1), \mu\delta t + \sigma\sqrt{\tau_t}B_0(1)\right) \\
&\quad = \sigma^2 Cov\left(\sqrt{\tau_{t+k\delta t}}B_k(1), \sqrt{\tau_t}B_0(1)\right) \\
&\quad = \sigma^2\left(\mathbb{E}[\sqrt{\tau_{t+k\delta t}}B_k(1)\sqrt{\tau_t}B_0(1)] - \mathbb{E}[\sqrt{\tau_{t+k\delta t}}B_k(1)]\mathbb{E}[\sqrt{\tau_t}B_0(1)]\right) \\
&\quad = \sigma^2(\mathbb{E}[B_0(1)])^2 Cov\left(\sqrt{\tau_{t+k\delta t}}, \sqrt{\tau_t}\right) \\
&\quad = 0
\end{aligned}$$

Similarly we get

$$\begin{aligned}
Cov\left(|\tilde{R}(t+k\delta t,\delta t) - \mu\delta t|, |\tilde{R}(t,\delta t) - \mu\delta t|\right) &= Cov\left(\sigma\sqrt{\tau_{t+k\delta t}}|B_k(1)|, \sigma\sqrt{\tau_t}|B_0(1)|\right) \\
&= \sigma^2(\mathbb{E}[|B_0(1)|])^2 Cov\left(\sqrt{\tau_{t+k\delta t}}, \sqrt{\tau_t}\right) \\
&= \sigma^2 Cov\left(\sqrt{\tau_{t+k\delta t}}, \sqrt{\tau_t}\right)
\end{aligned}$$

and

$$\begin{aligned}
Cov\left(|\tilde{R}(t+k\delta t,\delta t) - \mu\delta t|^2, |\tilde{R}(t,\delta t) - \mu\delta t|^2\right) &= Cov\left(\sigma^2\tau_{t+k\delta t}|B_k(1)|^2, \sigma^2\tau_t|B_0(1)|^2\right) \\
&= \sigma^4(\mathbb{E}[|B_0(1)|^2])^2 Cov\left(\sqrt{\tau_{t+k\delta t}}, \sqrt{\tau_t}\right) \\
&= \sigma^4 Cov\left(\tau_{t+k\delta t}, \tau_t\right)
\end{aligned}$$

□

In order to get $\tilde{R}(t,\delta t) \sim \mathcal{T}(\nu,\sigma,\mu)$, one needs to have the right distribution of the process T increments. Heyde and Leonenko [110], show that taking the increments of T inverse gamma distributed leads us to the wanted log-prices increments distribution (t-distributed log-prices increments) and so assumption (A_1) is satisfied. We develop next the properties of the process T. We first show that assumption (A_6) is satisfied asymptotically. Taking $T(t)/t \to 1$ almost surely as $t \to +\infty$, one has

$$\mathbb{E}|B(T(t+\delta t)) - B(T(t))| = \mathbb{E}[\mathbb{E}|B(T(\delta t))||T(\delta t) = t] = \sqrt{\delta t}\mathbb{E}\left[\sqrt{\frac{T(\delta t)}{\delta t}}\right]\mathbb{E}|B(1)|$$

Furthermore as $\mathbb{E}[T(\delta t)] < +\infty$, $\mathbb{E}\left[\sqrt{\frac{T(\delta t)}{\delta t}}\right] \to 1$ as $t \to +\infty$. The scaling property is given as δt goes to infinity.

8.3.2.2 Analysis of the Fractal activity time T

We need to generate the fractal activity time so assumptions (A_3),(A_4) and (A_5) are satisfied. Heyde [13] notes that the Fractal activity time process T exhibits long range dependence, which suggests that the time process to have a self-similar or multifractal[8] behaviour. One should have the following relation.

$$T(at) - at \overset{(fdd)}{=} M(a)(T(t) - t)$$

where a is a positive constant and M a random function independent of T. For self-similar behaviour, $M(a)$ is of the form of a^H, where H is the index of self-similarity of the process $T(t) - t$. Assuming for a real p strictly positive $\mathbb{E}|T(t) - t|^p < +\infty$, then for $q \in (0, p]$ one has

$$\mathbb{E}|T(t) - t|^q = a^{qH}\mathbb{E}|T(1) - 1|^q$$

and so H can be estimated by computing the slope of the linear regression of

$$\log(\mathbb{E}|T(t) - t|^q) = qH\log(a) + \log(\mathbb{E}|T(1) - 1|^q)$$

We test this hypothesis on our deseasonalised financial data. Using an Ito's formula, Heyde [13] proposed the following construction of the time process T

$$d\log(\tilde{S}(t)) - \frac{d\tilde{S}(t)}{\tilde{S}(t)} = \frac{\sigma^2}{2}dT(t)$$

where $\log(\tilde{S}(t)) = \tilde{X}(t)$. Taking the prices process $S = \{S(t)\}_{t\in[0,\tau]}$. The construction procedure of the process T is given as follows.

[8] see Appendix A for an introduction to multifractal processes

Set $X_0 := \log(\tilde{S}_0)$
Set $Y_0 := \tilde{S}_0$
Set $T_0 := 0$
Initialise δt
Initialise $k := 1$
REPEAT While $k\delta t \leq \tau$
$\quad$ *Set* $X_k := \log(\tilde{S}_{k\delta t})$
$\quad$ *Set* $Y_k := \tilde{S}_{k\delta t}$
$\quad$ *Set* $T_{k+1} := T_k + X_{k+1} - X_k - \dfrac{Y_{k+1} - Y_k}{Y_k}$
$\quad$ *Set* $k := k + 1$

The scaling properties of the time change process is shown in Figure 8.21 for $DS_{\hat{a}_m^3}$-AUD/USD and $DS_{\hat{a}_{med}^2}$-FTSE100, where $T(t) - t$ was constructed from the price $\tilde{S}(t)$ sampled at 30 minutes interval and $q = 1$. The scaling property is satisfied and we find a scaling exponent of 0.753 and 0.812 respectively for $DS_{\hat{a}_m^3}$-AUD/USD and $DS_{\hat{a}_{med}^2}$-FTSE100 fractal time processes. These correspond to the values mentioned in [13], which is $H \approx 0.8$. However this do not ensure the self-similarity of the process $T(t) - t$ and so to continue, we assume that the fractal activity time of our data satisfies Heyde's hypothesis that the process $T(t) - t$ is self-similar with index $H \in (1/2, 1)$. In that case, as the increments of the process T are stationary and $T(0) = 0$ almost surely, for two positive reals t and s, the covariance function of the process $\tau_t = T(t + \delta t) - T(t)$ is given by

$$Cov\left(\tau_{t+s}, \tau_t\right) = \frac{E|\tau_1|}{2}(|s+1|^{2H} + |s-1|^{2H} - 2|s|^{2H}) \tag{8.15}$$

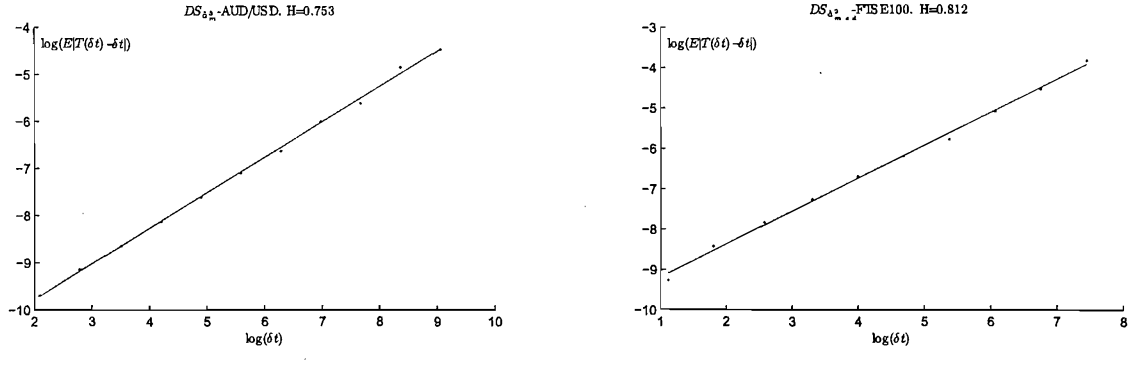

Figure 8.21: Estimation of the self-similarity index of the fractal activity time of the deseasonalised AUD/USD rate index (left panel) and the FTSE100 future contract (right panel)

As shown in [110], since the process T is increasing, $T(t) - t$ cannot be self-similar.

Indeed, for any $s \in (0,1)$, one would have

$$\begin{aligned} P(T(t+s) - T(t) < 0) &= P(T(s) - s < -s) \\ &= P(s^H(T(1) - 1) < -s) \\ &= P(T(1) < 1 - s^{1-H}) > 0 \end{aligned}$$

which contradict that T is increasing. This push Heyde [13] to write

$$T(at) - at \overset{(fdd)}{\cong} a^H(T(t) - t)$$

where $\overset{(fdd)}{\cong}$ stand for asymptotically equal in finite dimensional distribution and affirms the distribution of $T(t)$ can be approximated in the sense of finite dimensional distribution by the distribution of $t + t^H(\tau_1 - 1)$, where $\tau_1 \sim (\nu - 2)\mathrm{R}\Gamma(\nu/2, 1/2)$ so $\mathbb{E}[\tau_1] = 1$. In that case, the process $B(T(1)) \overset{(d)}{=} \sqrt{\tau_1}B(1) \sim \mathcal{T}(\nu, 1, 0)$ (see Appendix B). Using the relation between the inverse Gamma process and the Chi-squared given in Appendix B, the process τ_t can be generated as

$$\tau_t \sim \frac{\nu - 2}{\sum_{i=1}^{\nu} N_i(t)^2}$$

where the $N_i(t)$ are standard normal random processes independent from each other. Taking

$$Cov\left(N_i(t+s), N_i(t)\right) = \frac{1}{(1 + |s|^2)^{\alpha/2}}$$

where $\alpha \in (0, 1/2)$, Heyde and Leonenko [110] shows that we get the appropriate auto-correlation function of the process τ_t given by (8.15) and that the process $\dfrac{T(\lfloor nt \rfloor) - \lfloor nt \rfloor}{n^{1-\alpha}}$ converges weakly to a H-self-similar process $R(t)$ with $H = 1 - \alpha$, as n goes to infinity. It can be shown that [110]

$$\frac{T(\lfloor nt \rfloor) - \lfloor nt \rfloor}{n^H} \overset{(d)}{\to} \frac{1}{\nu} \sum_{i=1}^{\nu} R_i(t) \text{ as } n \to +\infty$$

where R_i are independent Rosenblatt processes. This satisfies the asymptotical self-similarity of the process $T(t) - t$.
The subordinated process $B(T(t))$, with the process T as described above, gave birth to Heyde's FATGBM model, to model stock price

$$\tilde{S}(t) = \tilde{S}(0)e^{\mu t + \sigma B(T(t))}$$

FATGBM stand for Fractal Activity Time Geometric Brownian Motion.

8.3.2.3 Comparison of real data to FATGBM model

We fit the logarithm of Heyde's FATGBM model to the $DS_{\hat{a}_m^3}$-AUD/USD and $DS_{\hat{a}_{med}^2}$-FTSE100 log-price. The FATGBM model is defined by the following parameters : ν, σ, μ and α . These parameters were already estimated in the previous sections. Here we remind their estimate in Table 8.13 for $DS_{\hat{a}_m^3}$-AUD/USD log-price with $\delta t = 30$ minutes and for $DS_{\hat{a}_{med}^2}$-FTSE100 log-price with $\delta t = 60$ minutes.

	$DS_{\hat{a}_m^3}$-AUD/USD				$DS_{\hat{a}_{med}^2}$-FTSE100			
Parameters	ν	σ	μ	α	ν	σ	μ	α
	4	$6.00e^{-4}$	$1.75e^{-5}$	0.25	5	$1.84e^{-3}$	$6.19e^{-5}$	0.19

Table 8.13: FATGBM fitted parameters.

Note that in theory, the FATGBM is valid for $\nu > 4$. However, in practice we allow $\nu = 4$. In Figure 8.22, we represent the $DS_{\hat{a}_m^3}$-AUD/USD and $DS_{\hat{a}_{med}^2}$-FTSE100 log-price and log-prices return. For comparison, we also draw a simulated log FATGBM and their increments. In Figure 8.24, we draw the autocorrelation function of the log-prices increments, as well as the absolute and squared log-prices increments of the $DS_{\hat{a}_m^3}$-AUD/USD and $DS_{\hat{a}_{med}^2}$-FTSE100 log-price with the autocorrelation function of their respective models.

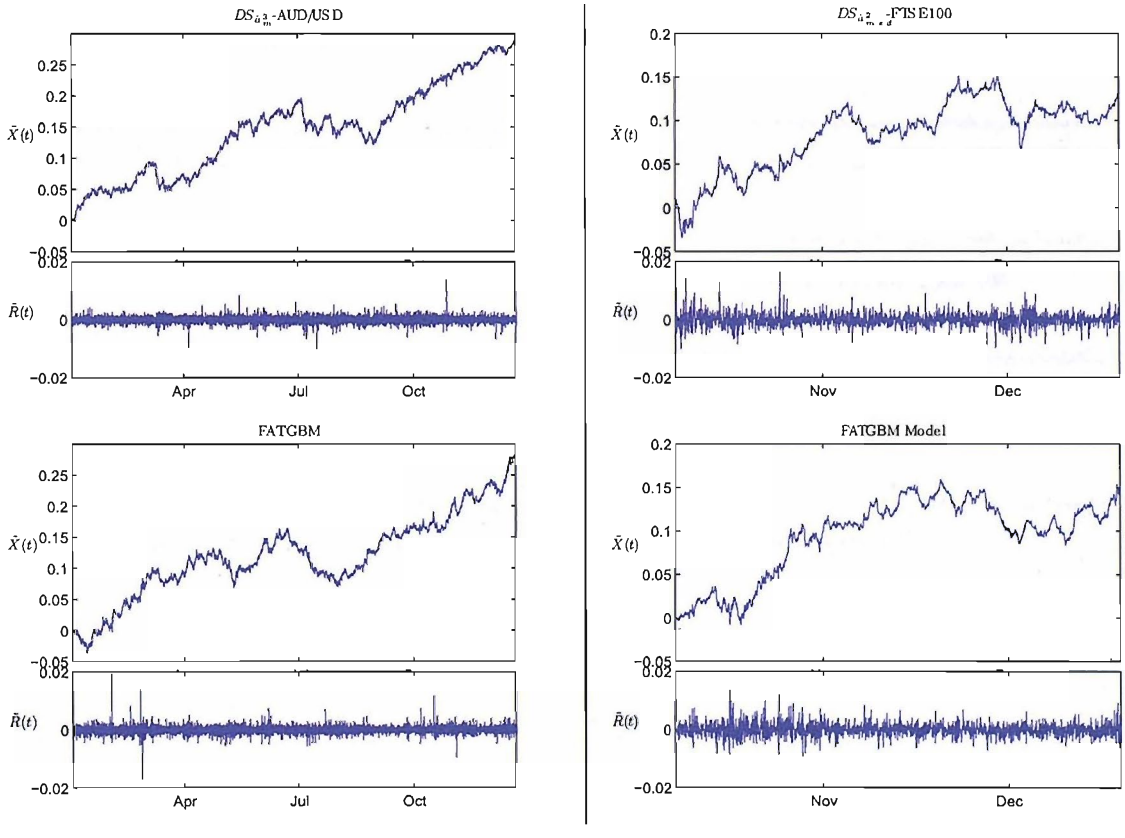

Figure 8.22: $DS_{\hat{a}_m^3}$-AUD/USD (left panels) and $DS_{\hat{a}_{med}^2}$-FTSE100 (right panels) log-prices and log-prices returns, and their calibrated log-FATGBM model below.

