

# Volatility is Rough<sub>1</sub>

University of Oxford

**V Courceau**

A thesis submitted in partial fulfillment of the MSc in  
*Mathematical and Computational Finance*

June 21, 2017

## Acknowledgements

This Masters dissertation was an intense 8-week project during which I benefited from the expertise and advice of my supervisor, Prof. Jan Oblój. I thank him for his support, feedback and every aspect of his involvement to make this report as thorough as possible. It has been very insightful to work alongside Jan and this will help me throughout my entire academic journey and beyond. It seems also natural to *thank* Jim Gatheral, although I have never had the opportunity to meet him, since his work has played an important role in my study of Mathematical Finance. Indeed, his ideas are at the inception of both my Bachelor's and Master's dissertations.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Definitions</b>	<b>3</b>
2.1	Overview of Gaussian Processes . . . . .	3
2.2	Definition of Fractional Brownian Motion . . . . .	4
2.2.1	Classical definition for non-negative times: . . . . .	4
2.2.2	Extension to negative times: . . . . .	5
2.3	A representation theorem for fBM . . . . .	5
2.4	Rough Fractional Stochastic Volatility model . . . . .	7
2.4.1	Approximating the fOU process . . . . .	10
2.5	Estimating the smoothness of the log-volatility process . . . . .	11
2.5.1	$\xi$ -estimation procedure: definition and intuition . . . . .	11
2.5.2	Besov spaces . . . . .	12
2.6	Predictions of fBM . . . . .	15
<b>3</b>	<b>Numerics: Real data and simulation</b>	<b>17</b>
3.1	Datasets . . . . .	17
3.2	The long-memory property . . . . .	19
3.3	Estimating H using $m(q, \Delta)$ . . . . .	19
3.3.1	Intuition . . . . .	19
3.3.2	Numerical results . . . . .	20
3.3.2.1	SPY . . . . .	20
3.3.2.2	OMI Datasets . . . . .	20
3.3.2.3	Other asset classes . . . . .	21
3.3.3	Numerical evidence of the RFSV Model . . . . .	23
3.4	Simulation study . . . . .	26
3.4.1	Procedure . . . . .	26
3.4.2	Testing the simulation paths via autocovariance function . . . . .	27
3.4.3	Testing the $\xi$ -estimation procedure . . . . .	29
3.5	Long-memory tests . . . . .	31
3.5.1	Failure of the Power-law method . . . . .	31
3.5.2	Variance of Realized Variance method . . . . .	32
3.5.3	Fractional differentiation method . . . . .	34

<b>4</b>	<b>Forecasting Volatility</b>	<b>37</b>
<b>5</b>	<b>Conclusion</b>	<b>40</b>
<b>A</b>	<b>Mandelbrot and Van Ness representation of fBM</b>	<b>42</b>
A.1	General results using the representation . . . . .	42
A.2	Equivalence between fBM and its representation . . . . .	46
<b>B</b>	<b>Technical proofs</b>	<b>47</b>
B.1	Proof of Definition 2.2.1 . . . . .	47
B.1.1	Ossiander and Waymire’s proof . . . . .	47
B.1.2	Gangolli’s proof: . . . . .	49
B.2	Proof of Equation (2.2) . . . . .	51
B.3	Proof of Proposition 2.4.1 . . . . .	51
B.4	Proof of Lemma 2.4.2 . . . . .	52
B.5	Proof of Proposition 2.4.4 . . . . .	55
B.6	Proof of the application of Proposition 2.5.6 . . . . .	56
<b>C</b>	<b>UZ model for high-frequency data</b>	<b>59</b>
<b>D</b>	<b>Real market dataset</b>	<b>61</b>
D.1	$\xi$ -estimation procedure . . . . .	61
D.2	Forecasting volatility: tables . . . . .	65
	<b>Bibliography</b>	<b>66</b>

# Chapter 1

## Introduction

In an already seminal paper, Gatheral, Jaisson and Rosenbaum [29] provided two important tools to the study of volatility by proposing a so-called *fractional* model, the Rough Fractional Stochastic Volatility (RFSV) model, as well as an estimation parameter of the *Hurst* exponent. It is defined as follows for  $S$  the price process,  $\sigma$  the volatility and  $X$  the log-volatility:

$$\begin{cases} dS_t &= S_t \sigma_t dW_t \\ \sigma_t &= \exp\{X_t\} \\ dX_t &= \nu dW_t^H - \alpha(X_t - m)dt \end{cases} \quad (1.1)$$

where  $\alpha > 0$ ,  $\nu > 0$ ,  $m \in \mathbb{R}$ . The intuition behind their ideas is the application of fractional Brownian Motion (fBM) in the modelling of log-volatility to support empirical evidence of roughness and fractal-like behaviour. Indeed, from its non-zero correlation between increments and self-similar scaling properties, *fBM* seems more suitable for this purpose as opposed to the standard Brownian motion. Also, they showed that the Hurst exponent was of order 0.1 using an Equity-based dataset. This has implications at two different levels: this demonstrates the roughness of volatility which was not incorporated in the FSV model by Comte and Renault [18] which assumed  $H > 1/2$ . Also, it entails that the volatility does not have the long-memory property although being a widely accepted fact (see the seminal works of Andersen et al. [1], [2], [3] based on this fact).

This report is based on this particular paper from Gatheral et al. Gatheral et al. [29] and complements it with additional numerical experiments and more explicit proofs of some of their theoretical results. More precisely, in Chapter 1 we will be interested in setting the theoretical foundations of the study of roughness of volatility: first, we introduce fractional Brownian motion as a Gaussian process and see the equivalence with the classical representation from Mandelbrot and Van Ness [44]. We will then define the RFSV model and its *fractional Ornstein-Uhlenbeck* modeling of log-volatility with  $H < 1/2$ . This will be accompanied by two distributional results (and their proofs) to approximate log-volatility by an affine function of fBM. We have not found in the literature such result for  $H < 1/2$  and we provide an argument using Jost's work [39] on Hurst parameter change. In addition, the RFSV essentially relies on the Hurst exponent and we need to estimate it. Gatheral, Jaisson and Rosenbaum propose a simple estimation procedure based on the  $q$ -th log-volatility increments moment. We will explain how this procedure is designed and justify its properties using Besov spaces by proving a convergence in probability result that was simply assumed in the original paper (from Gut's theorem, [32]). We will conclude this part by presenting

the tools to log-variance forecasting using the RFSV model.

To support this model, after presenting both an established and a newly-formed datasets, we will discuss empirically three points in Chapter 3. The new dataset is about other asset classes to which we would like to extend RFSV. We will first assess the estimation procedure for  $H$  by comparing with the results from the original paper and include some numerical evidence coming from market data which indicates that why the RFSV model should be preferred. Indeed, we observe the self-similarity property and Gaussian distributions at across all reasonable timescales (from 1 up to 120 days). Then, using simulation, we will compare results obtained using the Hurst exponent estimation procedure through the lenses of two volatility computation methods: the Uncertainty Zone method developed by Robert and Rosenbaum [59] and, depending on the data, either the standard Realized Variance estimator or the Minimal Realized Variance estimator as defined by Andersen et al. in [1]. We will retrieve the result on estimator smoothing presented in Appendix C, [29].

We will then be interested in the long-memory property from two points of view: first, this stylized fact for log-volatility (and sometimes for volatility) appears to be wrong. We will observe how our market and simulated data do not verify a power-law decay for the autocovariance function which may contradict long-memory property (but is not sufficient). This study will then be augmented by investigating two standard tests (as in [29]) under model misspecification: the Variance of Variance [1] and fractional differentiation [3] methods. Almost all of our market and simulated datasets pass those long-memory tests whilst having short-memory property given by Hurst exponents below  $1/2$ , leading to a contradiction.

Finally, one way to show if a model captures properly the underlying structure is to assess its predictive power. Here, we will be interested in forecasting log-variance and comparing our model to benchmark auto-regressive models as well as Corsi's HAR model [20]. We see that we obtain similar (yet not as strong) results as in [29]. More precisely, we found that over a set of 15 assets that includes the ones studied in the original paper and using both the realized variance and kernel and over forecasting horizons of 1, 5 and 20 days, we cannot distinguish any *clear* predictive domination of RFSV model over HAR (the AR models being almost every time outperformed).

# Chapter 2

## Definitions

### 2.1 Overview of Gaussian Processes

This part is based on the lecture notes of the course *Continuous martingales and stochastic calculus*, University of Oxford, written by Prof. Jan Oblój. For completeness, we shall comment that those definitions are identical (or at least, closely consistent) with those of Gallager, [27], Chapter 3.

We start with the definition of Gaussian random variable on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definition 2.1.1.** A random variable  $X : \Omega \rightarrow \mathbb{R}$  is called a centered standard Gaussian (or normal) if its distribution admits a density

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Additionally, a random vector  $X$  taking values in  $\mathbb{R}^d$  is said to be Gaussian if, and only if, any linear combination of its components is Gaussian (in  $\mathbb{R}$ ), that is, if and only if:

$$\forall u \in \mathbb{R}^d \quad \langle u, X \rangle := \sum_{i=1}^d u_i X_i \quad \text{is Gaussian}$$

The natural extension of this definition is to define Gaussian processes as follows:

**Definition 2.1.2.** A stochastic process  $(X_t)_{t \in \mathbb{R}_+}$  on  $(\Omega, \mathcal{F}, \mathbb{P})$  is called a centered Gaussian process if any finite linear combination of its coordinates is a Gaussian variable.