The log-prices return and the log FATGBM increments distribution are represented in Figure 8.23. The FATGBM reproduce the distribution of the log prices return of the deseasonalised AUD/USD rate index and the deseasonalised FTSE100 future contract.

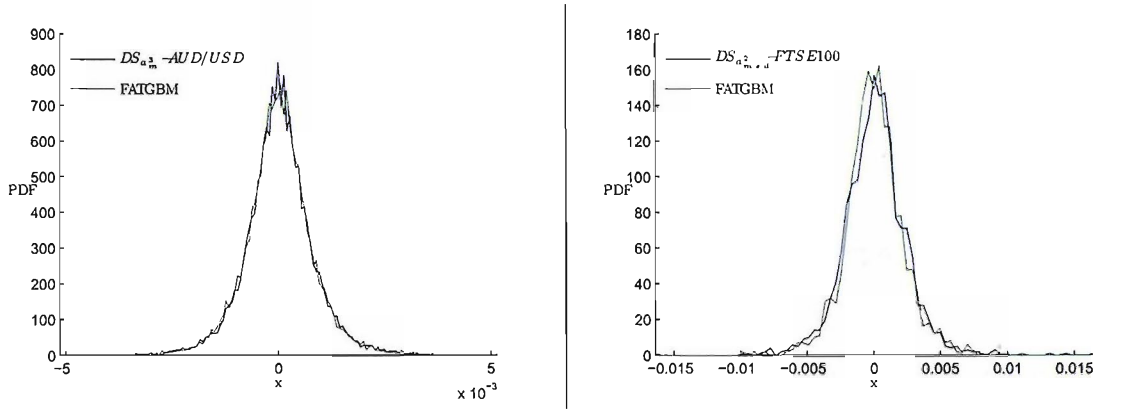

Figure 8.23: Log-prices return distribution of the $DS_{a_m^3} - AUD/USD$ (left panel) and the $DS_{a_{med}^2} - FTSE100$ (right panel) as well as their corresponding log FATGBM increments distribution (in green).

We note also that Heyde's FATGBM model describe well the autocorrelation structure of the data and so it seems to be a good model since it satisfies our assumptions on the log-prices characteristics described previously.

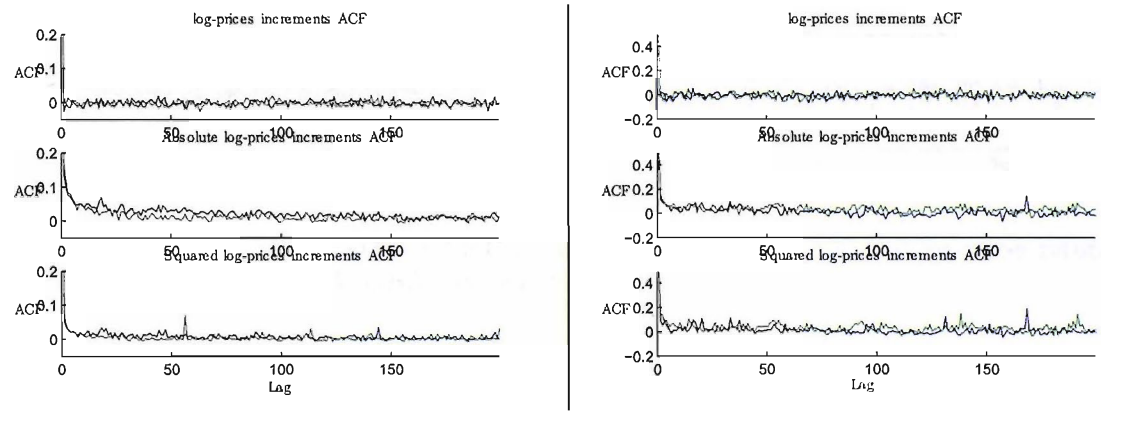

Figure 8.24: ACF of $DS_{\hat{a}_m^3}$-AUD/USD in blue (left panels) and $DS_{\hat{a}_{med}^2}$-FTSE100 in blue (right panels) compare to the ACF of their calibrated log-FATGBM model in green.

8.3.3 Time-changed subordinated Brownian motion

We consider the following time-changed FATGBM, noted θ-FATGBM and defined as

$$S(t) = S(0)e^{\mu\theta(t)+\sigma B(T(\theta(t)))}$$

We draw in Figures 8.25 the log-prices of our data and the log θ-FATGBM given by

$$X(t) = \mu\theta(t) + \sigma B(T(\theta(t)))$$

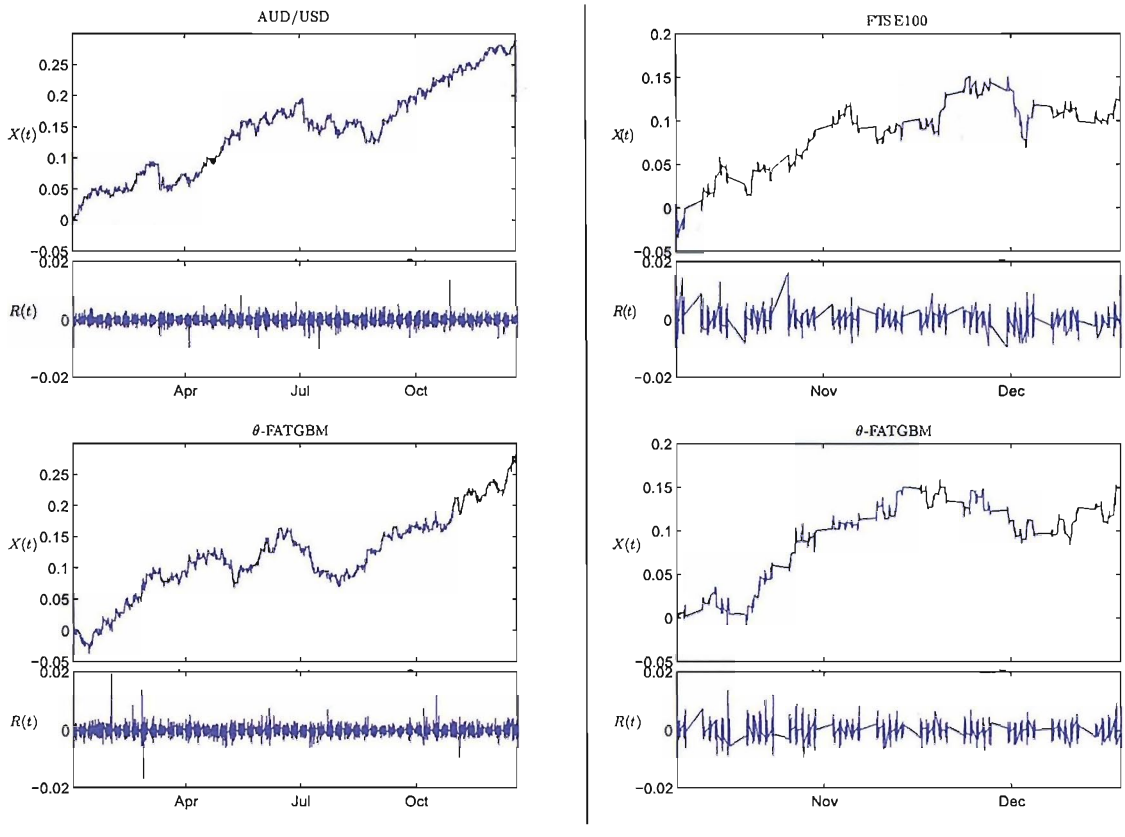

Figure 8.25: AUD/USD (left panels) and FTSE100 (right panels) log-prices and log-prices returns, and their calibrated $\log(\theta - FATGBM)$ model below.

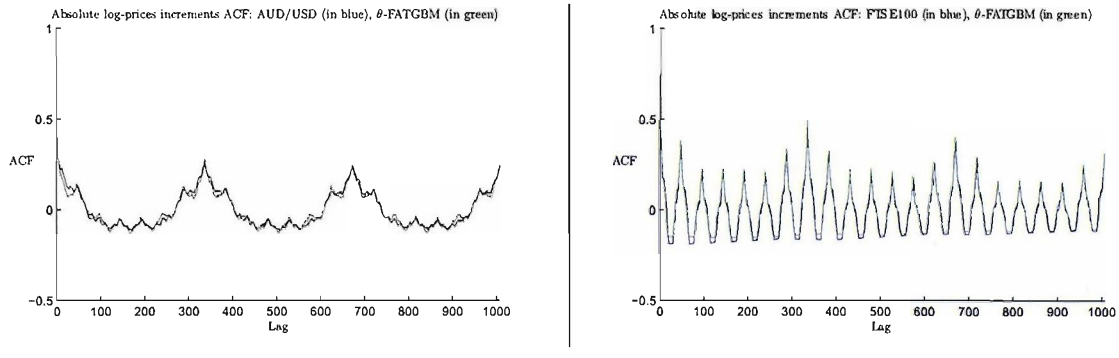

Figure 8.26: ACF of the absolute log-prices increments of AUD/USD (left), FTSE100 (right) and their respective θ-FATGBM.

Note that the weekly and daily periodicity of our high frequency financial data are reproduced into the model. This is confirmed by the autocorrelation function of the absolute log-prices increments and one of the log θ-FATGBM in Figures 8.26. The blue lines represent the ACF of AUD/USD (left panels) and FTSE100 (right panels). The green lines represent the ACF of the θ-FATGBM. Clearly, the periodicity exhibited by the absolute log-prices return of our financial data are also exhibited almost exactly by the time-changed FATGBM.

The θ-FATGBM seems to be a good model for high frequency financial data. It reproduces the characteristic required, which are given at the beginning of Section 8.3.1. Moreover, the θ-FATGBM can be seen as a model that includes a stochastic volatility component, due to the fractal time activity T and a deterministic volatility component intraday seasonality, brought by the function θ.

Conclusion

We assumed a process of the form $X(t) = X(\tilde{\theta}(t))$, where θ is a deterministic function strictly increasing and continuous and $\tilde{X}$ is a self-similar process of index H and has stationary increments.

Assuming we have observations of the process $X(t)$, we wanted to estimate the function θ. To do so, three methods estimating the time-changed function θ were developed; two of them using the path variation ($1/H$-variation and log-variation) and one using level crossings of the process.

For the estimators based on the path variation, the $1/H$-variation is more adapted to continuous sample paths processes such as the fractional Brownian motion. In comparison, the log-variation is adapted to jump processes such as the Lévy stable motion. For both of these estimators, the self-similarity index of $\tilde{X}$ needs to be estimated from the process X. Jones and Shen [7] estimator is theoretically the right estimator to use since it does not depend on the time-change function θ.

The crossing level estimator avoids the problem of estimating H. It has a natural estimate of H, while performing the time-change estimate. Moreover, the level crossings estimator may be used in a more general framework. The process $\tilde{X}$ does not need to be self-similar in that case, however the estimated time change will depend on the size of the level crossing chosen.

For the three estimators, we showed, under H-sssi and ergoditcity assumptions of $\tilde{X}$, convergence in probability to the time change function θ. Their performances have been tested on simulated time changed self-similar processes; in particular on time-changed fractional Brownian motion and time-changed Lévy motion. In conclusion, the three estimators have advantages and inconveniences. Only the $1/H$-variation method is unbiased. The log-variation performs the best with jump processes and the crossing method has the advantage to not need an estimate of the self-similarity index.

When applying the estimator to high frequency financial data, while the path variation estimators use the logarithmic price to estimate the activity (since some high frequency financial data exhibit a scaling property, when taking the logarithmic price), the crossing may be used directly on the prices itself. However, the level crossing estimator was performed on the log prices for the sake of comparison.

Note before applying our estimator on high frequency financial data, the data needed to be cleaned (Chapter 6). One need to remove spurious observations since the estimators are sensitive to outliers for example.

Other advantages: Our estimators do not need the use of years of HDF to estimate the activity, as the estimator of Dacorogna does [3], and so few weeks suffice to give a good estimate of the activity market.
Note in general, for financial data $H \approx 1/2$, so that the $1/H$-variation for $H = 1/2$ correspond to the time change estimator presented by Zhou in [93], which correspond to estimate the local volatile of the data.
Estimating the activity in high-frequency financial data, allows describing the market microstructure as well as removing the intraday seasonality from the data. The way that the time change removes the intraday seasonality is carried out by stretching periods with high fluctuation and compressing periods of low fluctuation (Chapter 7). The estimated time-change, which deform the physical time, is seen as the operational time. Since the estimator assume that the time changed data has stationary increments in the operational, the filtered data are easier to model. To obtain a model in physical time, it suffices to inject the time change function back into the model used in the operational time for the HFD. This procedure was described in Chapter 8, using a time change on Heyde's FATGBM model.
We saw that the de-seasonalised log-prices return follow a Student t-distribution. Also, the analysis of the data at different scales showed that the log-prices exhibit a scaling property. These characteristics are well described by the FATGBM.
Finally, the θ-FATGBM proposed for high frequency financial data in Chapter 8, is in some sense a model which exhibits a stochastic volatility as well as a deterministic volatility, where this last represents the intraday seasonality of the data. This model is a good model since it reproduces the desired periodicity in the autocorrelation function of the absolute logarithmic prices increments.
Further work will consist of, improving the level crossings estimator, and finding a way to make it an unbiased estimator. Also, to compare the performance of the crossing method on estimating the time change function from the prices and the log prices. We can also extend the use of the time-changed process to point processes. This last is an alternative for modelling high frequency financial data (see [84]).
A new topic would be, analysing how trading strategies performed on the operational time may be affected by the time change. Generally this will provoke fast trading during periods of high activity, which may be difficult to perform in reality, since we may end up with two consecutive trading times separated by a few seconds. An improvement of trading strategies may be done using time changed process as a model for high frequency finance. This last topic may be started by looking at the effect of the time change on the liquidity of the market and study how the volume of the transaction may be affected.

References

[1] O.D. Jones. Matlab source codes. http://www.ms.unimelb.edu.au/ odj/. (document), 1.3

[2] O.D. Jones. Fast, efficient on-line simulation of self-similar processes. *In Thinking in Patterns: Fractals and Related Phenomena in Nature, M.M. Novak Ed.*, pages 165–176, 2004. (document), 1.1.4, F.1

[3] M. M. Dacorogna, U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of international money and finance*, 12(4):413–438, 1993. (document), 6.1.4, 6.1.4, 7, 7.1.1, 7.1.2, 8, 8.3.3

[4] Edgar E. Peters. *Fractal Market Analysis.* Wiley, 1994. (document), 1.1.4

[5] C.K. Peng, S.V. Buldyrev, M. Simons, H.E. Stanley, and A.L. Goldberger. Mosaic organization of dna nucleatides. *Physical Review E*, 49:1685–1689, 1994. (document), 1.1.4, 1.1.4

[6] P. Abry and D. Veitch. Wavelet analysis of long range dependent traffic. *Trans. Info. Theory*, 44:2–15, 1998. (document), 1.1.4, 1.1.4, 1.1.4

[7] O.D. Jones and Y. Shen. Estimating the Hurst index of a self-similar process via the crossing tree. *Signal Processing Letters*, 11(4):416–419, 2004. (document), 1.1.4, 1.1.4, 1.1.4, 1.1.4, 3.3, 3.3.1, 5.1, 7.1.2, 7.1.5, 8.3.3

[8] J.H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulations*, 15:1109–1136, 1986. (document), 1.2.5, 1.2.5

[9] I.A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75:918–928, 1980. (document), 1.2.5, 1.2.5

[10] A.S. Paulson, E.W. Holcomb, and R.A. Leitch. The estimation of parameters of stable laws. *Biometrika*, 62:163–170, 1975. (document), 1.2.5

[11] G. Samorodnitsky and M.S. Taqqu. *Stable Non-gaussian Processes: Stochastic Models with Infinite Variance (Stochastic Modeling S.)*. Chapman & Hall/CRC, 1994. (document), 1.2.2, 1.2.2, 1.3.2, 1.3.2, 1.3.2, 1.3.2, 1.3.4, 4.6.2

[12] K. J. Falconer. *Fractal Geometry - Mathematical Foundations and Applications.* John Wiley, 2003. (document), 2.1.1, 2.1.1, 2.1.1

[13] C. C. Heyde. A risky asset model with strong dependence through fractal activity time. *Journal of Applied Probability*, 36(4):1234–1239, 1999. (document), 8.3.2.1, 8.3.2.1, 7, 8.3.2.2, 8.3.2.2

[14] P. Embrechts and M. Maejima. *Selfsimilar Processes.* Princeton University Press, 2002. 1, 1.1.1, 1.1.2, 1.3.2, 1.3.2, 1.3.8, 1.3.2, 1.3.2, 1.4

[15] J. W. Lamperti. Semi-stable stochastic processes. *Transaction of the American Mathematical Society*, 62:62–78, 1962. 1.1.1

[16] W. Vervaat. Sample path properties of self-similar processes with stationary increments. *Ann. Probability*, 13:1–27, 1985. 1.1.2

[17] G.L. O'Brien and W. Vervaat. Marginal distributions of self-similar processes with stationary increments. *Z. Wahrsch. verw. Geb.*, 64:129–138, 1983. 1.1.2

[18] G. Samorodnitsky and M.S. Taqqu. $(1/\alpha)$-self-similar α-stable processes with stationary increments. *J. Multivar. Anal.*, 35:308–313, 1990. 1.1.2

[19] B. Mandelbrot and J. Wallis. Noah, joseph and operational hydrology. *Water Resources Research 4*, 4(2):909–918, 1968. 1.1.3

[20] H. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of the Civil Engineers*, 116:770–799, 1951. 1.1.4

[21] M.S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3:785–798, 1995. 1.1.4, 1.1.4, 1.1.4, 1.1.4

[22] O.D. Jones and Y. Shen. A non-parametric test for self-similarity and stationarity in network traffic. *To appear in Fractals and Engineering*, 2005. 1.1.4, 1.1.4, 7.2.1, 7.4.1

[23] R. Fox and M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14:517–532, 1986. 1.1.4

[24] T. Higuchi. Approach to an irregular time on the basis of fractal theory. *Physica*, D, 31:277–283, 1988. 1.1.4

[25] B. Mandelbrot. Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift fr Wahrscheinlichkeitstheorie*, 31:271–285, 1975. 1.1.4

[26] M.S Taqqu and V. Teverovsky. On estimating the intensity of long-range dependence in finite and infinite variance time series. *A practical guide to heavy tails: statistical techniques and applications*, pages 177 – 217, 1998. 1.1.4, 1.1.4, 2

[27] P. Abry, D. Veitch, and P. Flandrin. Long-range dependence: Revisiting aggregation with wavelets. *Journal of Time Series Anal.*, 19:253–266, 1998. 1.1.4

[28] Z. Burq and O. Jones. Simulation of brownian motion at first passage times. Preprint, 2006 to appear. 1.1.4, 5.4.2, 5.5.2

[29] Jean Bertoin. *Levy Processes (Cambridge Tracts in Mathematics)*. Cambridge University Press, 1998. 1.2.1, 1.2.4, 1.3.3, 1.3.3

[30] Ken-Iti Sato. *Levy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. 1.2.1, 1.3.3

[31] William Feller. *An Introduction to Probability Theory and Its Applications (Probability & Mathematical Statistics S.)*. John Wiley & Sons Inc, 1971. 1.2.3

[32] V. M. Zolotarev. *One-Dimensional Stable Distributions*. American Mathematical Society, 1986. 1.2.1, 1.2.3

[33] J.P. Nolan. An algorithm for evaluating stable densities in zolotarev's (m) parametrization. *Mathematical and Computer Modelling*, 29:229–233, 1999. 1.2.3

[34] J.H. McCulloch and D.B. Panton. Precise tabulation of the maximally-skewed stable distribution and densities. *Computational statistics and data analysis*, 23:307–320, 1996. 1.2.3

[35] J.P. Nolan. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models*, 13:759–774, 1997. 1.2.3, 1.2.5

[36] J.M. Chambers, C.L. Mallows, and B.W. Stuck. A method for simulating stable random variables. *Journal of american statistical association*, 71(354):340–344, 1976. 1.2.4

[37] Rafal Weron. On the chambers-mallows-stuck method for simulating skewed stable random variables. *Statistics and Probability Letters*, 28:165–171, 1996. 1.2.4

[38] Rafal Weron. Correction to: On the chambers-mallows-stuck method for simulating skewed stable random variables. Technical report, 1996. 1.2.4

[39] E.F. Fama and R. Roll. Properties of symmetric stable distributions. *Journal of the American Statistical Association*, 63:817–836, 1968. 1.2.5

[40] E.F. Fama and R. Roll. Parameter estimates for symmetric stable distributions, journal of the american statistical association. *Journal of the American Statistical Association*, 66:331–338, 1971. 1.2.5

[41] R. Arad. Parameter estimation for symetric stable distribution. *International Ecomomic Review*, 21:209–220, 1980. 1.2.5

[42] V. M. Zolotarev. Statistical estimates of the parameters of stable laws. *Mathematical Statistics: Banack Center Publication*, 6:359–376, 1980. 1.2.5

[43] S. J. Press. Estimation in univariate and multivariate stable distribution. *Journal of the American Statistical Association*, 67:842–846, 1972. 1.2.5

[44] I.A. Koutrouvelis. An iterative procedure for the estimation of the parameters of the stable law. *Commun. Statist. - Simulation & Computation*, 10:17–28, 1981. 1.2.5

[45] J. Nolan. Maximum likelihood estimation and diagnostics for stable distribution. 1.2.5

[46] David Freedman. *Brownian motion and diffusion.* San Francisco, Holden-Day, 1971. 1.3.1

[47] N. Wiener. Differentiable space. *Journal of Mathematics and Physics*, 2:131 – 174, 1923. 1.3.1

[48] A.T.A. Wood and G. Chan. An algorithm of simulating stationary gaussian random fields. *Journal of Applied Statistics*, 46:171 – 181, 1997. 1.3.2

[49] Rama Cont and Pete Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, 2003. 1.3.3, C.1

[50] M. Taqqu. Weak convergence to fractional brownian motion and to the rosenblatt process. *Z. Wahrscheinlichkeitstheorie verw. Geb*, 31:287–302, 1975. 1.4

[51] K. J. Falconer. The local structure of random process. *J. London Math. Soc*, 67:657–672, 2003. 2.1, 2.2.1, 2.2.4, 2.2.2

[52] R. Peltier and J. Lévy-Vehel. Multifractional brownian motion: definition and preliminary results. *Rapport de recherche de l'INRIA 2645 1995*, 1996. 2.1

[53] Stéphane Seuret and Jaques Lévy Véhel. The local hölder function of a continuous function. *Projet Fractales, INRIA*, 2002. 2.1, 2.1.2

[54] K. Daoudi. Généralisations des systèmes de fonctions itérées: analyse de la régularité locale et modélisation multifractale des signaux. *Fractales et ondelettes*, 2:129–163, 2002. 2.1

[55] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics S.)*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1997. 2.1.3, 3.1.3

[56] A. Benassi, S. Cohen, and J. Istas. Local self-similarity and the hausdorff dimension. *C. R. Math. Acad. Sci. Paris*, 336(3):267–272, 2003. 2.2.2, 2.3.3

[57] A. Benassi, S. Cohen, and J. Istas. Identification of filtered white noise. *Stoch. Proc. Appl.*, 75:31–49, 1998. 2.3.2, 2.3.2, 2.3.2

[58] A. Benassi, S. Cohen, and J. Istas. Identification and properties of real harmonizable fractional lévy motion. *Bernouilli*, 8(1):97–115, 2002. 2.3.2

[59] A. Benassi, S. Cohen, and J. Istas. Identifying the multifractional function of a gaussian process. *Statistics & Probability Letter*, 39:337–345, 1998. 2.3.2

[60] E. Lukacs. *Stochastic Convergence*. Raytheon Education Company, 1968. 2.3.4

[61] A. Ayache and J. Lévy Véhel. Generalized multifractional brownian motion: Definition and preliminary results. *Fractals - Theory and Applications in Engineering*, pages 17–32, 1999. 2.3.4

[62] A. Ayache, P. Heinrich, L. Marsalle, and C. Suquet. Hölderian random functions. *Fractals in Engineering: new Trends in Theory and Applications*, pages 33–56, 2005. 3.1.1, 3.1.3

[63] J. Rosiński and W.A. Woyczyński. On itô stochastic integration with respect to p-stable motion; inner clock, integrability of sample paths, double and multiple integrals. *Annals of Probability 1*, 4:271–286, 1986. 3.2.1, 3.2.2

[64] A. Janicki and A. Weron. *Simulation and Chaotic Behavior of α-Stable Stochastic processes.* Marcel Dekker Ltd, 1994. 3.2.1, 3.2.1

[65] S. Cohen and J. Istas. A universal estimator of local self-similarity. *Preprint*, 2004. 3.3, 3.3.1, 3.3.1, 3.3.1, 4.3.1

[66] K. Takashima. Sample path properties of ergodic self-similar processes. *Osaka J. Math*, 26:159–189, 1989. 4.2.2

[67] G. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences U.S.A.*, 17:656–660, 1931. 4.2.2, 4.2.2, 4.3.2, 4.3.2

[68] P.Billingsley. *Convergence of probability measure*. Wiley, 1968. 4.2.2, 4.2.2, 4.2.3, 4.3.1, 4.3.2, 4.3.2

[69] D. Ylvisaker. The expected number of zeros of a stationary gaussian process. *The Annals of Mathematical Statistics*, 36(3):1043–1046, 1965. 5.1

[70] H. Cramer and M.R. Leadbetter. *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications.* Dover Publications, 2004. 5.1, 5.2.2

[71] D. Ylvisaker. A note on the absence of tangencies in gaussian sample paths. *The Annals of Mathematical Statistics*, 39(1):261–262, 1968. 5.1

[72] M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes.* Springer, 1983. 5.1.3

[73] L. Breiman. *Probability.* Addison-Wesley Educational Publishers, 1968. 5.2.7

[74] Y. Derriennic. Un théorème ergodique presque sous additif. *The Annals of Probability*, 11(3):669–677, 1983. 5.3, 5.3.1, 5.3.2

[75] D. Foata and A. Fuchs. *Processus stochastiques : Processus de poisson, chanes de Markov et martingales.* Sciences sup, 2004. 5.3.2

[76] R. F. Engle. The econometrics of ultra high frequency data. *Econometrica*, 68:1–22, 2000. 6

[77] T. G. Andersen. Some reflections on analysis of high-frequency data. *Journal of Business and Economic Statistics*, 18:146–153, 2000. 6

[78] C.A.E. Goodhart and M. OHara. High frequency data in financialmarkets: Issues and applications. *Journal of Empirical Finance*, 4:73–114, 1997. 6

[79] R. A. Wood. Market microstructure research databases: History and projections. *Journal of Business and Economic Statistics*, 18:140–145, 2000. 6

[80] C. Gourieroux and J. Jasiak. *Financial Econometrics: Problems, Models, and Methods.* Princeton University Press, 2001. 6

[81] R. S. Tsay. *Analysis of Financial Time Series.* John Wiley & Sons, Inc, 2001. 6, 8

[82] S.J. Taylor. *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, 2005. 6