Equivalently,  $X$  is a (centered) Gaussian process if:

$$\forall n \in \mathbb{N}, \forall 0 \leq t_1 < t_2 < \dots < t_n, (X_{t_1}, \dots, X_{t_n}) \quad \text{is a (centered) Gaussian vector}$$

Given the specific properties of Gaussian processes and Daniell-Kolmogorov's theorem, it is possible to prove that for a Gaussian process is characterised by its covariance function  $\Gamma : [0; +\infty[ \rightarrow \mathbb{R}$  defined as follows:

$$\Gamma(t, s) := \text{Cov}(X_t, X_s)$$

We can also prove the converse - that is, given a suitable function  $\Gamma$ , there exists an appropriate Gaussian process with  $\Gamma$  as its covariance function. More precisely, we reformulate this result in the following theorem:

**Theorem 2.1.3** (Characterisation of Gaussian processes via the covariance structure).

Let  $(X_t)$  be a centered Gaussian process with  $\Gamma(t, s) := \text{Cov}(X_t, X_s)$  its covariance function. Then  $\Gamma$  is symmetric and positive semi-definite in the sense that for all  $n \in \mathbb{N}$ ,  $0 \leq t_1 < \dots < t_n$ :

$$\forall u \in \mathbb{R}^n \quad \sum_{1 \leq i, j \leq n} u_i u_j \Gamma(t_i, t_j) \geq 0$$

Conversely, for any such function  $\Gamma$ , there exists a centered Gaussian process  $(X_t)$  such that  $\Gamma$  is its covariance function.

*Remark.* Given this definition in terms of covariance, it implied that the covariance matrix  $(\Gamma(t_i, t_j)_{1 \leq i, j \leq n})$  is symmetric and positive semi-definite.

We can now study a few different definitions or points of view on Fractional Brownian Motion using these results.

## 2.2 Definition of Fractional Brownian Motion

### 2.2.1 Classical definition for non-negative times:

We will use the convention that, unless stated otherwise,  $0 \leq t \leq T < \infty$ . Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space which we complement with a  $\mathcal{F}$ -filtration  $(\mathcal{F}_t)_{t \in [0; T]}$  generated by an H-fBM. Denote  $(t, \omega) \mapsto W(t, \omega)$  the Brownian Motion from Bachelier, Wiener and Lévy that we will eventually adapt to be extended to negative times as well. The most common definition of Fractional Brownian Motion (further referred to as fBM) is given by saying that it is a centered Gaussian process with a given covariance structure. More explicitly, we can find in Hu et al. [34] the definition stated below:

**Definition 2.2.1.** The fractional Brownian Motion (fBM) with Hurst parameter  $H \in ]0; 1[$ , denoted  $W^H$ , is a zero mean (centered) Gaussian process with covariance:

$$\forall s, t \geq 0 \quad \Gamma_H(t, s) := \mathbb{E}[W_t^H W_s^H] = \frac{1}{2} [t^{2H} + s^{2H} - |t - s|^{2H}] \quad (2.1)$$

and such that  $W_0^H = 0$ .

*Remark.* This definition of fBM is said to have its variance normalised at time  $t = 1$ .

*Proof.* Theorem 2.1.3 ensures that such an object is well-defined. Indeed, this result yields that a Gaussian process is characterised by its covariance structure. Here, the given covariance function is symmetric and it remains to prove the positive semi-definiteness. For this purpose, two choices are offered to us: we can either use the very general theory developed by Gangolli [28] in 1967 or rely on more recent development by Ossiander and Waymire [53] where multidimensional time parameters in  $\mathbb{R}^d$  are studied and consider the special case  $d = 1$ . We will review both in Appendix B, Section B.1 as a short introduction to the extensive generalization efforts about fBM.  $\square$

### 2.2.2 Extension to negative times:

We take this opportunity to define the *two-sided* Brownian Motion used in the next section in the representation result of fBM from Mandelbrot and Van Ness [44] stated in Definition 2.3.1. More precisely, in their book [33] (Remark 6.1, page 61), Gusak et al. specified that the particular case  $H = 1/2$  in Definition 2.2.1 (where (2.1) was extended to  $\mathbb{R}^2$  using  $\mathbb{E}[W_t^H W_s^H] = \frac{1}{2} [|t|^{2H} + |s|^{2H} - |t-s|^{2H}]$ ) defines the two-sided Brownian motion. Spodarev et al. [64] make a similar observation in the Example 3 of Gaussian Random Fields section. In [40], Kaarakka et al. use the same idea to define the two-sided fBM for any  $H \in ]0; 1[$ .

*Remark.* In honour of the works of Hurst [37] in the 1950s, the parameter  $H$  is called the *Hurst exponent* and depending if  $H > 1/2$ ,  $H = 1/2$  or  $H < 1/2$ , the fBM showcases very different properties. For  $H > 1/2$  (resp.  $H < 1/2$ ) we have the persistence (resp. anti-persistence) property of fBM - that is, positive (resp. negative) correlation between increments. More precisely, we have:

$$\mathbb{E}[(W^H(1) - W^H(0))(W^H(s+1) - W^H(s))] \sim H(2H-1)s^{2H-2} \quad (2.2)$$

See Appendix B, Section B.2 for the derivation of the formula. For  $H = 1/2$ , we have uncorrelated jointly Gaussian increments, hence independent. In addition, the *larger  $H$  the smoother the path* in the sense that for  $H > 1/2$  (resp.  $H < 1/2$ ), we have long-range (resp. short-range) dependence (as in Definition 3.2.1). The words of Gatheral et al. [29], page 3, are as follows: *there is a one to one correspondence between regularity and long memory through the Hurst parameter  $H$* . This is what motivated the RFSV model from empirical analysis of financial data. Before that, we shall take a look at the historical representation of fBM as follows.

## 2.3 A representation theorem for fBM

The term *Fractional Brownian Motion* first appeared in Mandelbrot and Van Ness' seminal paper [44] in which they propose a more constructive representation of fBM which was proved to be unique for  $H > 1/2$  in [31] as a decreasing, right-continuous concave function on  $\mathbb{R}^+$ . They define fBM as follows:

**Definition 2.3.1** (Reduced Fractional Brownian Motion).

Let  $0 \leq H < 1$  and let  $b_0 \in \mathbb{R}$  be an arbitrary real number. Define  $W^H(t, \omega)$  the reduced fractional Brownian motion with parameter  $H$  and starting value  $b_0$  at time  $t = 0$ .  $W^H(t, \omega)$  is defined for all  $\omega \in \Omega$  by:

$$\left\{ \begin{array}{l} W^H(0, \omega) = b_0 \\ W^H(t, \omega) - W^H(0, \omega) = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^0 [(t-s)^{H-1/2} - (-s)^{H-1/2}] dW(s, \omega) \right. \\ \left. + \int_0^t (t-s)^{H-1/2} dW(s, \omega) \right\} \quad \forall t > 0 \end{array} \right.$$

*Remark.* If  $H = 1/2$  and  $b_0 = 0$ , then we retrieve the two-sided Brownian Motion:

$$W^{1/2}(t, \omega) = W(t, \omega)$$

Without loss of generality, we will consider that  $b_0 = 0$  in the rest of the report. Also, as hinted above, this is the original definition but we usually normalise the variance at time 1 by multiplying by a constant given in the proof of Lemma A.1.8, (A.3).

For completeness, we placed in Appendix A details and results on the equivalence between Definition 2.2.1 and Definition 2.3.1. However, those developments are not central in this report hence not placed in the main body.

One key feature of fBM is *self-similarity*. From Nuzman et al. [51], we can define self-similarity as stated below:

**Definition 2.3.2** (H-self-similarity (H-s.s.)).

A stochastic process  $(Y(t))_{t \in \mathbb{R}}$  is said to be H-self-similar (i.e. H-s.s.) for  $H \geq 0$  if it verifies the scaling property:

$$(Y(at))_{t \in \mathbb{R}} \stackrel{d}{=} (a^H Y(t))_{t \in \mathbb{R}} \quad \forall a > 0 \quad (2.3)$$

*Remark.* Those previous equalities are in the sense of equality in distribution. We also use the classical definition of the Gamma function:

$$\Gamma(x) := \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad x > 0$$

Gatheral et al. [29] use a different definition for fBM and we propose to go over that give equivalence of definition between the Definition 2.3.1 and the following:

**Definition 2.3.3** (Fractional Brownian Motion via self-similarity).

Let  $0 < H < 1$ .  $(W_t^H)_{t \in \mathbb{R}}$  is the Fractional Brownian Motion with Hurst exponent H if and only if  $W^H$  is a centered H-self-similar Gaussian process verifying for any  $t \in \mathbb{R}, \Delta \geq 0, q > 0$ :

$$E[|W_{t+\Delta}^H - W_t^H|^q] = K_q \Delta^{qH} \quad (2.4)$$

where  $K_q$  is the q-th moment of the absolute value of a standard Gaussian variable.

A process verifying Definition 2.2.1 verifies Definition 2.3.3 from the equivalence with Definition for two essential features we proved: stationarity (Lemma A.1.3) and self-similarity (Remark A.1.2) of increments. Indeed, we obtain

$$W_{t+\Delta}^H - W_t^H \stackrel{d}{=} W_{\Delta}^H \stackrel{d}{=} \Delta^H W_1^H$$

where  $W_1^H \sim N(0, 1)$  if we have normalized at time  $t = 1$ . We conclude by taking the q-th moment.

The other implication requires to use the standard trick when it comes to covariance estimation. From the definition, in particular for  $q = 2$ , for  $s \leq t \in \mathbb{R}$ , set  $\Delta := t - s \geq 0$ , yielding:

$$\mathbb{E}[(W_t^H - W_s^H)^2] = K_2(t - s)^{2H}$$

Therefore:

$$\begin{aligned}\mathbb{E}[W_t^H W_s^H] &= \frac{1}{2} [\mathbb{E}[(W_t^H)^2] + \mathbb{E}[(W_s^H)^2] - \mathbb{E}[(W_t^H - W_s^H)^2]] \\ &= \frac{K_2}{2} [|t|^{2H} + |s|^{2H} - |t - s|^{2H}]\end{aligned}$$

Recalling that  $K_q$  is the  $q$ -th moment of the absolute value of a standard Gaussian, we know that  $K_2 = 1$  which concludes the proof by continuity of Gaussian and their characterisation through their covariance structure.