[83] M. M. Dacarogna, R. Gencay, U. A. Müller, R.B. Olsen, and O.V. Pictet. *An Introduction to High-Frequency Finance.* Academic Press, 2001. 6, 6.1.4, 6.2.1, 6.2.3, 6.3, 6.4.4, 7.1.4, 7.2.2

[84] N. Hautsch. *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models (Lecture Notes in Economics & Mathematical Systems*

S.). Springer-Verlag Berlin and Heidelberg GmbH & Co. K (April 2004), 2004. 6.1.2, 8, 8.3.3

[85] T.G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–158, 1997. 6.1.2

[86] K. Oya. Measurement of volatility of diffusion processes with noisy high frequency data. In A. In Zerger and R.M. Argent, editors, *MODSIM 2005 International Congress on Modelling and Simulation*, pages 940–945. Modelling and Simulation Society of Australia and New Zealand, December 2005. 6.1.2

[87] Yacine Aït-Sahalia. How often to sample a continuous-time process in the presence of market microstructure noisetsay. *Review of Financial Studies*, 18:351–416, 2005. 6.1.2

[88] F. M. Bandi and J. R. Russell. Microstructure noise, realized volatility, and optimal sampling. *University of Chicago*, 2003. 6.1.2

[89] A. Madhavan. Market microstructure: A survey. *Journal of Financial Mathematics*, 3:205–258, August 2000. 6.1.3

[90] M. O'Hara. *Market Microstructure Theory*. Blackwell Publishers, November 1997. 6.1.3

[91] O. Ap Gwilym and C. Sutcliffe. *High-frequency Financial Market Data (Risk Technology Reports)*. Risk Books, 1999. 6.1.4, 7.4.2

[92] E. Ghysels, C. Gourieroux, and J. Jasiak. Trading patterns, time deformation and stochastic volatility in foreign exchange markets. Technical report, Papers 9655, Institut National de la Statistique et des Etudes Economiques, 1996. 6.1.4

[93] B. Zhou. Forecasting foreign exchange rates subject to de-volatilization. *Forecasting Financial Markets: Exchange Rates, Interest Rates and Asset Management*, pages 51–68, 1996. 6.1.4, 8, 8.3.3

[94] T.N. Falkenberry. High frequency data filtering: A review of the issues associated with maintaining and cleaning a high frequency financial database. *http://www.tickdata.com/FilteringWhitePaper.pdf*, 2002. 6.2.1, 6.2.2, 6.2.3

[95] M. Lundin, M. Dacorogna, and U. A. Müller. *Correlation of high-frequency financial time series (in Financial Markets Tick by Tick)*, chapter 4, pages 91–126. Wiley & Sons, 1999. 6.2.1

[96] M. M. Dacorogna, U. A. Müller, C. Jost, O. V. Pictet, R. B. Olsen, and J. Robert Ward. Heterogeneous real-time trading strategies in the foreign exchange market. *The European Journal of Finance*, 1:383–403, 1995. 6.2.1

[97] U. A. Müller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, and C. Morgenegg. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law and intraday analysis. *Journal of Banking and Finance*, 64(4):1189–1208, 1990. 6.2.2, 8.1.1.2

[98] G. Morgan. Cleaning financial data. Originally published by Financial Engineering News, June/July 2002 (www.fenews.com). Numerical Algorithms Group. 6.2.2, 6.2.3

[99] H.G. Green, B. Schmidt, and K. Reher. Algorithms for filtering of market price data. *Computational Intelligence for Financial Engineering (CIFEr), Proceedings of the IEEE/IAFE*, pages 227–231, 1997. 6.2.3

[100] F. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21., 1969. 6.4.4

[101] Stefan Mittnik, Svetlozar T. Rachev, and et al. The distribution of test statistics for outlier detection in heavy-tailed samples. *Mathematical and computer modelling*, 34, 2001. 6.4.4

[102] J.M. Bardet. Testing for the presence of self-similarity of gaussian time series having stationary increments. *Journal of Time Series Analysis*, 21:497–516, 2000. 7.1.5, 7.1.5

[103] J.M. Bardet. Les cours d'actifs financiers sont-ils autosimilaires? *Journal de la société francaise de statistique*, 141:137–148, 2000. 7.1.5

[104] B. Mandelbrot and H.M. Taylor. On the distribution of stock price differences. *Operations Research*, 15(6):10571062, 1967. 8

[105] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41:135–155, 1973. 8, 8.3.2.1

[106] E. Ghysels and J. Jasiak. Stochastic volatility and time deformation: An application to trading volume and leverage effects. Technical report, CIRANO Working Papers 95s-31, CIRANO, 1995. 8

[107] W. Wasserfallen and H. Zimmermann. The behavior of intra-daily exchange rates. *Journal of Banking and Finance*, 9(1):55–72, 1985. 8.1.1.2

[108] C.A.E. Goodhart and L. Figliuoli. Every minute counts in financial markets. *Journal of International Money and Finance*, 1:23–52, 1991. 8.1.2.1

[109] T. G. Andersen and T. Bollerslev. Dm-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. Technical report, NBER Working Papers 5783, National Bureau of Economic Research, Inc, 1996. 8.1.2.1

[110] C.C. Heyde and N.N. Leonenko. Student process. *Adv.appl. prob.*, 37:342–365, 2005. 7, 8.3.2.1, 8.3.2.1, 8.3.2.2, C.1

[111] L. Calvet and A. Fisher. Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics*, 84:381–406, 2002. A.1

[112] L. Calvet, A. Fisher, and B. Mandelbrot. A multifractal model of asset returns. *Cowles Foundation Discussion Paper No. 1164. http://ssrn.com/abstract=78588*, 1997. A.1, A.1

[113] N.A.J. Hastings, B. Peacock, and M. Evans. *Statistical Distributions.* John Wiley & Sons Inc, 1993. B.2, B.3

[114] K. Pearson. Contribution to the mathematical theory of evolution. *Philos. Trans. Roy. Soc. London Ser. A 185*, pages 71–110, 1894. C.1

[115] R. R. Harris and G. K. Kanji. On the use of minimum chi-square estimation. *The Statistician*, 32(4):379–394, 1983. C.2

[116] C. Hsiao. Minimum chi-square. *Encyclopedia of Statistical Sciences. ed. S. Kotz and N. Johnson, (John Wiley)*, 5:518–522, 1985. C.2

[117] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc. London*, A 222:309–368, 1922. C.3

[118] A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949. C.3

[119] N. Sagara. Nonparametric maximum-likelihood estimation of probability measures: Existence and consistency. *Journal of Statistical Planning and Inference*, 133(2):249–271, 2005. C.3

[120] Y. Bar-Shalom. On the asymptotic properties of the maximum likelihood estimate obtained from dependent observations. *J. Royal Statistical Society Ser. B*, 33(1):72–77, 1971. C.3

[121] B. R. Bhat. On the method of maximum-likelihood for dependent observations. *Journal of the Royal Statistical Society. Series B (Methodological),*, 36(1):48–53, 1974. C.3

[122] T. J. Sweeting. Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.*, 8:1375–1381, 1980. C.3

[123] C. Gouriéroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, 1984. C.3

[124] D. Levine. A remark on serial correlation in maximum likelihood. *J. Econometrics*, 23:337–342, 1983. C.3

[125] C. Gouriéroux, A. Monfort, and A. Trognon. A general approach to serial correlation. *Econom. Theory*, 1(1):315–340, 1985. C.3

[126] C. Choirat, C. Hess, and R. Seri. A functional version of the birkhoff ergodic theorem for a normal integrand: A variational approach. *Ann. Probab.*, 31(1):63–92, 2003. C.3

[127] A. Kolmogorov. Sulla determinazione empirica di una legge di distributione. *Giornale dell'Istituto Italiano degli Attuari*, 4:83–91, 1933. D.1

[128] N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distributions for two independent samples. *Bul. Math. de l'Univ. de Moscou*, 2:3–14, 1939. D.1

[129] Z. W. Birnbaum. Numerical tabulation of the distribution of kolmogorov's statistic for finite sample size. *Journal of the American Statistical Association*, 47(259):425–441, 1952. D.1, D.3

[130] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974. D.1, D.2, D.3

[131] T. W. Anderson and D. A. Darling. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23:193–212, 1952. D.2

[132] R. B. D'Agostino and M. A. Stephens. *Goodness-of-FitTechniques.* Marcel Dekker Inc. New York., 1986. D.3

[133] B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7:1–26, 1979. D.3

[134] R.C.H. Cheng, W. Holland, and N.A. Hughes. Selection of input models using bootstrap goodness-of-fit. *Simulation Conference Proceedings, 1996. Winter*, pages 199–206, 1996. D.3

Appendix A

An overview of Multifractal processes

A.1 Definition and properties

Generally, multifractal measures are proposed to model temporal heterogeneity of self-similar processes. Definition A.1.1 of a Multifractal process is given by Calvet and Fisher in [111].

Definition A.1.1. *A random process $X = \{X(t)\}_{t\geq 0}$ is called Multifractal process if it has stationary increments and satisfies*

$$\mathbb{E}(|X(t)|^q) = c(q)t^{\tau(q)+1} \tag{A.1}$$

where for $q > 0$, $\tau(q)$ and $c(q)$ are deterministic functions of q.

$\tau(q)$ is called the scaling function of the Multifractal process. Note for a self-similar process the function τ is linear. Indeed suppose X is H-ss process, one has $\mathbb{E}|X(t)|^q = t^{qH}\mathbb{E}|X(1)|^q$, this implies $\tau(q) = qH - 1$.

Proposition A.1.2. *The function τ is concave.*

Proof. Let $\lambda \in [0,1]$ and q_1, q_2 two strictly positive constant. We set $q = (1-\lambda)q_1 + \lambda q_2$ Using the Hölder inequality one has

$$\mathbb{E}|X(t)|^q \leq \left(\mathbb{E}|X(t)|^{q_1}\right)^{1-\lambda}\left(\mathbb{E}|X(t)|^{q_2}\right)^{\lambda}$$

taking the logarithm and using (A.1), one has

$$\log(c(q)) + \tau(q)\log(t) \leq (1-\lambda)\log(c(q_1)) + \lambda\log(c(q_2)) + ((1-\lambda)\tau(q_1) + \lambda\tau(q_2))\log(t) \tag{A.2}$$

Dividing by $\log t$ for $t < 1$ and tending t to 0, one gets

$$\tau(q) \geq (1-\lambda)\tau(q_1) + \lambda\tau(q_2)$$

this implies the function τ is concave. □

Proposition A.1.3. *The multiscaling property holds for a bounded interval* $\mathrm{T} \subseteq \mathbb{R}^+$.

Proof. Note that the proof of Proposition A.1.2 contains additional information. Considering relation (A.2), taking $t > 1$ and tending t to infinity, one gets

$$\tau(q) \leq (1-\lambda)\tau(q_1) + \lambda\tau(q_2)$$

which contradicts the concavity of τ, we deduce that the multiscaling property holds for a bounded interval $\mathrm{T} \subseteq \mathbb{R}^+$. □

A large class of multiscaling processes can be defined (see [112]). We consider a process X and we assume the existence of an independent positive process $M = \{M(c)\}_{c\geq 0}$ that satisfies

$$\{X(ct)\}_{t\geq 0} \overset{(fdd)}{=} \{M(c)X(t)\}_{t\geq 0},\ \forall t,\ 0 < c \leq 1 \tag{A.3}$$

If $(a,b) \in (0,1]^2$, the process M takes positive values and satisfies [112]

$$\{M(ab)\} \overset{(fdd)}{=} \{M_1(a)M_2(b)\} \tag{A.4}$$

where M_1 and M_2 are two independent copies of M.
From the multiplicative property and the independency of M_1 and M_2, Equation (A.4) yields to $\mathbb{E}|M(ab)|^q = \mathbb{E}|M(a)|^q\mathbb{E}|M(b)|^q$ for all $q > 0$. When these moments are finite, the process M satisfies the scaling relationship $\mathbb{E}[M(c)]^q = c^{\tau(q)+1}$. Using (A.3) one has

$$\mathbb{E}|X(t)|^q = \mathbb{E}|M(t)X(1)|^q = \mathbb{E}|M(t)|^q\mathbb{E}|X(1)|^q = c(q)t^{\tau(q)+1}$$

assuming that X has stationary increments, we deduce by Definition A.1.1 that X is multifractal. By analogy to self-similar processes, one can replace $M(c)$ by $c^{H(c)}$, where $H(c)$ is a stochastic function of c. We can rewrite (A.3) as

$$\{X(ct)\}_{t\geq 0} \overset{(fdd)}{=} \{c^{H(c)}X(t)\}_{t\geq 0} \tag{A.5}$$

Moreover under strict stationarity of the increments of X, one has for $\delta t > 0$

$$\{X(t+c\delta t) - X(t)\}_{t\in \mathrm{T}} \overset{(fdd)}{=} \{c^{H(c)}(X(t+\delta t) - X(t))\}_{t\in \mathrm{T}}$$

A.2 Moments and scaling law

Definition A.2.1. *X is a fractal process if for $t \geq 0$ and $q > 0$, then X satisfies the scaling law*

$$M(q,\delta) = \mathbb{E}[|\Delta X(t,\delta)|^q] \sim C(q)\delta^{\zeta(q)} \tag{A.6}$$

where $\Delta X(t,\delta) = X(t+\delta) - X(t)$ and $C(q)$ and $\zeta(q)$ are deterministic function of q.

The nature of the process X depends on the function ζ, one has the following remark

Remark A.2.2. *Let X a process satisfying (A.6), then*

(i) If the function theta ζ is linear ($\zeta(q) = qH$), then X is a monofractal process, example H-ss process

(ii) If the function theta ζ is non linear, then X is a multifractal process.

Note if the process X has stationary increments and $X(0) \overset{(a.s.)}{=} 0$, then (A.6) becomes for $\delta > 0$

$$\mathbb{E}[|X(\delta)|^q] \sim C(q)\delta^{\zeta(q)} \tag{A.7}$$

Appendix B

Symmetric scaled Student t-distribution

B.1 The symmetric scaled Student t-distribution

The symmetric scaled Student t-distribution $\mathcal{T}(\nu, \sigma, \mu)$ is represented by the following density function (σ and μ represent respectively the scale and the shift parameter).

$$f(x) = \frac{\Gamma\left((\nu+1)/2\right)}{\sigma\sqrt{\pi}\Gamma(\nu/2)\left(1+\left((x-\mu)/\sigma\right)^2\right)^{(\nu+1)/2}}$$

The parameter ν is called the degree. Most studies in the literature tend to take ν as an integer; but in fact ν belong to the positive real line. For $\nu = \infty$ the student distribution becomes the normal distribution.