**Remark 2.3.4. Integration by parts for integral w.r.t to fBM:**

As explored in McGonegal [45], Fractional Brownian Motion is not a *semi-martingale* if  $H \neq 1/2$ : its quadratic variation is infinite for  $H < 1/2$  and zero for  $H > 1/2$ . The latter contradicts the Doob-Meyer decomposition since  $W^H$  has unbounded 1-variation. This very unboundedness of variation renders ineffective the traditional definition via Riemann sum of the stochastic integrals with respect to fBM. We shall redirect ourselves towards to the more general (path-wise) Riemann-Stieljes definition (see Section 6.1, [45]).

Additionally, the author gives that for  $f$  a deterministic function such that it has bounded  $p$ -variation sample paths for all  $p < 1/(1 - H)$ , we have the following *integration by parts* relationship:

$$\int_0^T f(t) dW^H(t) = f(T)W^H(T) - \int_0^T W^H(t) df(t) \quad (2.5)$$

## 2.4 Rough Fractional Stochastic Volatility model

We would like to specify a model for the log-volatility that respects some empirical observations made originally by Gatheral et al. [29] on increments of log-volatility. More precisely, we model the price dynamics using the following SDE for  $t \geq 0$ :

$$\begin{cases} dS_t &= S_t \sigma_t dW_t \\ \sigma_t &= \exp\{X_t\} \\ dX_t &= \nu dW_t^H - \alpha(X_t - m)dt \end{cases} \quad (2.6)$$

where  $m \in \mathbb{R}$ ,  $\nu > 0$ ,  $\alpha > 0$ ,  $(\sigma_t)$  is the volatility process,  $(X_t)$  is the log-volatility process, and  $(W_t)$  a standard Brownian Motion. Note that the risk-free rate is ignored for simplicity of notation but its implementation does not entail any technical difficulty. The RFSV model is essentially about demanding fractional Ornstein-Uhlenbeck dynamics for the log-volatility  $(X_t)$  to correspond to empirical observations. Indeed, the authors illustrate the stylized fact that increments of log-volatility have an approximate normal distribution. The naive model is then:

$$\log(\sigma_{t+\Delta}) - \log(\sigma_t) = \nu(W_{t+\Delta}^H - W_t^H)$$

where  $\nu > 0$  and  $W^H$  being a fractional Brownian Motion with Hurst exponent  $H$  - which is, essentially, a Gaussian. Unfortunately, this model does not respect stationarity which may be a problem in two situations: it entails a the lack of tractability and over long period of time, volatility should be stationary. In order to impose a stationarity model, Gatheral

et al. use a two-step model for volatility using a modified Ornstein-Uhlenbeck process to generate the log volatility, further referred as *fractional Ornstein-Uhlenbeck (fOU)*, using the following SDE:

$$dX_t = \nu dW_t^H - \alpha(X_t - m)dt \quad (2.7)$$

with  $m \in \mathbb{R}$ ,  $\nu > 0$ ,  $\alpha > 0$ . According to the development in [16] and [40], we will reformulate the problem without the mean-reverting feature and in terms of the Langevin equation in integral form as follows:

$$X_t = \phi - \alpha \int_0^t X_s ds + \nu W_t^H, \quad t \geq 0 \quad (2.8)$$

where  $\phi$  is a random variable that represents Mandelbrot-Van Ness' definition in the sense that fBM originates from the infinite past.

To specify an initial condition, the authors observe that for all  $H \in ]0; 1[$ ,  $a \in [-\infty; +\infty[$ , the integral

$$\int_a^t e^{\lambda u} dW_u^H, \quad t > a$$

exists as a path-wise Riemann-Stieltjes integral which makes the next observation well-defined, in our case define  $\lambda := \alpha > 0$ . As  $\alpha$  and  $\nu$  are constant, we fulfil the local Lipschitz and linear growth conditions which yield that there exists an a.s. unique strong solution to (2.8) that happens to be a.s. continuous. Finally, to transfer from (2.8) to (2.7), observe that  $dX_t = d(X_t - m)$ . Indeed, for  $X$  verifying (2.7), we have by *integration by parts* (Remark 2.3.4, (2.5)):

$$\begin{aligned} d(X_t e^{\alpha t}) &= dX_t e^{\alpha t} + \alpha X_t e^{\alpha t} dt \\ &= e^{\alpha t} [\nu dW_t^H + \alpha m dt] \end{aligned}$$

So, it follows that

$$\begin{aligned} X_t e^{\alpha t} &= \nu \int_0^t e^{\alpha s} dW_s^H + m \int_0^t \alpha e^{\alpha s} ds + X_0 \\ X_t &= e^{-\alpha t} \left[ \nu \int_0^t e^{\alpha s} dW_s^H + m(e^{\alpha t} - 1) + X_0 \right] \end{aligned}$$

We can then reformulate this using the notation of (2.8):

$$X_t = e^{-\alpha t} \left[ \nu \int_0^t e^{\alpha s} dW_s^H + \phi + m \right] \quad (2.9)$$

where  $\phi$  is now part of the initial condition  $X_0 = \phi + m$ . By property of the Riemann-Stieltjes integral w.r.t. Gaussian process, it is a.s. continuous as afore-mentioned. We would like to incorporate the whole past history to ensure stationarity in our solution (so that it is not dependent on the value at time 0) without affecting the distribution. As hinted by Mandelbrot and Van Ness' definition (Definition 2.3.1), fBM is defined by the whole past history - we then consider the two-sided fBM (as in Section 2.2.2). Define:

$$\phi := \nu \int_{-\infty}^0 e^{\alpha u} d\hat{W}_u^H$$

As explicitly treated in Kaarakka et al. [40],  $\phi$  is well-defined. Indeed, recall using *integration by parts* (as in (2.5)), we have in integral form for any  $s \in \mathbb{R}_+$

$$\int_0^s e^{\alpha s} dW_u^H = e^{\alpha s} W_s^H - \int_0^s \alpha e^{\alpha u} W_u^H du$$

This can be extended to  $s \in \mathbb{R}_-$  using  $\hat{W}^H$  and its definition on the negative semi-infinite interval. Explicitly, we obtain:

$$\int_s^0 e^{\alpha s} d\hat{W}_u^H = -e^{\alpha s} \hat{W}_s^H - \int_s^0 \alpha e^{\alpha u} \hat{W}_u^H du$$

Since  $\alpha > 0$ , the first term will see its variance shrink to zero as  $s$  gets larger. For the second term, see [40], Section 2.2, for the final details of the limit  $s \rightarrow -\infty$ . Therefore:

$$X_t := \nu \int_{-\infty}^t e^{-\alpha(t-u)} d\hat{W}_u^H + m$$

still verifies the equation but with the initial condition set to be  $X_0 = \phi + m$ . Note that the last term  $m$  is due to the mean-reverting component of the SDE (2.7). From the stationarity of the increments of fBM (Lemma A.1.3), such reformulation is *stationary*. By the usual  $L^2$  argument associated with Kolmogorov-Čentsov's continuity criterion (Theorem A.1.6) and closed-ness of Gaussians under countable sums,  $X_t$  is a.s. Gaussian. Also, we let ourselves reformulate this initial value  $\phi$  in terms of the initial value of  $X$ , namely  $X_0$ , so that the starting value of the volatility  $\sigma_t$  defined as follows is as desired:

$$\sigma_t = \exp\{X_t\}, \quad \text{where } , t \geq 0$$

In the rest of the report we will use the *stationary* solution of (2.8) (note that we integrated from  $-\infty$  to 0) expressed formally by:

$$X_t^\alpha := \nu \int_{-\infty}^t e^{-\alpha(t-s)} dW_s^H + m \tag{2.10}$$

### 2.4.1 Approximating the fOU process

One of the most important results in [29] is the fact that the log volatility defined in (2.7) behaves like fBM for  $\alpha \ll 1/T$  and that such  $\alpha$  is indeed expected to be small as it will be used as a measure of the granularity (i.e. tick size). They formalise this idea in Proposition 3.1 in [29] as stated below:

**Proposition 2.4.1.** *Let  $W^H$  be a fBM and  $X^\alpha$  be an fOU process (i.e. a solution of (2.7) or (2.8) for a given  $\alpha > 0$ ). Then:*

$$\mathbb{E} [\sup_{t \in [0; T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H|] \xrightarrow{\alpha \rightarrow 0} 0 \quad (2.11)$$

*Proof.* Presented in Appendix B, Section B.3 using Lemma 2.4.2.  $\square$

To prove such proposition, we need to find a bound for the supremum process of fBM. Gatheral et al. cited the works of Novikov and Valkeila [49] in which they carried out the calculations in a more general context of stopping times. Unfortunately, it seems that the result of supremum process bound for  $H < 1/2$  Gatheral et al. used is not in the cited paper because, as Novikov and Valkeila said: *we do not have any upper bound for the case  $H < 1/2$ , but we conjecture than we have a similar result.* Actually, even some of the recent developments elude the case  $H < 1/2$  for the supremum (e.g. [22], [65] both published in 2009).

More precisely, Novikov and Valkeila proved for  $H > 1/2$  the following result in Theorem 1.2 and Proposition 2.1 of [49]. We propose a reformulation as follows:

**Lemma 2.4.2** (Supremum process upper bound).

*Let  $H \in ]0; 1[$  and  $p \geq 1$ . Then, there exists a function  $C(p, H)$  such that*

$$\mathbb{E} \left[ \left( \sup_{t \in [0; T]} |W_t^H| \right)^p \right] \leq C(p, H) T^{pH}$$

*Proof.* See Appendix B, Section B.4.  $\square$

*Remark 2.4.3.* We proved that, under the assumption that  $\alpha$  is small enough, our fOU process behaves almost like a affine function of fractional Brownian Motion which is what we looked for. Indeed, our naive non-stationary model is retrieved when setting formally  $\alpha = 0$ .

We wish to replicate the properties of fBM as well as incorporating new ones. As an example, the mean reversion of the fOU process which is a stylized fact for traded assets leaves us with a more realistic model (as studied in [26]). In addition, Gatheral, Jaisson and Rosenbaum [29] presented another interesting result: the fOU process produces a similar scaling property as the fBM when  $\alpha$  goes to zero (Corollary 3.1 of the original paper). They state it as follows:

**Proposition 2.4.4** (Approximate scaling property as  $\alpha \rightarrow 0$ ).

*Let  $q > 0$ ,  $t > 0$ ,  $\Delta > 0$ . We have that:*

$$\mathbb{E} [|X_{t+\Delta}^\alpha - X_t^\alpha|^q] \xrightarrow{\alpha \rightarrow 0} \nu^q K_q \Delta^{qH} \quad (2.12)$$

where  $K_q$  is the  $q$ -th absolute moment of a standard normal.

To prove such result, Gatheral et al. needed to express the covariance between  $X_{t+\Delta}^\alpha$  and  $X_t^\alpha$  and use the fact that Gaussians are characterised by their mean and variance and used the associated properties.

*Proof.* See Appendix B, Section B.5. □

*Remark 2.4.5.* As explained in Cabrera and Volodin [15] (p. 646), another proof could have been to use the classical result of La Vallée - Poussin that a sequence  $\{X_n\}$  of random variables is uniformly integrable if, and only if, there exists a function  $\phi$  that such that  $\mathbb{E}[\phi(X_n)] < \infty$  where  $\phi : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$  which verifies  $\phi(t)/t \xrightarrow{t \rightarrow +\infty} 0$ . We would have chosen  $\phi(x) = |x|^{(q+1)/q}$ .

## 2.5 Estimating the smoothness of the log-volatility process

### 2.5.1 $\xi$ -estimation procedure: definition and intuition

We are interested in studying market data to estimate the Hurst parameter  $H$  that we believe to be essential in understanding the roughness of volatility. This will be done by reproducing the results given by [29] - that is, observing that Hurst exponents lie between 0 and 1/2 and of order 0.1 across all assets. In order to do so, we shall look into a two-step process which is based on a similar estimator to the absolute log-volatility  $q$  -  $th$  moment. We will further refer to this estimator as  $m(q, \Delta)$  defined as follows:

**Definition 2.5.1.** Suppose we are in the RFSV framework given by (2.6) and that we have data on a uniform time grid of mesh  $\Delta > 0$ , namely  $\{t_n\}_{n=0}^N := \{n\Delta\}_{n=0}^N$ , where  $N$  verifies  $N := \lfloor T/\Delta \rfloor$ . Define for  $q \geq 0$ :

$$m(q, \Delta) := \frac{1}{N} \sum_{k=1}^N |\log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta})|^q$$

*Remark 2.5.2.* Such estimator implicitly expects to observe the volatility at regular intervals but it is reasonable to assume so as many financial signals are (re-)constructed this way.

The properties of such object are not trivial and although under stationarity assumptions it looks like the empirical counterpart of  $\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q]$ .

Let us ignore the technical details for now. Gatheral, Jaisson and Rosenbaum [29] take another approach using fBM and its Hurst parameter  $H$ . Indeed, they use the self-similarity of fBM and the approximate scaling property of their RFSV model (Proposition 2.4.4) to infer a value of  $H$  without making any additional assumption. In practice, using (2.12), they use a range of values for  $q$  and  $\Delta$  to estimate different values for  $m(q, \Delta)$  as a function of  $\Delta$ . Then, if the RFSV model is true, taking the logarithm yields that a simple linear regression of  $\log(m(q, \Delta))$  against  $\log(\Delta)$  for a given  $q$  should give

$$\log(m(q, \Delta)) \approx \xi_q \log(\Delta) + C \quad \text{for } \alpha \ll 1/T$$

where  $C$  is a constant and more importantly, if the approximation (2.12) holds, we should have

$$\xi_q \approx qH$$

Finally, to estimate  $H$  as one value, we will take the mean value over all the values of  $q$  considered. Indeed, there is no reason why a particular value of  $q$  should be preferred but smaller or large values could polarize numerical bias in the data. With the hope to smooth out this effect, we could use the mean value. We will further refer to this numerical recipe as the  $\xi$ -*estimation procedure*. To motivate this argument about the independence of  $\xi$  w.r.t  $q$ , we plot numerical values of  $\xi_q$  against  $q$ , as shown in Appendix ??, Figure 3.9. We see an almost straight line that goes through the origin. The slight convex curvature was deemed as a consequence to the finiteness of data that implies a small bias in both the estimations of the moments and the linear regression estimation procedure. It is discussed further in Section 3.3.3.

This procedure has several advantages:

- As afore-mentioned, it does not depend on a prior belief regarding the value of  $H$ ;
- Because we are working with first order differences on the log-scale, it is also *independent* from the volatility scaling (which is not the case when we will try to predict volatility);
- It is quite simple both to understand and to implement;

This process seems very appealing and has proved to be very interesting given its simplicity. However, it is not obvious why this argument should hold. Actually, our RFSV imposes a specific log-volatility process which verifies the stationarity assumption in the increments to approach  $\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q]$  numerically. However, our main assumption about  $m(q, \Delta)$  is related to *Besov spaces*. Informally, we are interested in using a result related to the *strength* of convergence of  $m(q, \Delta)$ . Indeed, if we could prove that there exists a finite number  $y_q$  or an a.s. finite random variable  $Y_q$  so that

$$N^{qH} m(q, \Delta) \xrightarrow{p} Y_q \tag{2.13}$$

then, we would know that  $m(q, \Delta)$  informs us on the roughness of a path that is at least as rough as the real one. This is what we are interested in since we are *ultimately* interested in knowing if  $H < 1/2$  or not. To quote Rosenbaum in his related paper [61], Besov spaces are *a natural framework to study smoothness of the sample paths of a continuous time random process [in  $L^p$ ]*. We will now give a brief overview of Besov spaces.

## 2.5.2 Besov spaces

Cohen's definition [17] of Besov spaces is similar to the following:

**Definition 2.5.3** (Besov spaces).

Let  $f$  be a real function defined on  $[0; 1]$ . Let  $\Delta_h^n$  be the operator defined recursively by

$\Delta_h^1 : f \mapsto f(\cdot + h) - f(\cdot)$  and  $\Delta_h^{n+1} : f \mapsto \Delta_h^1(\Delta_h^n(f))$ . The  $n$ -th order  $L^p$  modulus of smoothness of  $f$  on  $[0; 1]$  is

$$\omega_n(f, t)_p := \sup_{|h| \leq t} \|\Delta_h^n f\|_{L^p(\Omega_{h,n})}$$

where  $\Omega_{h,n} := \{x \in [0; 1] : x + kh \in [0; 1] \quad \forall k = 0, \dots, n\}$ .

For  $p, q \geq 1, s > 0$ , the Besov space  $\mathcal{B}_{p,q}^s([0; 1])$  consists of the functions  $f \in L^p[0; 1]$  such that:

$$\{2^{sj}\omega_n(f, 2^{-j})_p\}_{j \leq 0} \in l^q$$

with  $n \in \mathbb{N}$  such that  $s < n$ .

*Remark 2.5.4.*  $\mathcal{B}_{p,q}^s([0; 1])$  is a Banach space equipped with the norm

$$\|f\|_{\mathcal{B}_{p,q}^s([0; 1])} := \|f\|_{L^p} + \|\{2^{sj}\omega_n(f, 2^{-j})_p\}_{j \leq 0}\|_{l^q}$$

Actually, they can be extended for  $p$  or  $q$  less than 1 and  $s > \max\{1/p - 1, 0\}$ , however the resulting Besov space  $\mathcal{B}_{p,q}^s([0; 1])$  is only a quasi-Banach space.

Definition 2.5.3 makes sense in term of smoothness of paths as it measures the  $n$ -th order variation of  $f$  directly in  $L^p$  at dyadic points which is then brought up to  $l^q$ .

As formally studied by Rosenbaum in [61],  $p$ -variations come up naturally in Mathematical Finance. For instance, the particular case of  $p = 2$  gives the so-called *realized volatility* or a link between first order  $p$ -variations and market microstructure has also been studied by Rosenbaum [62]. This can also be complement with the following result which is not exactly what we look for but gives a good intuition of the ongoing dichotomy around the order  $s$  of the Besov spaces.

**Proposition 2.5.5.** *Let  $(X_t)_{t \in [0; 1]}$  be a zero mean Gaussian process with stationary increments. Let  $\sigma(h) := \mathbb{E}[(X_{t+h} - X_t)^2]^{1/2}$ . Additionally, assume that for some  $0 < r < 1$  and  $0 < \alpha < \infty$ , we have:*

$$\lim_{h \rightarrow 0} \sigma(h)/h^r = \alpha$$

*Then, we obtain in particular:*

- For  $0 < s \leq r, 1 \leq q < \infty$   $X \in \mathcal{B}_{q,\infty}^s$
- For  $s > r, 2 \leq q < \infty, rq > 1$   $X \notin \mathcal{B}_{q,\infty}^s$

Given that the fBM has a finite second moment at time  $t = 1$ , we obtain by H-self-similarity and stationarity of increments that  $\sigma(h) = h^H \mathbb{E}[X_1^2]^{1/2}$ . Therefore, the aforementioned Proposition 2.5.5 can be applied with  $r = H$ . The details on how to handle the cases  $0 < q < 1$  (or  $0 < q < 2$  for the second bulletpoint) are not explicitly treated so, when choosing  $q$  to perform the regression as in Section 2.5.1, we will assume that the result still hold in the quasi-Banach spaces we then obtain. We see that the parameter  $r$  contains information in terms of regularity of the fBM since all paths belong to  $\mathcal{B}_{q,\infty}^r$  almost surely for  $r \leq H$ . Given that  $r = H$ , it leads us to believe that this is the link between fBM and its regularity parameter  $H$  by making use of Besov spaces. Our hope is that we can do

exactly the same as we did with fBM but using a first order q-variation of our *fOU* process given by (2.7) which is the exact content of the next result as presented in Corollary 1, [61]. First, we define the first order q-variation on the time interval  $[0; 1]$  for  $N = 2^j$  data points equally spaced (and time-scaled by  $T$  if necessary) by:

$$V_j^q(X) := \sum_{k=1}^{2^j} |X_{k2^{-j}} - X_{(k-1)2^{-j}}|^q$$

We can now see the Proposition that will yields a result similar as (2.13) that we could apply on our estimator  $m(q, \Delta)$

**Proposition 2.5.6** (First order p-variation convergence and Besov spaces dichotomy).

*Let  $(X_t)_{t \in [0;1]}$  be a continuous process such that for  $q > 1$ ,  $1/q \leq s < 1$  and  $Y$  an almost surely finite and positive random variable. Suppose that*

$$2^{j(qs-1)} V_j^q(X) \xrightarrow{p} Y$$

*Then, for any  $\epsilon > 0$ , we have  $X \in \mathcal{B}_{q,\infty}^{s-\epsilon}([0; 1])$  and  $X \notin \mathcal{B}_{q,\infty}^{s+\epsilon}([0; 1])$ .*

As explained in the  $\xi$ -estimation procedure presentation, this Proposition can hopefully be applied to  $m(q, \Delta)$  as assumed in [29]. We interested in proving that we have a Weak Law of Large Numbers (WLLN) as stated in Proposition 2.5.6 with an arbitrary  $q$  and  $X$  being the log-volatility process specified in the RFSV model (2.6). Such process  $X$  has Gaussian correlated increments albeit identically distributed. This correlation is the non-classical component of the problem.