B.2 Other distributions related to the t-distribution

We present three distributions that are related (closely or not) to the t-distribution. Gamma, Inverse Gamma and Chi-squared. The distributions and their links to each other are taken from [113].

- The Gamma distribution $\Gamma(\alpha, \beta)$ is defined by the following density function

$$f_\Gamma(x; \alpha, \beta) = x^{\alpha-1}\frac{e^{-x/\beta}}{\beta^\alpha\,\Gamma(\alpha)}$$

  where the strictly positive reals α and β are respectively the shape and the scale parameters. A random variable gamma distributed has mean $\alpha\beta$ and variance $\alpha\beta^2$ for $\beta > 2$.

- The Inverse Gamma distribution $\mathrm{R}\Gamma(\alpha, \beta)$ is defined by the following density function

$$f_{\mathrm{R}\Gamma}(x; \alpha, \beta) = x^{-\alpha-1}\frac{\beta^\alpha e^{-\beta/x}}{\Gamma(\alpha)}$$

where the strictly positive reals α and β are respectively the shape and the scale parameters. A random variable inverse gamma distributed has mean $\beta/(\alpha-1)$ and variance $\beta^2/(\alpha-1)^2(\alpha-2)$ for $\beta > 2$. This distribution is related to the Gamma distribution by $\mathrm{R}\Gamma(\alpha,\beta) \sim 1/\Gamma(\alpha, 1/\beta)$.

- Chi-squared distribution $\chi^2(\nu)$

$$f_{\chi^2}(x;\nu) = x^{\nu/2-1}\frac{e^{-x/2}}{2^{\nu/2}\,\Gamma(\nu/2)}$$

where the strictly positive real ν is the degree of freedom. The link to the Gamma distribution is $\chi^2(\nu) \sim \Gamma(\nu/2, 2)$ and to the inverse Gamma is $\chi^2(\nu) \sim 1/\mathrm{R}\Gamma(\nu/2, 1/2)$. Note that $\chi^2(\nu)$ has same the distribution as the sum of the squares of ν-independent unit normal random variables.

B.3 Generating scaled Student t-distribution random variables

The random variable $X \sim \mathcal{T}(\nu, 1, 0)$ has the representation (see for example [113] page 146)

$$X \overset{(d)}{=} \frac{Y}{\sqrt{Z/\nu}}$$

where Y has normal standard distribution $\mathcal{N}(0,1)$ and Z has chi-square distribution with ν degrees of freedom independent of Y. $X' \sim \mathcal{T}(\nu, \sigma, \mu)$ is obtained by the following representation

$$X' = \mu + \sigma X$$

Another representation can be given by $X' \overset{(d)}{=} \mu + \sigma\sqrt{V}Y$, where V, independent of Y, has mean 1 and $V \sim \dfrac{\nu-2}{\nu}\mathrm{R}\Gamma(\nu/2, \nu/2)$.

B.4 Moments of $\mathcal{T}(\nu, \sigma, \mu)$ random variable

The n^{th} moment $\mu_n = E[(X-\mu)^2]$ of $\mathcal{T}(\nu, \sigma, \mu)$ random variable are given by

$$X = \begin{cases} \mu_n = 0 & \text{if } n < \nu \text{ and } n \text{ odd} \\ \mu_n = \sigma^n \dfrac{1.3.5\ldots(n-1)\nu^{n/2}}{(\nu-2)(\nu-4)\ldots(\nu-n)} & \text{if } n < \nu \text{ and } n \text{ even} \\ \mu_n = \infty & \text{if } n \geq \nu \end{cases} \tag{B.1}$$

Appendix C

Estimating density parameters

Several technics can be found in the literature to estimate density parameters. The main estimators are: the Method of Moments (MM), the Minimum chi-square (MCS) estimation, and the maximum likelihood estimator MLE.

C.1 Method of moments

The method of moments was introduced by Pearson [114]. Generally, only moments of order one to four of the observed data are chosen as these moments are well defined. The method consists on matching empirical moments with theoretical moments. It can be applied to dependent random variables, generally assumed to be stationary and ergodic for which the sum of the sample powered by n divided by the sample size converges to its theoretical n^{th} moment, as long as this last exists.
For example, for the symmetric scaled t-distribution, one uses (B.1) for $n = 1, \ldots, 4$ to estimate the density parameters. However, the disadvantage of this method for the t-distribution, is that it cannot be applied to sample data for which its empirical distribution has a degree of freedom less than or equal to 4, since the 4^{th} moment of such sample does not exists. For financial data, generally the logarithmic price return follows a t-distribution with $2 < \nu < 4$ (see for example [49] or [110]).

C.2 Minimum chi-square estimator

The Minimum Chi-Square (MCS) estimator (see for example [115] and [116]) consists on minimising the following statistics

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where θ is the density parameters, O_i the number of observations falling within the i^{th} $\frac{1}{n}$% sample quantile and E_i the expected number of observations falling within the same quantile.

For large sample data, this method can be applied to dependent observations assumed to be stationary and ergodic [116]. However, the disadvantage of this method is that it performs poorly for continuous random variables and the choice of n for the percentage quantile may affect the estimator.

C.3 Maximum likelihood estimator

The Maximum likelihood estimator (MLE) method was introduced by Fisher [117]. Let $X_1, \ldots, X_n$ with joint distribution $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta)$ where θ is the parameter of the distribution. The likelihood function is defined by

$$L(\theta|x_1, \ldots, x_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) \tag{C.1}$$

The method of maximum likelihood estimates θ by finding the value of θ that maximises $L(\theta|x_1, \ldots, x_n)$. Assuming $X_1, \ldots, X_n$ are i.i.d. with common density function f, (C.1) becomes

$$L(\theta|x_1, \ldots, x_n) = \prod_{j=1}^{n} f(x_1|\theta) \tag{C.2}$$

If the sequence $X_1, \ldots, X_n$ are not i.i.d. then (C.1) can be rewritten as

$$L(\theta|x_1, \ldots, x_n) = f(x_n|x_{n-1}, \ldots, x_1; \theta) f(x_{n-1}|x_{n-2}, \ldots, x_1; \theta) \ldots f(x_1|\theta) f(x_1|\theta) \tag{C.3}$$

To maximise the likelihood function, one maximise its logarithm function. For example, using the log of (C.2), one gets

$$\log(L(\theta|x_1, \ldots, x_n)) = \sum_{j=1}^{n} \log(f(x_1|\theta)) \tag{C.4}$$

Wald in [118], shows that in the case of i.i.d. sequence, the MLE estimator is strongly consistent and asymptotically normal. For dependent observations, the likelihood function is not the product of the densities. However, for large samples, some assumption can be made to show the consistency of the MLE. Assuming the observations are stationary and ergodic, Sagara [119] shows that the MLE is consistent; He uses the Birkhoff's ergodic Theorem and the martingale convergence Theorem. For previous work on MLE for dependent observations, we refer to Bar-Shalom [120] and Bhat [121]. For a survey on the application of the MLE on more general processes see Sweeting [122]. In practice, one can use (C.4) for dependent observations with some additional assumptions. However (C.4) will not be called likelihood function but "pseudo"-likelihood function. The term "pseudo" is given for example by Gouriéroux et al. in [123]. We also find the term "quasi" instead of "pseudo" in some econometrical papers. The principe is similar to the MLE. We maximise the pseudo-likelihood function to esti-

mate θ. Under some assumptions of the dependent observations, Levine [124] and Gouriéroux et al. [125] showed the consistency of the the maximum pseudo likelihood estimator MPLE. Assuming stationary and ergodic observations, Choirat et al [126] showed that the MPLE is consistent.

C.4 Comparison

Here we present a comparison of the estimators. We generate 10000 independent random variables t-distributed, from which we estimate the parameters ν, σ and μ using the three estimators described above. We repeat the process 50 times in order to get the mean and twice the standard deviation of the estimator. The results of the simulation are shown below in Table C.1. The mean, and twice the standard deviation into brackets are displayed for some parameters. We note the good performance of the MLE and the bad performance of the MM for $\nu \leq 4$.

Methods	ν	σ	μ
True parameters	10	20	5
MM	10.188(2.101)	20.022(0.531)	5.012(0.517)
MCS	9.160(1.464)	20.165(0.453)	5.000(0.505)
MLE	9.964 (1.841)	19.972(0.463)	5.002(0.488)
True parameters	4	1	0
MM	4.683(0.719)	1.070(0.042)	0.000(0.034)
MCS	3.990(0.812)	1.076(0.130)	0.000(0.090)
MLE	4.001(0.399)	1.000(0.026)	0.002(0.026)
True parameters	2	0.5	-0.01
MM	4.032(0.061)	1.130(0.4673)	-0.009(0.031)
MCS	2.144(0.893)	0.395(1.2990)	0.044(0.526)
MLE	2.025(0.071)	0.502(0.0117)	-0.010(0.021)

Table C.1: Comparison of MM, MCS and MLE on t-distributed sample data.

Appendix D

Goodness of fit

We present two statistical tests the Kolmorov-Smirnov test and the Anderson-Darling test. In both tests, the null hypothesis H_0 is "The data follow the specified distribution". H_1: " The data do not follow the specified distribution". We consider $\{X_i\}_{i\in\{1,\dots,n\}}$ the observations used to define the the tests.

D.1 Kolmogorov-Smirnov test

The Kolmogorov test [127] measures the maximum absolute distance between the empirical distribution function F_n of a sample and the empirical distribution function of a theoretical function F

$$D_n = \sup_x |F(x) - F_n(x)| \tag{D.1}$$

This test is often called Kolmogorov-Smirnov test. In fact Smirnov in [128], presents an extension of the Kolmogorov test. The test is based on Kolmogorov statistics and has the aim to compare two independent samples that can be of different size. This test tells us whether the two samples have the same distribution. Assuming F_n^1 and F_m^2 are the empirical distribution of both samples respectively of size n and m, Smirnov statistics is given by

$$D_{n,m} = \sup_x |F_n^1(x) - F_m^2(x)| \tag{D.2}$$

The Kolmogorov-Smirnov test in both cases can be applied only on random variables for which the empirical distribution function is continuous. The null hypothesis H_0 is rejected if $D_{n,m}$ is greater than the critical value. Tabled values can be found in Birnbaum [129]. A more applicable formula of the Kolmogorov-Smirnov statistics, which can be found in [130], is given by

$$D_n = \max_{i\in\{0,\dots,n\}} \left(F_n(X_i) - \frac{i-1}{n}, \frac{i}{n} - F_n(X_i) \right) \tag{D.3}$$

Where $F_n(X_i)$ is the value of the fitted empirical distribution at the i^{th} ordered observation.

D.2 Anderson-Darling test

Anderson-Darling test [131] measures the quadratic deviations between the empirical distribution function F_n of a sample, and the empirical distribution function of a theoretical function F, multiplied by a weighting function $\psi(x)$

$$A^2 = \int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 \psi(x) \mathrm{dF} \tag{D.4}$$

Where $\psi(x) = \dfrac{1}{F_n(x)(1 - F_n(x))}$. This test has the advantage of being more sensitive to the tail of distributions (e.g. α-stable distribution, Student t-distribution). As the Anderson-Darling test makes use of the specific distribution to calculate the critical value, this last must be calculated for each distribution. Tabulated values have been published for a few specific distributions by Stephens [130]. If the test statistic A^2 is greater than the critical value, then the null hypothesis H_0 is rejected.
A more applicable expression of (D.4) is given by (this can be found in [130] for example)

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} (\log(F_n(X_i)) + \log(1 - F_n(X_{n+1-i}))) \tag{D.5}$$

Where $F_n(X_i)$ is the value of the fitted empirical distribution at the i^{th} ordered observation. Note that the Anderson-Darling test statistics can be applied for both discrete and continuous distributions.

D.3 Bootstrap method goodness of fit

Critical values of the Anderson-Darling and the Kolmogorov-Smirnov statistic for some distributions can be found in the literature, see for example [129], [130] and [132]. Unfortunately, for the t-distribution and α-stable distribution we could not find them. However one can estimate the statistics of this test using a Monte Carlo simulation. This is the case when the distribution of the sample are known.
The distribution of observed real data, such as financial data, are unknown. Generally we assume that the sample have a distribution $F(X;\theta)$ and we estimate θ (using one of the methods cited in C) for which $\hat{F} = F(X;\hat{\theta})$ fits better the empirical distribution of the observed data. The question is how well the distribution fits the one of the sample. In that case we use a bootstrap method.
The bootstrap method introduced by Efron [133] is a general resampling procedure for estimating the distributions of statistics based on the observations. The bootstrap goodness of fit [134] is used to check if the a sample observations come from a particular distribution. Let $F(x,\theta)$ be the distribution chosen to fit our sample distribution. The procedure consists of (see [134]):

1 Estimate θ from the sample. The estimator is given by $\hat{\theta}$.

2 Calculate the goodness of fit statistic T for the given distribution $F(x, \hat{\theta})$.

3 Use bootstrap sampling to estimate the distribution of the sample T.

   a Generate B sample $X^{(i)}$ from the fitted distribution $F(x, \hat{\theta})$.

   b Fit $F(x, \theta)$, by computing $\hat{\theta}^{(i)}$ an estimator of θ for each sample $X^{(i)}$.

   c Calculate the goodness of fit statistic $T^{(i)}$, for the fitted $F(x, \hat{\theta}^{(i)})$.

   d From the EDF of $T^{(i)}$ we compute the p-value p of T. The model with the smallest value of p is selected.

$$p = \frac{\#T^{(j)} \leq T}{B}$$

Note that as B is bigger as the precision of the p-value increases. One needs $B > 500$, for a good estimate of p.

Appendix E

Time-changed self-similar processes and diffusion processes

We present a short overview on how to include intraday seasonality using the time-change function θ to some well known asset prices models.