Actually, this research topic is still very active. Existing results involve a Central Limit Theorem for additive functional of fBM but these are limited to  $H \in \left[\frac{1}{d+2}; \frac{1}{d}\right]$  found in [50] (improving the result first obtained for  $H \in \left[\frac{1}{d+1}; \frac{1}{d}\right]$  in [35] of the same authors). Here,  $d$  is the dimension of the parameters space : in our case,  $d = 1$  and therefore, we know Proposition 2.5.6 can be applied for  $H \in [1/3; 1]$ . The lower bound  $1/(d+2)$  is found to be optimal since lower values of  $H$  would lead to an infinite constant in their proof. Still, these results are stronger that what we need since we only need convergence in probability and not in distribution, which will hopefully relax the lower bound assumption of  $1/3$  for  $H$ . We will therefore rely on WLLN results in the context of dependent variables, namely given by the theorem of Section 3 from Gut [32], page 50. Its application is presented in Appendix B, Section B.6.

Hence, we can apply Proposition 2.5.6 on the continuous equivalent to fBM we obtained by using Kolmogorov's continuity criterion for  $s = H$ . Therefore, it is very insightful in the understanding of the estimator  $m(q, \Delta)$ : the Hurst parameter characterises the maximum regularity parameter verified by fBM. This is the essence of the interest behind fractional Brownian Motion models as they inform on the roughness of volatility. Through  $m(q, \Delta)$ , we are *given to observe* a path that is at least as rough as the original path which will help up know if  $H < 1/2$  regarding the implied properties of the log-volatility (i.e. short-memory property). Because  $H$  is the extremal value we can reasonably hope to capture this particular roughness parameter using a linear regression for different value of  $q$  (See Section 2.5 for further details and Section 3.3.2 for the numerical results).

## 2.6 Predictions of fBM

We would like to assess how well our model capture the structure of real world data. We perform this by investigating the predictive power of the RFSV model. Hence, we shall need a method to predict the value of Fractional Brownian Motion given its past values, namely  $\hat{W}_{t_0+a}^H := \mathbb{E}[W_{t_0+a}^H | W^H(t), t < t_0]$ , where  $t_0$  is arbitrary and represents the last date we observed from the fBM and note that  $a > 0$ . As Gatheral et al. [29] did, we are going to rely on the study of Nuzman and Poor [51] and their linear processing of fBM using Lamperti's transform. We will briefly cover the main results we are going to use as well as the idea behind their breakthrough - that is, prediction of fBM in the context of  $H < 1/2$ . From Theorem 4.2 [51]:

**Theorem 2.6.1** (Prediction of fBM for  $H < 1/2$ ).

Let  $W^H$  be a fractional Brownian Motion with Hurst parameter  $H \in ]0; 1[$ . For any  $a > 0$  and  $t_0 \in \mathbb{R}$ , we have:

$$\hat{W}^H(t_0 + a) = W^H(t_0) + \frac{\cos(\pi H)}{\pi} a^{H+1/2} \int_0^{+\infty} \frac{W^H(t_0 - s) - W^H(t_0)}{s^{H+1/2}(s+a)} ds$$

This can be reformulated in the case  $H < 1/2$  as follows:

$$\hat{W}^H(t_0 + a) = \frac{\cos(\pi H)}{\pi} a^{H+1/2} \int_0^{+\infty} \frac{W^H(t_0 - s)}{s^{H+1/2}(s+a)} ds$$

*Remark 2.6.2.* There exists also a reformulation for  $H > 1/2$  but this case is not central in this report.

Nuzman and Poor [51] provide a comprehensive and technical argument which is beyond the scope of this work but they propose an original proof using Lamperti's transform [42]. This mathematical object is an invertible map between H-self similar processes and shift-stationary processes. They define it in a similar way as what follows (Definition 2.1, [51]):

**Definition 2.6.3.** Let  $I \subset ]0; +\infty[$  be a positive real index set. For each  $H \in \mathbb{R}$ , the Lamperti transformation with parameter H on  $\mathbb{R}$ , denoted  $L_H$ , is an invertible map between  $\mathbb{R}$ -valued functions on I and  $\mathbb{R}$ -valued functions on  $\log(I)$ . Indeed, for each  $y : I \rightarrow \mathbb{R}$  the function  $L_H y : \log(I) \rightarrow \mathbb{R}$  is given by

$$L_H y(\tau) = e^{-H\tau} y(e^\tau), \quad \tau \in \log(I)$$

The inverse transform is then given by

$$L_H^{-1} x(t) = t^H x(\ln t), \quad t \in I$$

The authors also explain that we are interested in predicting using the information contained a finite time interval containing the origin. Hopefully, the particular case of fBM has stationary increments and therefore we can *jump* to a future time  $t_0$ . From Nuzman and Poor's words ([51], p. 14), *the information contained in a self-similar process on a positive or negative neighborhood of the origin is equivalent to the information contained in*

a stationary process on a semi-infinite interval. They made a link between the study of the frequency domain (as found in signal processing methods) and the prediction of fBM via spectral factorisation. Indeed, from Definition 2.4 by Mandelbrot and Van Ness [44], we see that we need the semi-infinite past information to generate fBM. As a measure of the quality of prediction, they use the renowned Minimum Mean Square Error (further referred as MMSE and denoted  $D^2$ ) measure of fit defined as follows:

$$D^2 := \mathbb{E}[(Z - \hat{Z})^2]$$

where  $Z$  is the random variable we are trying to predict and  $\hat{Z}$  the (optimal) prediction itself.

*Remark 2.6.4.* We can define  $d^2$  as a *the normalized MMSE*. Using  $H$ -self similarity of fBM, we obtain for an arbitrary fixed  $a > 0$  that:

$$d^2(W^H(a)) := \frac{D^2(W^H(a))}{\mathbb{E}[W^H(a)]^2} = \frac{a^{2H}}{a^{2H}} \frac{D^2(W^H(1))}{\mathbb{E}[W^H(1)]^2} = d^2(W^H(1))$$

Therefore, the choice of  $a$  is not crucial and will be chosen to be 1 in the technical details.

Nuzman and Poor ([51], p. 21-24) proved that the normalized MMSE for fBM verifies for any  $0 < H < 1$ :

$$d^2(W^H(1)) = \frac{\Gamma(3/2 - H)}{\Gamma(2 - 2H)\Gamma(H + 1/2)} \quad (2.14)$$

See Figure ?? below. Observe that for the extreme cases,  $H \rightarrow 0$  yields that we can only explain half of the variance using predictions since  $d^2(W^H(1)) = 1/2$  and we also retrieve that, for  $H = 1/2$  we have  $d^2(W^H(1)) = 1$ . Indeed, this coincides with the particular case when the fBM is  $\beta$  impossible.

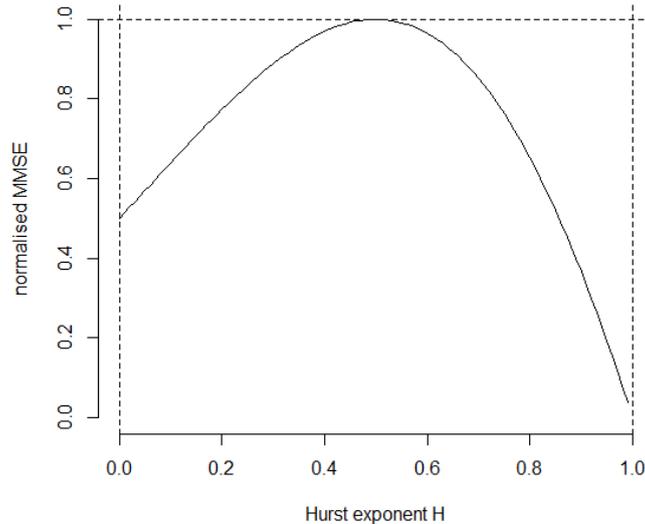


Figure 2.1: Normalised Minimised Mean Squared Error (MMSE) of fBM prediction against H

## Chapter 3

# Numerics: Real data and simulation

In this chapter, our aim is, first, to assess the  $\xi$ -estimation procedure using real data. Then, we will study if our volatility data demonstrates the widely accepted long-memory property and how standard statistical procedures might wrongfully capture it under model misspecification. In order to do so, we will study conjointly real and RFSV-model simulated data by comparing how well the model replicates the autocovariance structure of the data. Additionally, we will see three classical long-memory tests that lead to contradictory conclusions: our data is seen to have long-memory by two of those tests but the power-law decay and our RFSV model say otherwise. Finally, we will be interested in comparing the predictive power of the RFSV model against Auto-regressive benchmark model as well as HAR model (Corsi, [20]).

### 3.1 Datasets

For this report, we will use data provided by the *Oxford-Man Institute (OMI)* [54]. More precisely, we will propose a study of realized variance (RV) that was estimated using Univariate and Multivariate HEAVY models from Noureldin, Shephard and Sheppard (respectively [63] and [48]) as well as Realized Kernel estimations using seminal works of Barndorff-Nielsen and Hansen ([5]) and more recent developments (see Literature reference page on the OMI website for the full list).

In addition, we will supplement the study using two others datasets. The first one is *OMI's Realized Library* from which we picked 30 Equity and MSCI Indices with daily 5-minute Realized Variance (Univariate HEAVY model) and Realized Kernel estimates from 03/01/1996 up to 03/01/2009 and can be characterized by a large number of missing values. The second one is the so-called *Live Data Library* which contains daily Realized Variance (Multivariate HEAVY) and Kernel estimates for 15 Assets from both Equity and Foreign Exchange rate for a larger period of time spanning from 03/01/2000 up to 19/05/2017 with reasonable number of missing values. Both datasets have assets in common which will be helping with regards to the discussion on whether H is time-dependant or not.

The original paper by Gatheral, Jaisson and Rosenbaum provides an extensive presentation

of Equity-based data. Therefore, we will present a refreshed version of their results using the Realized and Live Data libraries but will focus on presenting a comparison between Equity and three different classes of assets:

- *Equity*: SPY and BAC 5-minute Realized Volatility data (using multivariate HEAVY 1-minute subsampling model) for its high quality and very few missing values. The dataset spans from 01/02/2001 up to 31/12/2009 with daily estimations;
- *Foreign Exchange rates*: EUR/USD 5-minute minimum Realized Variance (MinRV, [4]) from 26/07/2005 up to 03/05/2017. Bitcoin(BTC)/USD 5-minute Realized Variance from 14/08/2011 to 27/04/2017;
- *Commodities*: 5-minute MinRV for West Texas Intermediate (WTI) Oil price from 18/12/2003 to 26/04/2017 and for Gold price in USD from 27/04/2017 to 20/04/2006. The latter is named XAUUSD for historical and conventional reasons.
- *Fixed Income*: The 5-minutes MinRV 2, 5 and 10 Years U.S Treasury Bonds (resp. from 12/09/2005 to 15/03/2017, from 21/09/2010 to 21/04/2017 and from 13/04/2009 to 20/04/2017);

The complementary data was provided by OANDA using their Python REST-v20 API (see [19]) and a wrapper we coded for the occasion. Because the data is available under Japanese candlesticks, we used the Close-to-Close returns on the mid-price. We obtained from 400,000 up to 900,000 data points, each one correspond to 5 minutes periods. For Bitcoin data, the rate BTC/USD was extracted from BitCoin Charts ([11]) using a BitCoin marketplace trading activity, namely of BTC-e, and obtaining 30,000,000 data points that uses the most recent traded price.

We can observe the quality of the data is different from both sources as one provides a summary every 5 minutes, implying jumps, whereas the other yields the last traded price of any transaction (within seconds). This is what motivated the use of two different measures of volatility.

*Remark.* The idea to include Realized Kernel estimates is to provide a comparison with a noise-robust estimate of volatility (see [5]). On the other hand, MinRV from Andersen et al. [4] is a jump-robust RV estimator. We used an implementation from the R package **highfrequency**, [12]. Finally, when not stated otherwise, the Realized Variance is computed using the standard approximation (see [59], Example 1) with S being the observed price process:

$$\sum_{k : 0 \leq t_0 < \dots < t_k < \dots < t_N \leq t,} \left( \frac{S_{t_k} - S_{t_{k-1}}}{S_{t_{k-1}}} \right)^2$$

which converges in probability to  $\int_0^t \sigma_s^2 ds$  (see pioneering work by Barndorff-Nielsen and Shephard, [6]).

## 3.2 The long-memory property

A stylized fact about volatility is the *long-memory* property (see Definition 3.2.1). A large part of the literature comment such behaviour and finds models that satisfy this property (see [25], [2] partially written by, respectively, Engle and Andersen two of the main figures in the field of volatility. Also, [18] for the FSV model, [20] for the HAR model). From the seminal paper on long-memory property by McLeod and Hipel [46] (used for instance in [29] and [18]), we have the following definition:

**Definition 3.2.1** (Long-memory or long-range dependence property).

A stochastic process  $(X_n)_{n \in \mathbb{N}}$  with covariance stationary time series (where  $j$  is called the lag)

$$(\rho(j))_{j=-\infty}^{+\infty} := (\text{Cov}(X_n, X_{n+j}))_{j=-\infty}^{+\infty} \quad \text{for } n \in \mathbb{Z}$$

is said to verify the *long-memory property* if

$$\sum_{j=-\infty}^{+\infty} |\rho(j)| = +\infty$$

On the other hand, it verifies the *short-memory property* if

$$\sum_{j=-\infty}^{+\infty} |\rho(j)| < +\infty$$

*Remark.* In practice, stationarity is in the sense of *weak* (second-order) stationarity, that is to say that the mean and variance are constant in time and  $\text{Cov}(X_k, X_j)$  depends only on  $|k - j|$  (See Section 7.4, p 266, Walden and Percival [55]).

*Remark 3.2.2.* The intuition behind this definition is quite clear but the problem with real data is that we cannot be sure that such property holds from the finiteness of data. A common proxy to verify this property is the use of an approximate power law for the covariance function (Dieker, [24]). More precisely, using the same notation as in Definition 3.2.1, suppose as  $|j|$  goes to  $+\infty$ , we have:

$$\rho(j) \sim K|j|^{-\gamma}, \quad K > 0$$

where  $f(x) \sim g(x) \Leftrightarrow \lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = 1$  for  $g(x)$  non-zero for  $x$  large enough. Then, if  $0 < \gamma < 1$ , we have that  $X$  verifies the long-memory property.

## 3.3 Estimating H using $m(q, \Delta)$

### 3.3.1 Intuition

It is not immediately clear why having an information on the regularity of paths (as discussed in Section 2.5) would immediately lead to a result concerning the long-range dependence of the path. This relies on a popular result given in (2.2) (used for example by Nuzman and Poor [51], Section 2.2) that we have for  $s$  large enough and  $H \neq 1/2$ :

$$\mathbb{E}[W^H(1)(W^H(s+1) - W^H(s))] \sim H(2H - 1)s^{2H-2}$$

Therefore, having  $H > 1/2$  (respectively  $H < 1/2$ ) yields that  $W^H$  verifies the long-memory property (resp. short-memory property).

### 3.3.2 Numerical results

#### 3.3.2.1 SPY

Applying the afore-mentioned  $\xi$ -estimation procedure on SPY, it yields that  $H \approx 0.16$ . Given this and under the assumption that the volatility verifies the RFSV model (using the notations of Section 2.4), Cheridito et al. [16], Section 2, provide a *inverted* formula to estimate  $\nu$ , the so-called volatility of volatility in the model:

$$\text{Var}[\log(\sigma_t)] = \frac{H(2H - 1)\nu^2}{\alpha^{2H}} \Gamma(2H - 1) \quad (3.1)$$

where  $\Gamma(x) := \frac{1}{x}\Gamma(x + 1)$  for  $x < 0, x \notin \mathbb{Z}$ . In our example, using the mean value for H, we have  $\nu \approx 0.23$ . This looks reasonable as the authors of [29] found  $H = 0.14$  and  $\nu = 0.3$  on their SPX data ranging from 2000 to 2014: as SPX and SPY are not the same assets but represents the same financial structure we would expect different although comparable estimated values of H and  $\nu$ . See Figure 3.1.

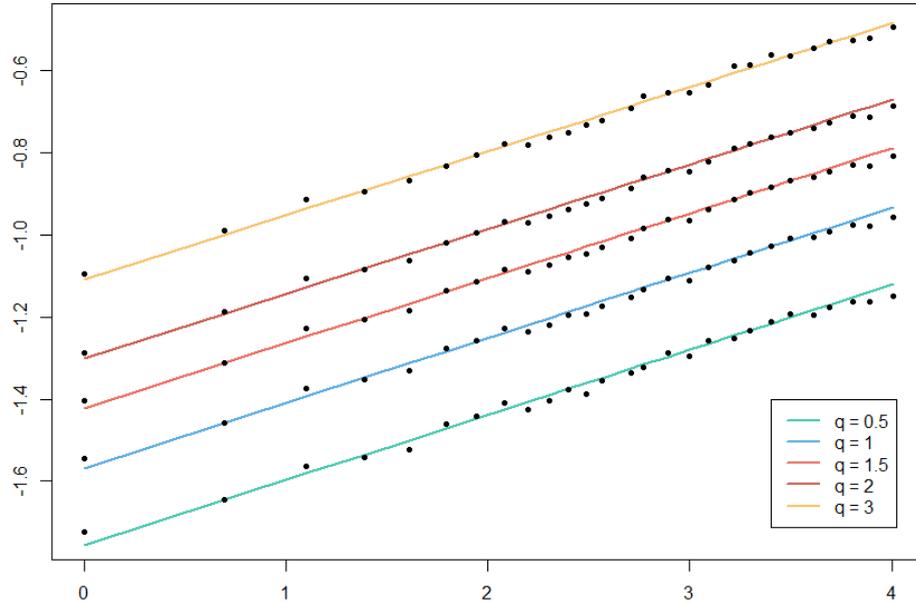


Figure 3.1:  $\log(m(q, \Delta))/q$  vs  $\log(\Delta)$ :  $\xi$ -estimation procedure on real market SPY data with OMI Realized Variance estimates. If the model were correct, we should see the parallel lines we observe here. Here, the slopes are the estimated parameter  $H$  for each value  $q$ , similar for all  $q$ .