E.1 Motivation

Assuming the basic model of an asset price in θ-time $\tilde{S}(t)$ is given by the following Stochastic Differential Equation (SDE)

$$d\tilde{S}(t) = \mu\tilde{S}(t)dt + \sigma\tilde{S}(t)dW(t) \tag{E.1}$$

where the drift μ and the volatility σ are constants and $W(t)$ is a Brownian motion. The solution of (E.1) is

$$\tilde{S}(t) = \tilde{S}(0)\exp\left(\mu - \sigma^2/2\right)t + \sigma W(t)$$

The asset price $S(t)$ in calendar time is then

$$S(t) = \tilde{S}(\theta(t)) = \tilde{S}(0)\exp\left((\mu - \sigma^2/2)\theta(t) + \sigma W(\theta(t))\right) \tag{E.2}$$

The deterministic function θ can be estimated from the log-prices $X(t) = \log(S(t)/\tilde{S}(0))$. The process X is locally self-similar if $\mu - \sigma^2/2 \neq 0$, and self-similar otherwise, in both cases the index of self-similarity is $1/2$, which is the self-similarity index of the Brownian motion W. Moreover, X has independent increments and so using Theorem in Chapter 4, one can estimate the time function θ from X. So the asset price model represented by (E.2) can be applied to high frequency financial data, where the function θ represents the intraday seasonality component of the model.

E.2 Time changed Asset price model as a deterministic volatility model

The time change function θ in the asset price model (E.2), can be seen as a time-dependent volatility model represented by the following SDE

$$dS_v(t) = \mu_v S_v(t)dt + \sigma_v(t)S_v(t)dW(t) \tag{E.3}$$

where μ_v is a constant and σ_v is a deterministic function of time, and for which the solution is

$$S_v(t) = S_v(0)\exp\left(\mu_v t - \frac{1}{2}\int_0^t \sigma_v^2(u)du + \int_0^t \sigma_v(u)dW(u)\right) \tag{E.4}$$

Using an integral representation of the time-changed Brownian in (E.2), one has

$$S(t) = \tilde{S}(0)\exp\left((\mu - \sigma^2/2)\int_0^t \theta'(u)du + \sigma\int_0^t \sqrt{\theta'(u)}dW^*(u)\right) \tag{E.5}$$

By analogy to (E.4), we find

$$\sigma_v(t) = \sigma\sqrt{\theta'(t)} \text{ and } \mu_v = \mu\theta(t)$$

Note in that case μ_v is time-depend function.

Appendix F

Appendices related to Chapters

F.1 Formation of the crossing tree (Chapter 1)

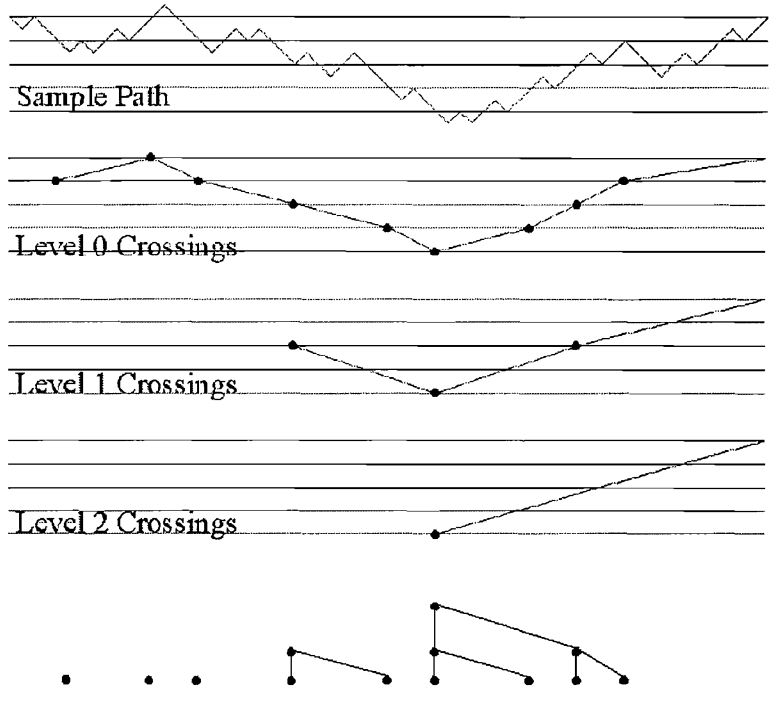

Figure F.1: Construction of the crossing tree for the EBP estimator. Graph taken from [2]

F.2 DuMouchel Tables (Chapter 1)

$\alpha = \Psi_1(\nu_\alpha, \nu_\beta)$

	ν_β						
ν_α	0.00000	0.10000	0.20000	0.30000	0.50000	0.70000	1.0000
2.4390	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
2.5000	1.9160	1.9240	1.9240	1.9240	1.9240	1.9240	1.9240
2.6000	1.8800	1.8130	1.8290	1.8290	1.8290	1.8290	1.8290
2.7000	1.7290	1.7300	1.7370	1.7450	1.7450	1.7450	1.7450
2.8000	1.6640	1.6630	1.6630	1.6680	1.6760	1.6760	1.6760
3.0000	1.5630	1.5600	1.5530	1.5480	1.5470	1.5470	1.5470
3.2000	1.4840	1.4800	1.4710	1.4600	1.4480	1.4380	1.4380
3.5000	1.3910	1.3860	1.3780	1.3640	1.3370	1.3180	1.3180
4.0000	1.2790	1.2730	1.2660	1.2500	1.2100	1.1840	1.1500
5.0000	1.1280	1.2100	1.1140	1.1010	1.0670	1.0270	0.97300
6.0000	1.0290	1.0210	1.0140	1.0040	0.97400	0.93500	0.87400
8.0000	0.89600	0.89200	0.88700	0.88300	0.85500	0.82300	0.76900
10.000	0.81800	0.81200	0.80600	0.80100	0.78000	0.75600	0.69100
15.000	0.69800	0.69500	0.69200	0.68900	0.67600	0.65600	0.59500
25.000	0.59300	0.59000	0.58800	0.58600	0.57900	0.56300	0.51300

Table F.1: DuMouchel Table to estimate α, note that $\Psi_1(\nu_\alpha, -\nu_\beta) = \Psi_1(\nu_\alpha, \nu_\beta)$.

$\beta = \Psi_2(\nu_\alpha, \nu_\beta)$

	ν_β						
ν_α	0.00000	0.10000	0.20000	0.30000	0.50000	0.70000	1.0000
2.4390	0.00000	2.1600	1.0000	1.0000	1.0000	1.0000	1.0000
2.5000	0.00000	1.5920	3.3900	1.0000	1.0000	1.0000	1.0000
2.6000	0.00000	0.75900	1.8000	1.0000	1.0000	1.0000	1.0000
2.7000	0.00000	0.48200	1.0480	1.6940	1.0000	1.0000	1.0000
2.8000	0.00000	0.36000	0.76000	1.2320	2.2290	1.0000	1.0000
3.0000	0.00000	0.25300	0.51800	0.82300	1.5750	1.0000	1.0000
3.2000	0.00000	0.20300	0.41000	0.63200	1.2440	1.9060	1.0000
3.5000	0.00000	0.16500	0.33200	0.49900	0.94300	1.5600	1.0000
4.0000	0.00000	0.13600	0.27100	0.40400	0.68900	1.2300	2.1950
5.0000	0.00000	0.10900	0.21600	0.32300	0.53900	0.82700	1.9170
6.0000	0.00000	0.096000	0.19000	0.28400	0.47200	0.69300	1.7590
8.0000	0.00000	0.082000	0.16300	0.24300	0.41200	0.60100	1.5960
10.000	0.00000	0.074000	0.14700	0.22000	0.37700	0.54600	1.4820
15.000	0.00000	0.064000	0.12800	0.19100	0.33000	0.47800	1.3620
25.000	0.00000	0.056000	0.11200	0.16700	0.28500	0.42800	1.2740

Table F.2: DuMouchel Table to estimate β, note that $\Psi_2(\nu_\alpha, -\nu_\beta) = -\Psi_2(\nu_\alpha, \nu_\beta)$.

$\nu_c = \Phi_3(\alpha, \beta)$

	β				
α	0.00000	0.25000	0.50000	0.75000	1.0000
2.0000	1.9080	1.9080	1.9080	1.9080	1.9080
1.9000	1.9140	1.9150	1.9160	1.9180	1.9210
1.8000	1.9210	1.9220	1.9270	1.9360	1.9470
1.7000	1.9270	1.9300	1.9430	1.9610	1.9870
1.6000	1.9330	1.9400	1.9620	1.9970	2.0430
1.5000	1.9390	1.9520	1.9880	2.0450	2.1160
1.4000	1.9460	1.9670	2.0220	2.1060	2.2110
1.3000	1.9550	1.9840	2.0670	2.1880	2.3330
1.2000	1.9650	2.0070	2.1250	2.2940	2.4910
1.1000	1.9800	2.0400	2.2050	2.4350	2.6960
1.0000	2.0000	2.0850	2.3110	2.6240	2.9730
0.90000	2.0400	2.1490	2.4610	2.8860	3.3560
0.80000	2.0980	2.2440	2.6760	3.2650	3.9120
0.70000	2.1890	2.3920	3.0040	3.8440	4.7750
0.60000	2.3370	2.6350	3.5420	4.8080	6.2470
0.50000	2.5880	3.0730	4.5340	6.6360	9.1440

Table F.3: DuMouchel Table to estimate ν_c, note that $\Phi_3(\alpha, -\beta) = \Phi_3(\alpha, \beta)$.

$\nu_\mu = \Phi_4(\alpha, \beta)$

	β				
α	0.00000	0.25000	0.50000	0.75000	1.0000
2.0000	0.00000	0.00000	0.00000	0.00000	0.00000
1.9000	0.00000	−0.017000	−0.032000	−0.049000	−0.064000
1.8000	0.00000	−0.030000	−0.061000	−0.092000	−0.12300
1.7000	0.00000	−0.043000	−0.088000	−0.13200	−0.17900
1.6000	0.00000	−0.056000	−0.11100	−0.17000	−0.23200
1.5000	0.00000	−0.066000	−0.13400	−0.20600	−0.28300
1.4000	0.00000	−0.075000	−0.15400	−0.24100	−0.33500
1.3000	0.00000	−0.084000	−0.17300	−0.27600	−0.39000
1.2000	0.00000	−0.090000	−0.19200	−0.31000	−0.44700
1.1000	0.00000	−0.095000	−0.20800	−0.34600	−0.50800
1.0000	0.00000	−0.098000	−0.22300	−0.38300	−0.57600
0.90000	0.00000	−0.099000	−0.23700	−0.42400	−0.65200
0.80000	0.00000	−0.096000	−0.25000	−0.46900	−0.74200
0.70000	0.00000	−0.089000	−0.51200	−0.52000	−0.85300
0.60000	0.00000	−0.078000	−0.27200	−0.58100	−0.99700
0.50000	0.00000	−0.061000	−0.27900	−0.65900	−1.1980

Table F.4: DuMouchel Table to estimate ν_μ, note that $\Phi_4(\alpha, -\beta) = -\Phi_4(\alpha, \beta)$.

F.3 Appendixes of Chapter 6

F.3.1 MATLAB GUI UHFD FX Data cleaner

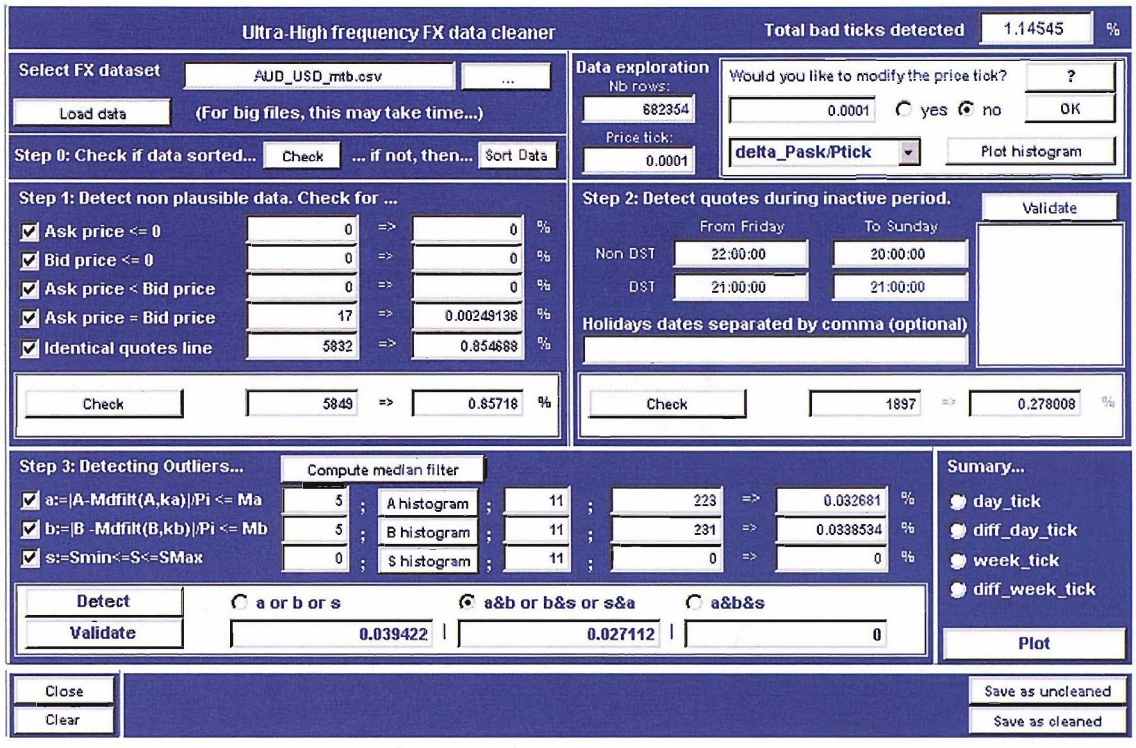

Figure F.2: MATLAB GUI: UHFD FX Data cleaner.

Select FX dataset: Load the UHFD dataset file. The data must be in the following format: time in Matlab number, bid price, ask price.

Step0: Check if dataset is sorted. If not then sort data.

Step1: Detect non plausible ticks.

Step2: Detect quotes during inactive periods (Week-ends and Holidays (optional)).

Step3: Detect outliers.

Data exploration: Shows the nuber of row in the dataset, the Pip value of the data. This last can be wrong so one can change this value. We can also plot some Histogram of price increments (bid and ask) and also the bid ask spread.

Summary: At any moment this block can be used. It is used to check graphically the number of quotes removed after each step.