#### 3.3.2.2 OMI Datasets

Ignoring missing values, we can do the same procedure on the OMI's Realized Library. See Appendix D.1, Tables D.4 and D.5 respectively using Realized Variance and Realized

Kernels. As previously mentioned, we observe that the  $\xi$ -estimation procedure gives (for the most part) an almost monotone progression as  $q$  gets larger and on some examples like the *FTSE Milano Italia Borsa MIB30\_rv*, the ratio  $\xi_q/q$  is constant. This seems reasonable as the data is quite limited in length so some numerical bias is expected in the estimation. On the other hand, there is no reason why the ratio  $\xi_q/q$  should increase or decrease (or remain constant in the best-case scenario) given this bias and we also see that such increasingness is not always verified, which is entirely fine. The important observation is that all assets' volatility seem to be best-described by a Hurst parameter comprised between 0.09 and 0.15 which implies a short-memory property which contradict the stylized fact of long-range dependence of log-volatility. Comparing those results with the Live Data Library (Tables D.1 and D.2), we observe three main features:

- *SPX2.rv* was used in [29] and we obtained very similar  $H$  estimates (they found  $H \approx 0.14$ , we are close to  $H \approx 0.15$  here) as they have although it was on data ranging from 2000 to 2014 (here from 2000 to 2017). This goes in the direction of the authors' views on why we should not be too worried about  $H$  being dependent on time at least at on larger time scales.
- We cannot find any substantial difference in the behaviour of new assets of the Live Data Library compared to the ones shared with Realized Library. Also, MSCI indices perform in a similar fashion as other Equity indices.
- Across all the datasets, we can observe that Hurst exponent estimates using Kernels are systematically larger than the estimates using Realized Variance which is what we should expect. Indeed, a larger  $H$  implies smoother paths and from the robustness of Realized Kernel estimates with respect to noise, they should only capture the significant movement in price. On the other hand, Realized Variance estimates take noise as significant movements and therefore represent rougher paths.

### 3.3.2.3 Other asset classes

When looking at the additional asset classes we introduced, we made the observation that one of the particularity of the data obtained from OANDA are *jumps*. They come from the fact we have access to data with a fixed time granularity (as opposed to trade-based data like BTCUSD). Also, for example the U.S. Treasury Bonds can be largely and quickly affected by some news coming from the U.S. Federal Reserve. To tackle this issue, we used the Minimum Realized Variance (minRV) estimator given by Andersen et al. [4] (Prop. 1 and 2). For a signal  $Y$  of which we would like to estimate its variance:

$$\text{MinRV}_N := \frac{\pi}{\pi - 2} \left( \frac{N}{N - 2} \right) \sum_{i=1}^{N-1} \min(|Y_i - Y_{i-1}|, |Y_{i+1} - Y_i|)^2$$

Quoting the authors, this estimator is not *particularly efficient* because jumps are truncated from the time series therefore ignoring the information they contain about the variability of the asset. Therefore, although we cannot trust the values of  $H$  in the absolute sense, we know that such estimator look at a *smoothed* version of the price path which is all we

need. Indeed, all in all, we are interested in inferring if the long-memory property holds (see Section 3.2 for details), that is to say if  $H < 1/2$  or not. Hence, if we obtain an estimate  $\hat{H}$  such that  $\hat{H} \ll 1/2$ , we can reasonably hope that most accurate estimates will also yield  $H < 1/2$  implying such long-range dependence does not hold. Because we rely on real data, some minRV estimates were very close to zero and we needed to redeem such values as not consistent. To do so, we used the usual trick to ignore the values smaller than the  $\alpha$ -th quantile. The evidence that made us believe that our minRV estimates captured the underlying structure behind volatility is that the value of  $\alpha$  did not really matter in the estimation of  $H$  as long as  $\alpha > 10\%$ . More precisely we could not notice any significant change in the output value of our  $\xi$ -estimation procedure. Therefore, we choose to ignore the smallest 20% of the estimates.

Finally, concerning the BTC/USD data, because it was generated after each trade two features are observed: first, we do not have significant and large jumps because we have access to ultra high frequency data (a few seconds maximum between each couple of data points), therefore we do not need to use MinRV estimator. Also, we do have a good convergence to the theoretical daily integrated variance from the abundance of data for each day. Therefore, we chose to use the standard RV estimator.

*Remark.* When using the standard RV estimator on the data from OANDA, we obtain negative values for  $H$  to cope with abrupt jump-like changes in volatility which is captured as an extreme roughness. Of course, these results are not correct and the reason behind this feature is the continuity assumption of the underlying price path that is required to have a consistent standard RV estimator (briefly implied in [59] in the specific case of the Uncertainty Zones model).

Looking at our result presented in Table D.3, we see the difference between the jump-robust MinRV estimator and the simple RV estimator that includes noise as significant. For the latter, we obtain reasonable values of  $H$  from 0.091 up to 0.159 that is very close to the values we obtained from the Equity-based indices. On the other hand, for the former, we obtain estimates ranging from 0.08 to 0.035 which seems very small. As explained above, we cannot trust those values in the absolute sense but we can reasonably think that the true value for  $H$  would be below  $1/2$ . During a conference at Cornell University<sup>1</sup>, Jim Gatheral explained using Appendix C, [29], concerning the smoothing of integrated variance estimators, that a particular scaling of integrated variance could lead to values of  $H$  as small as 5%. In addition, the MinRV estimator is quite inaccurate from its tendency to be affected by an *exposure to small (zero) returns* as stated by the authors Andersen et al., we believe that the small estimated values for  $H$  are not surprising.

Note that in the rest of the report, we will focus on comparing SPY with the other assets classes, namely the assets presented in Table D.3. To see a comprehensive study of the equity-based indices, see the original paper [29].

---

<sup>1</sup>[www.orie.cornell.edu/engineering2/customcf/iws\\_events\\_calendar/files/RoughVolatilityCornell2015.pdf](http://www.orie.cornell.edu/engineering2/customcf/iws_events_calendar/files/RoughVolatilityCornell2015.pdf)

### 3.3.3 Numerical evidence of the RFSV Model

The RFSV model is not only intuitive in its construction (mean-reverting behaviour in the log-volatility, convergence in distribution to an affine function of fBM), it is also based on insightful observations made by Gatheral, Jaisson and Rosenbaum [29].

We can observe the following features already built in the model:

- The mean-reverting property has been thoroughly observed and studied with many models being equipped with this property (see [26] for a classic study of the matter). It also relaxes any implicit conditions we would apply to  $H$  to fit the behaviour of the data to let  $H$  be only there to represent roughness;
- The increments of log-volatility showcase the traditional Gaussian distribution (as discussed by Andersen et al. [2]) and seem to verify the *self-similarity property* that is central in the definition of fBM. In order to show this, we use *SPY* real data as pictured in Figure 3.2. The histograms of the log-volatility increments with time difference of 1, 5, 20 and 120 days (representing one day, a week, a month and 6 months). In addition, we plotted in *green* a Normal pdf fitted to the data using standard unbiased MLE estimators for the mean and the standard deviation. Finally, we plotted the Normal pdf obtained for  $\Delta = 1$  day in *orange*, scaled using the  $H$ -self-similarity property of fBM and the Hurst parameter estimated by the  $\xi$ -estimation procedure. As noted in [29], we observe that the scaled and fitted Gaussians are *very*

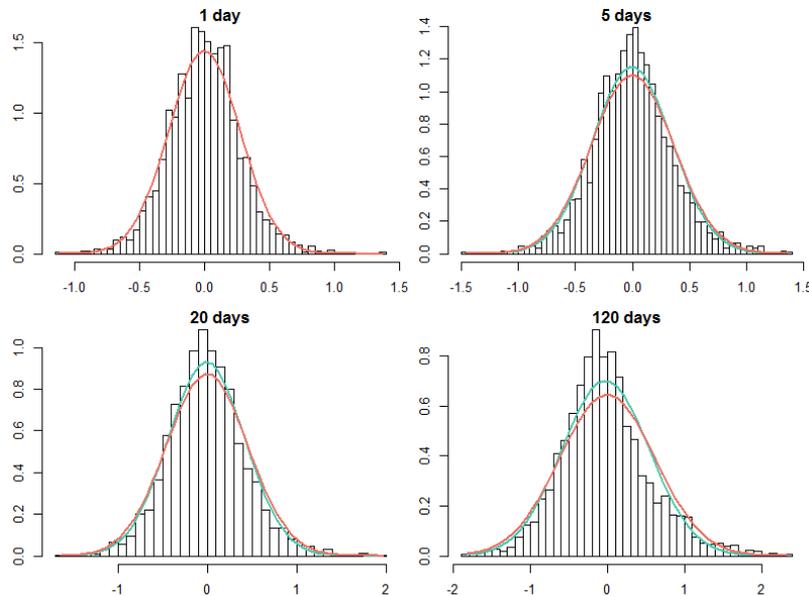


Figure 3.2: SPY Data: Increments of log-volatility against Gaussian with corresponding  $\Delta$  (5, 20 or 120 days) (in green) and Gaussian for  $\Delta = 1$  (in orange, scaled using self-similarity)

*close* to each other. This encourages to pick a value of the parameter  $\alpha$  in the RFSV model very close to 0 to achieve a good approximation given by Proposition 2.4.4. This would model both *normality and self-similarity at the same time*.

Now that we have shown why defining the RFSV model this way made sense in theory, we can also see that the data showcases what the model would predict if it were true in terms of *covariance* in both log-volatility and volatility. However, it is interesting to observe that *those properties are not universally verified across the different assets*.