F.3.2 Trading hours of some markets (Chapter 6)

	Market	Pre-Open	Trading hours	After hours Trading	DST	Non-DST
Asian Markets	**ASX**	07:00-10:00	10:00-16:00	16:12-17:00	GMT+10	GMT+11
	TSE	-	09:00-11:00 & 13:00-15:00	-	GMT+9	GMT+9
	HKEx	-	10:00-12:30 & 14:30-16:00	-	GMT+8	GMT+8
	SGX	08.30-09.00	09:00-12:30 & 14:00-17:00	-	GMT+8	GMT+8
European Markets	**Euronext**	07:00-09:00	09:00-17:30	-	GMT+2	GMT+1
	LSE	-	08:00-16:30	-	GMT+1	GMT+0
	SWX	-	08:30 - 17:30	-	GMT+2	GMT+1
American American	**TSX**	-	09:30-16:00	16:15-17:00	GMT-4	GMT-5
	AMEX	08:00-09:30	09:30 -16:00	-	GMT-4	GMT-5

Table F.5: Trading hours for some markets over the world

Asian Markets (Australia, Hong Kong, Japan, Singapore)	**European Markets** (Europe, Switzerland, United Kingdom)	**American Markets** (Canada, United states)
Australian Stock exchange Sydney (ASX)	Euronext West. Europe	Toronto Stock exchange (TSX)
Tokyo Stock exchange (TSE)	London Stock Exchange (LSE)	American Exchange (AMEX)
Hong Kong Stock exchange (HKEx)	Swiss Exchange (SWX)	
Singapore Exchange (SGX)		

F.4 Activity estimation in HFD data (Chapter 7)

F.5 Activity estimation in UHFD data

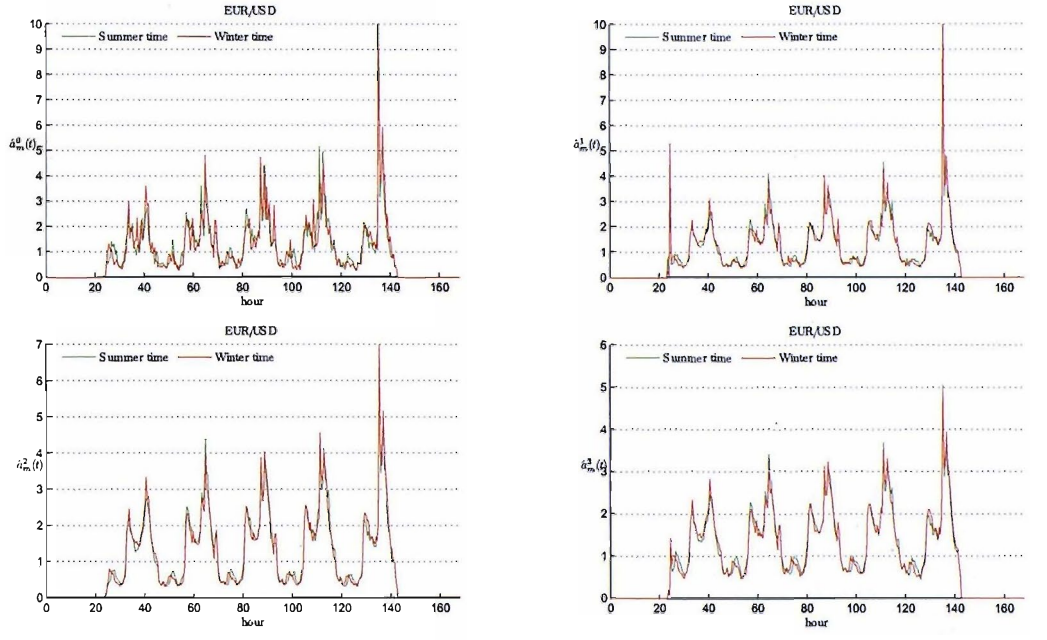

Figure F.3: Activity estimation of the EUR/USD rate index taking the averaged activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

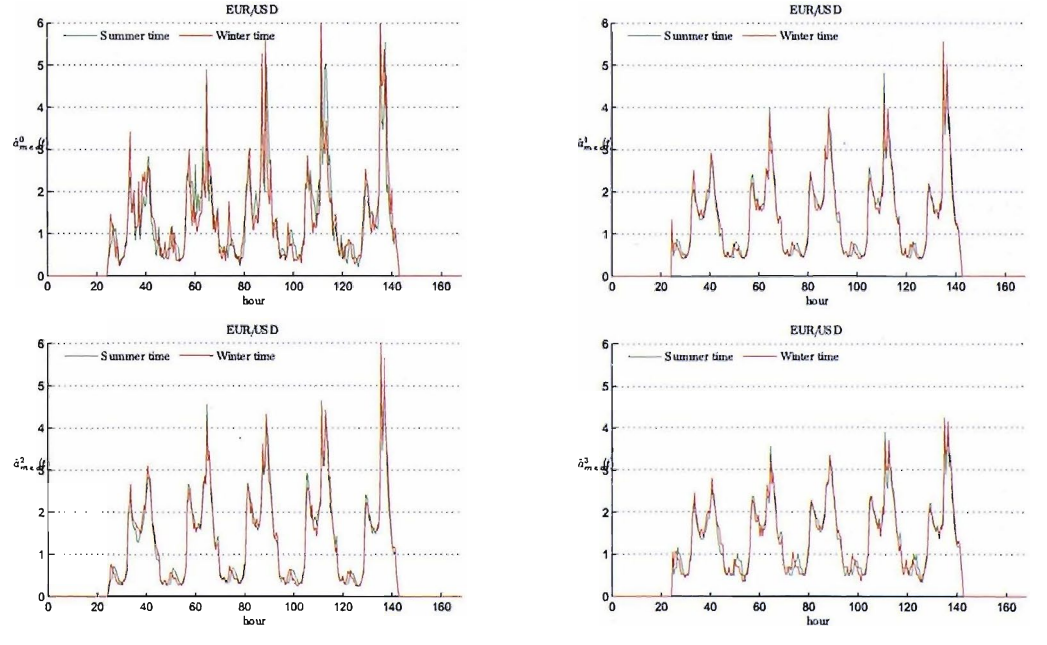

Figure F.4: Activity estimation of the EUR/USD rate index taking the median activity estimated by: the scaling law method(top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

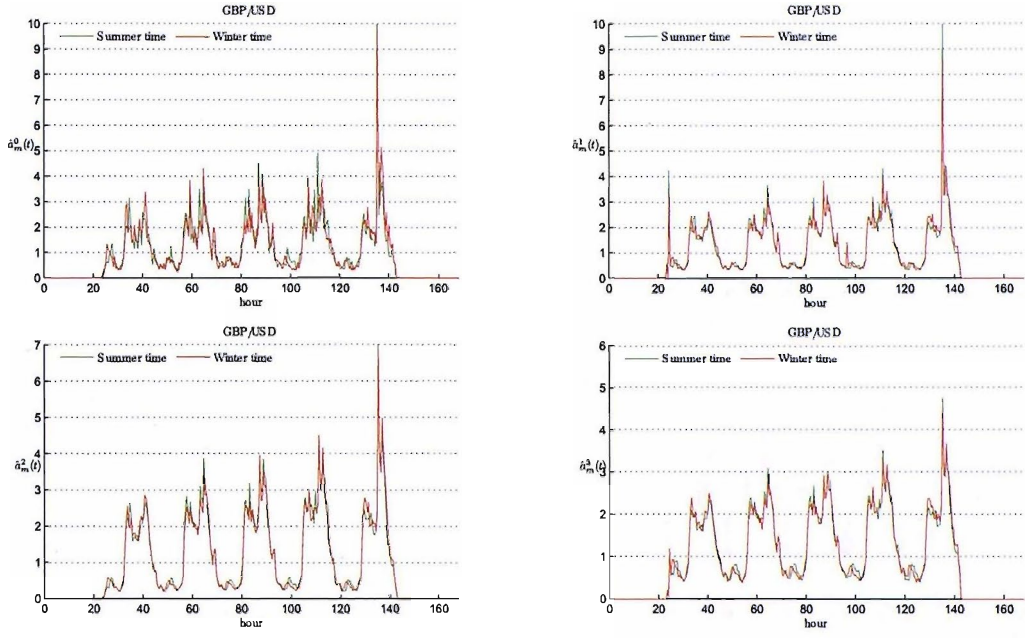

Figure F.5: Activity estimation of the GBP/USD rate index taking the averaged activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

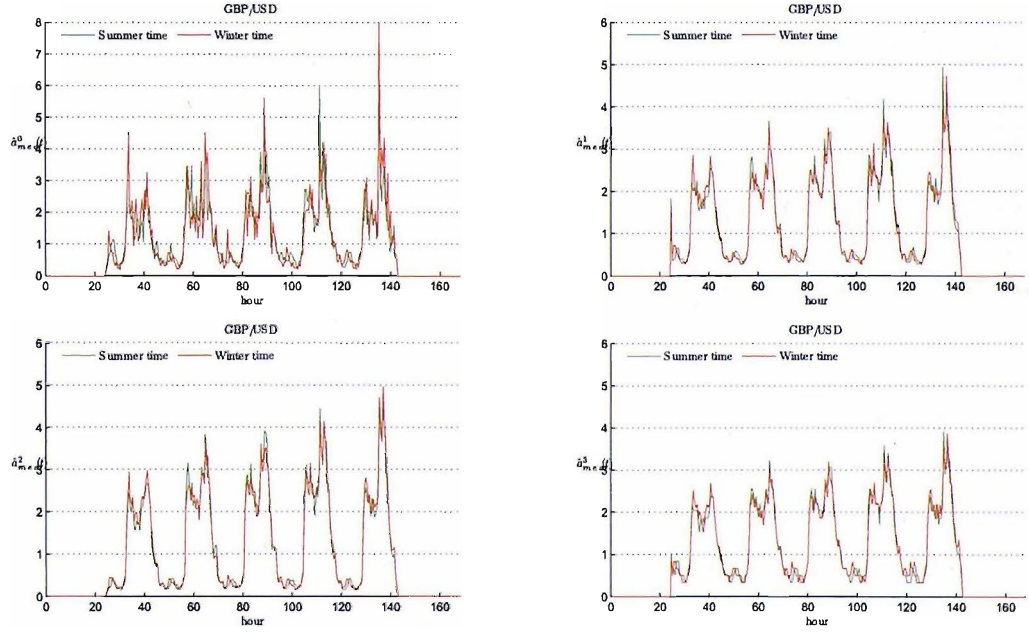

Figure F.6: Activity estimation of the GBP/USD rate index taking the median activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

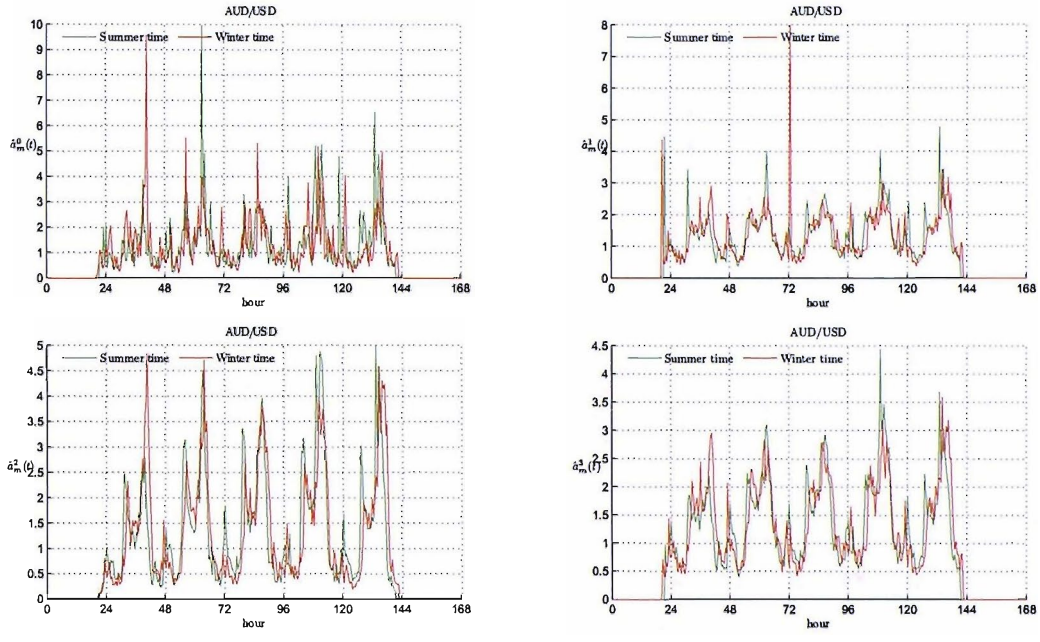

Figure F.7: Activity estimation of the AUD/USD rate index taking the averaged activity estimated by: the scaling law method(top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

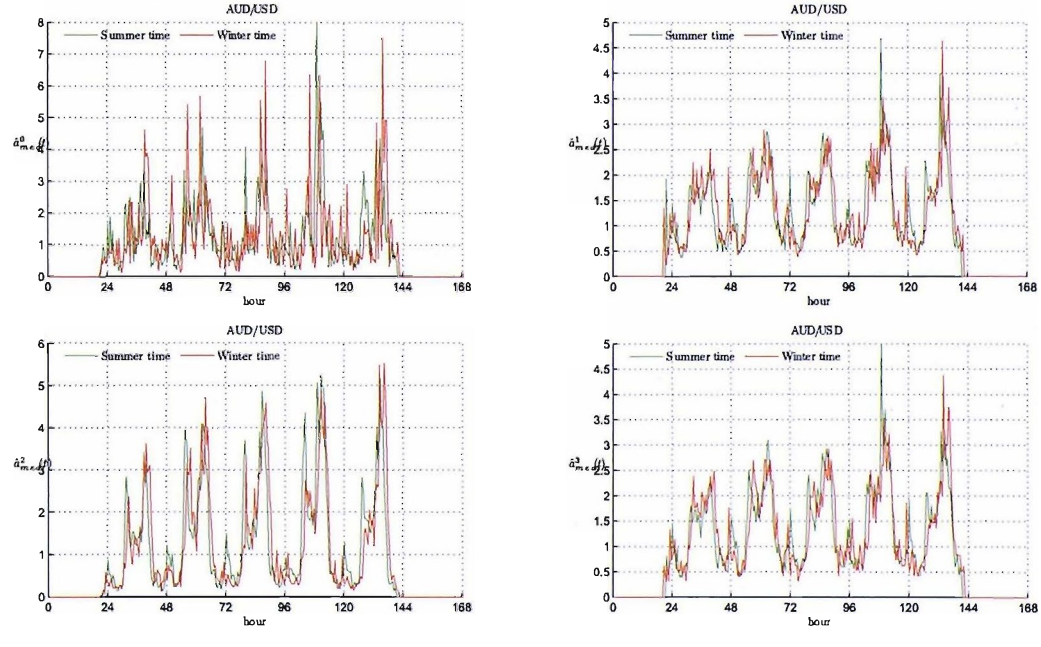

Figure F.8: Activity estimation of the AUD/USD rate index taking the median activity estimated by: the scaling law method(top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

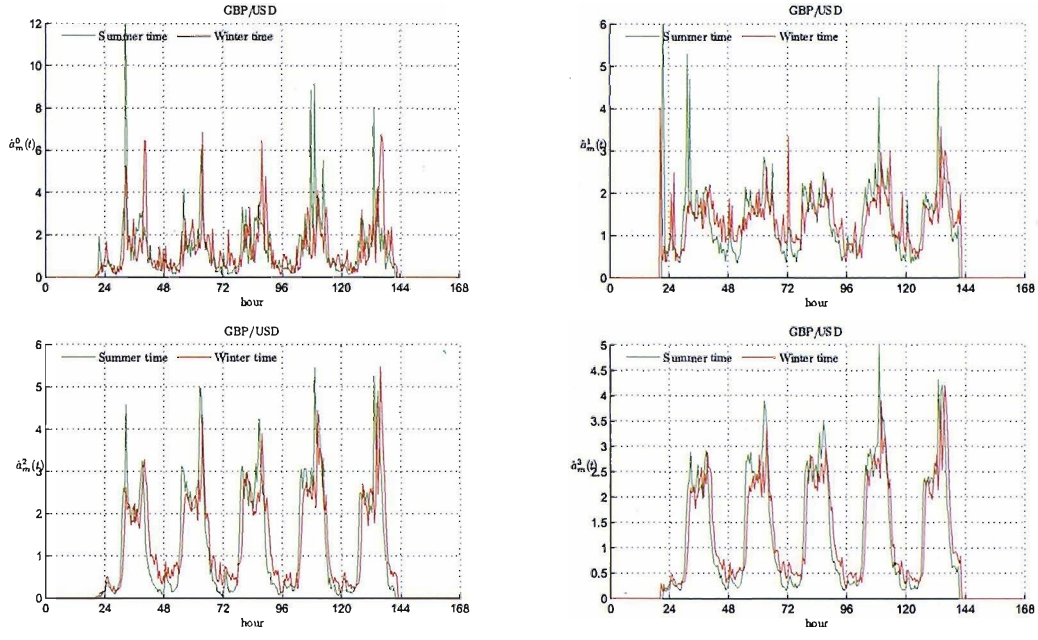

Figure F.9: Activity estimation of the GBP/USD rate index taking the averaged activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

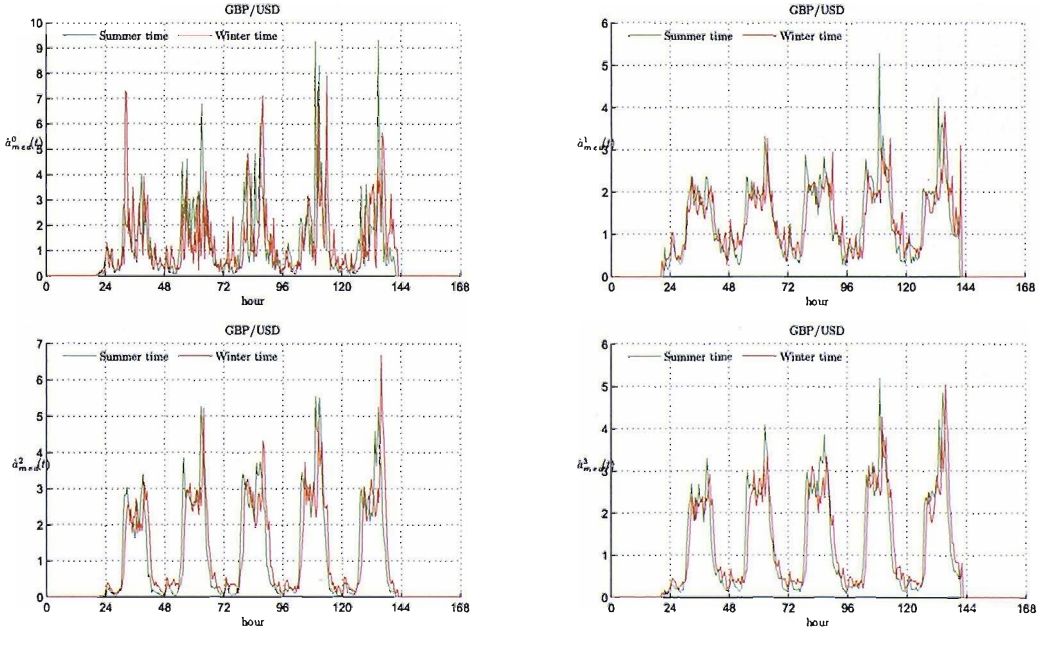

Figure F.10: Activity estimation of the GBP/USD rate index taking the median activity estimated by: the scaling law method(top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

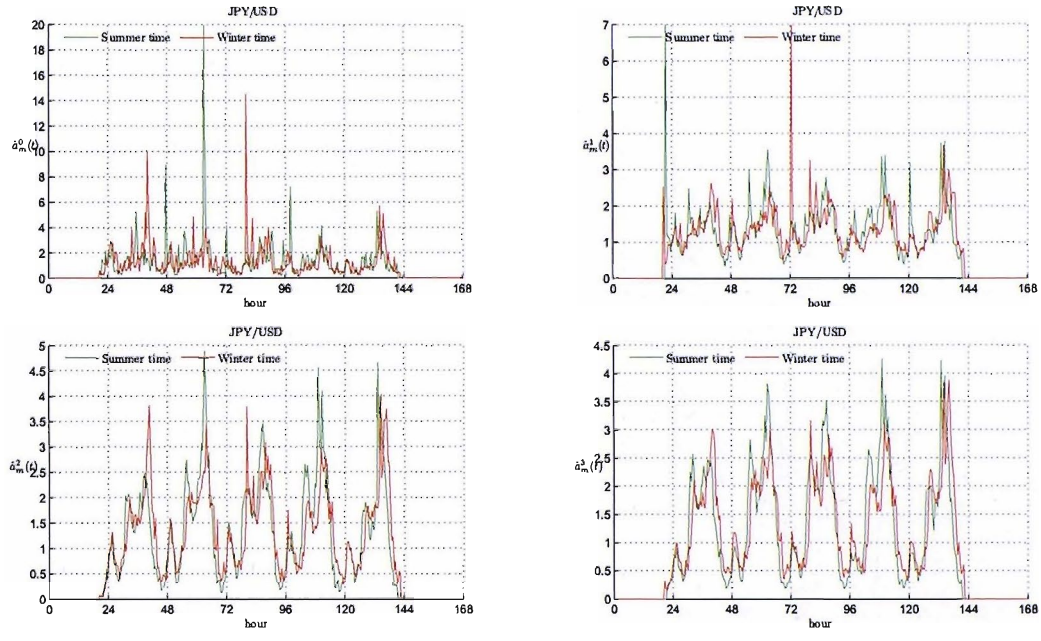

Figure F.11: Activity estimation of the JPY/USD rate index taking the averaged activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

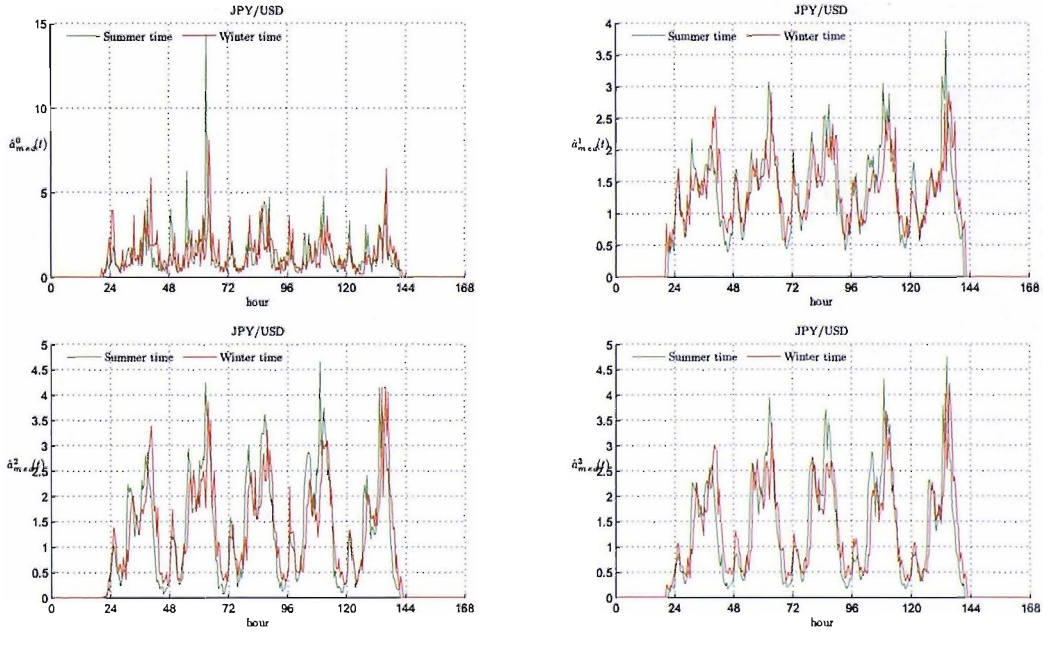

Figure F.12: Activity estimation of the JPY/USD rate index taking the median activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

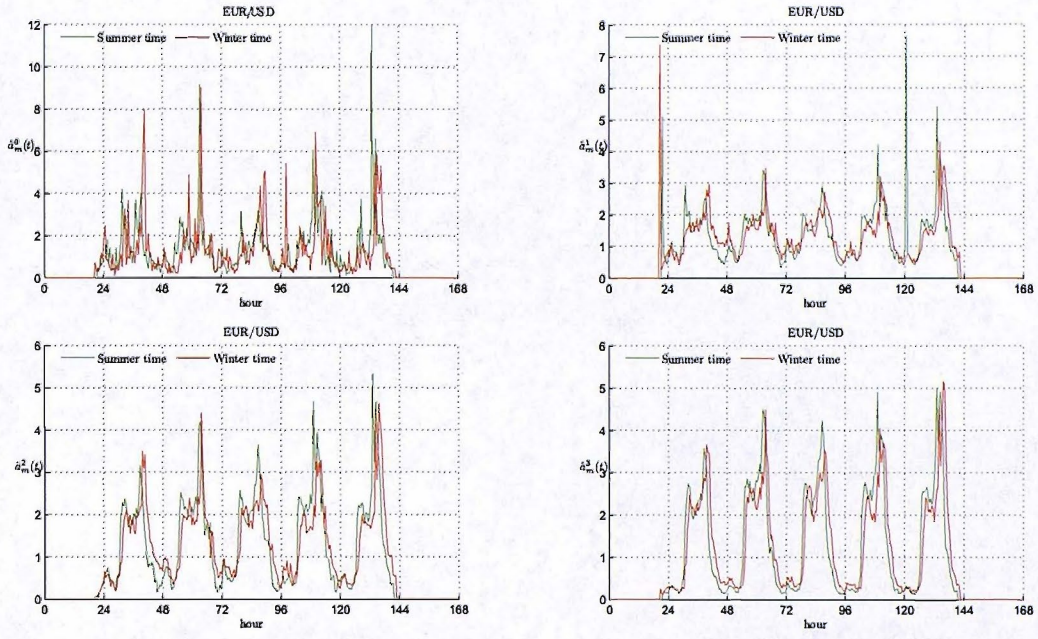

Figure F.13: Activity estimation of the EUR/USD rate index taking the averaged activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.

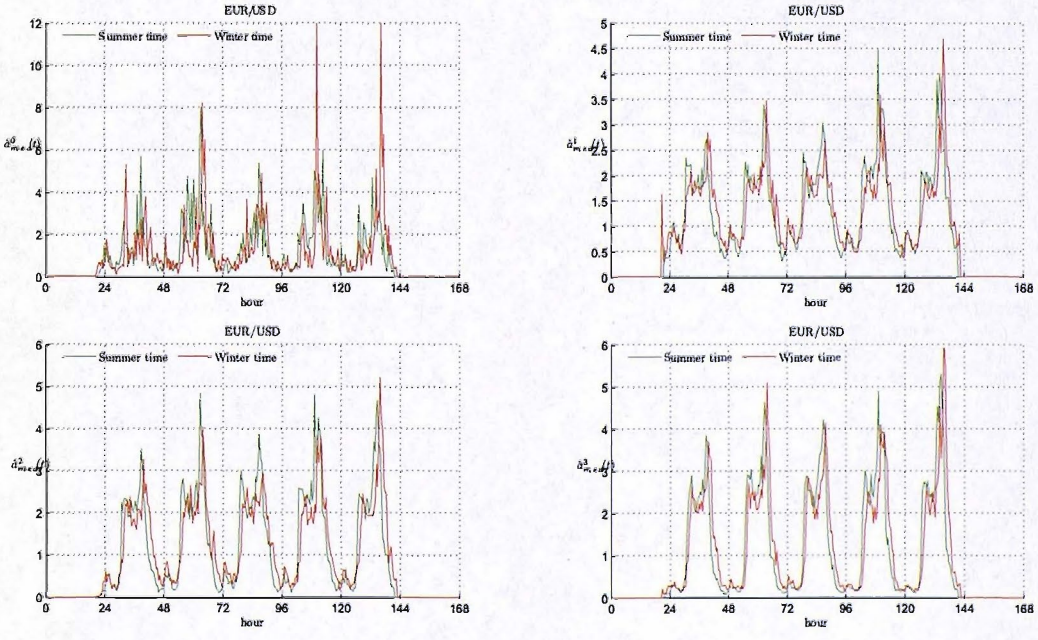

Figure F.14: Activity estimation of the EUR/USD rate index taking the median activity estimated by: the scaling law method (top-left), the $1/H$-variation (top-right), the log-variation (bottom-left) and crossing number (bottom-right) estimators.