- **Covariance of log-volatility:**

Following [29] and Proposition 2.4.1, we have *by stationarity* for  $\alpha$  close to zero:

$$\text{Cov}(X_{t+\Delta}^\alpha, X_t^\alpha) = \text{Var}(X_t^\alpha) - \frac{1}{2}\nu^2\Delta^{2H} + o(1) = -\frac{1}{2}\nu^2\Delta^{2H} + A + o(1), \quad A > 0 \quad (3.2)$$

Therefore, on log autocovariance against  $\Delta^{2H}$  plots, we should see a straight line with negative slope which is exactly what we observe in Figure 3.3 for SPY and BTCUSD. However, volatility estimates coming from the MinRV estimator, namely for XAU-

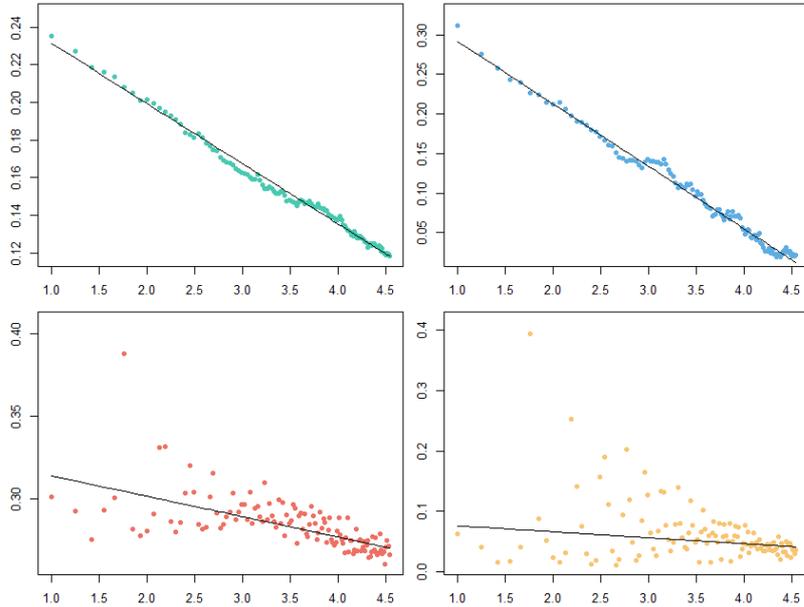


Figure 3.3: Log-volatility autocovariance  $\text{Cov}(\log(\sigma_{t+\Delta}), \log(\sigma_t))$  against  $\Delta^{2H}$ : SPY data (top-left), BTCUSD data (top-right), US 2Y Bond (bottom-left), XAUUSD (bottom-right).

USD and U.S. 2-Year Bond, are more spread out in a cone shape (in addition, other assets like EUR-USD behaved the same way as the bottom plots but is not reported). This said, we can still see that the general shape of a affine function seems reasonable.

- **Volatility cross-moments and autocovariance:**

Using Proposition 2.4.4, they also derive that approximately as  $\alpha \rightarrow 0$ , we have that there exists  $A, B > 0$ :

$$\mathbb{E}[\sigma_{t+\Delta}\sigma_t] = Ae^{-B\Delta^{2H}} \quad (3.3)$$

This would imply that, on a semi-log scale, we shall observe a straight line for  $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$  against  $\Delta^{2H}$ . This what we obtain in Figure 3.4 although BTCUSD

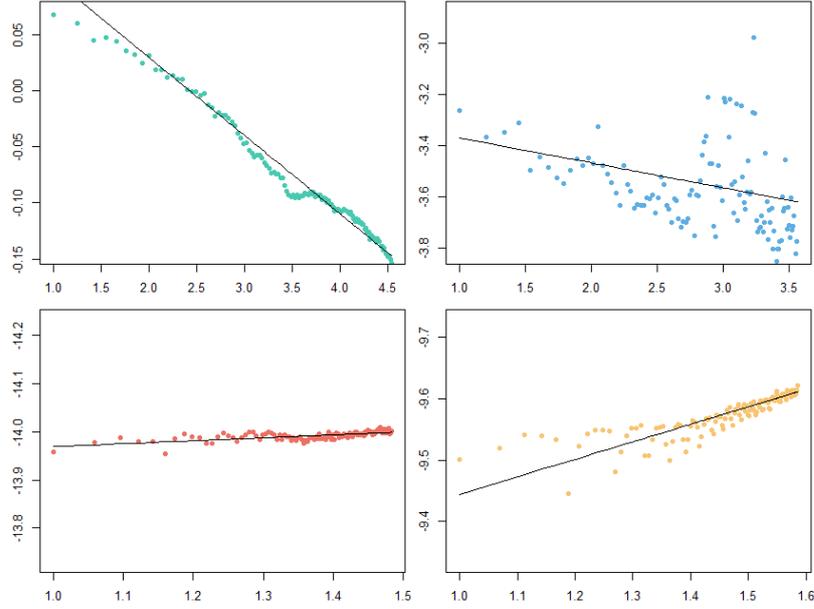


Figure 3.4:  $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$  against  $\Delta^{2H}$  on a semi-log scale: SPY data (top-left), BTCUSD data (top-right), US 2Y Bond (bottom-left), XAUUSD (bottom-right).

is slightly more spread out (but, again, the straight line is still capturing the general shape). Observing that the mean of volatility is not time-dependent, we shall expect for  $D > 0$ :

$$\text{Cov}(\sigma_{t+\Delta}, \sigma_t) \approx Ae^{-B\Delta^{2H}} - D \quad (3.4)$$

We can further see that a log-log scale for  $\text{Cov}(\sigma_{t+\Delta}, \sigma_t)$  against  $\Delta$  would not yield a straight line for SPY, BTCUSD as pictured in Figure 3.5.

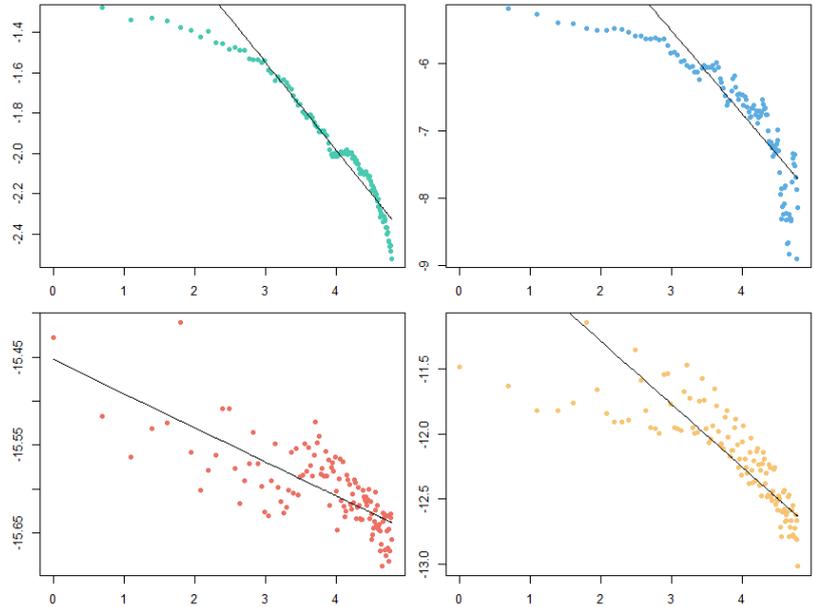


Figure 3.5: Volatility autocovariance  $\text{Cov}(\sigma_{t+\Delta}, \sigma_t)$  against  $\Delta$  on a log-log scale: SPY data (top-left), BTCUSD data (top-right), US 2Y Bond (bottom-left), XAUUSD (bottom-right).

We clearly see a downward bell shape for SPY, BTCUSD and XAUUSD which seems to correspond to (3.4). See Section 3.5.1 to know what this entails regarding the long-memory property. Still, it is worth noticing that the U.S. 2-Year Bond seems to be relatively well described by a straight line which would imply a power-law decay (and potentially, long-memory property) contradicting the Hurst parameter value below 1/2. This is probably one of the elements that hinted the authors of [29] to investigate classical long-memory tests.

*Remark.* The afore-mentioned properties are all verified by the other asset classes data. Due to length limitations, we could not report those results.

The RFSV model seems plausible from many points of view: from its *built-in* features (e.g. the log-volatility fractal-like/self-similarity scaling from the fBM) to its implied properties (e.g. covariance as  $\alpha \rightarrow 0$ ) are both observed in real data. Still, it remains to compare the real data integrated volatility paths we have against simulations of the RFSV model. Indeed, it will let us compare directly what the model captures the autocovariance structure.

## 3.4 Simulation study

### 3.4.1 Procedure

From its definition, the RFSV model can be simulated given only the initial value of log-volatility and the Hurst exponent. In order to do so, we need a two-step process to estimate the fractional Ornstein-Uhlenbeck process for the log-volatility  $X^\alpha$  and then estimate the price process  $S$ . Following [29], we will use Euler-Maryuama discretisation scheme on the coupled SDEs from (2.6):

$$\begin{cases} dX_t^\alpha &= \nu dW_t^H + \alpha(m - X_t)dt \\ dS_t &= S_t \exp\{X_t^\alpha\}dW_t \end{cases}$$

where  $W$  is a standard Brownian Motion and  $m$  is the initial condition on  $X$  at time 0. It gives the following set of discrete equations:

$$\begin{cases} X_{(n+1)\delta}^\alpha &= X_{n\delta}^\alpha + \nu(W_{(n+1)\delta}^H - W_{n\delta}^H) + \alpha\delta(m - X_{n\delta}) \\ S_{(n+1)\delta} &= S_{n\delta} + S_{n\delta} \cdot \exp\{X_{n\delta}^\alpha\} \cdot (W_{(n+1)\delta} - W_{n\delta}) \end{cases} \quad (3.5)$$

where  $\delta$  represents the time interval between two intraday values of price and volatility. We chose the granularity so that we estimate  $2^{11} = 2048$  days with  $2^{14} = 16'384$  intraday timesteps. This will give us a sample path made of  $2^{25} = 33'554'432$  data points to estimate 2'048 integrated variance estimates. Therefore, we chose  $\delta = 2^{-14}$  and we set  $\alpha = 5 \cdot 10^{-4}$  as the tick size of our data.

For the different simulations, we use the mean value of estimated Hurst exponents for  $q = 0.5, 1, 1.5, 2, 3$  as in [29]. More precisely, we used:

	SPYUSD	BTCUSD	EURUSD	XAU	WTI	US 2Y	US 5Y	US 10Y
H	0.157	0.133	0.035	0.048	0.078	0.041	0.047	0.043

Table 3.1: Estimated Hurst exponents for each asset used in the simulation.

*Remark.* The *spot* volatility for day  $k$  is defined and approximated by:

$$\sigma_{t_k} := \exp\{X_{t_k}^\alpha\} \approx \frac{1}{\delta} \int_{t_k}^{t_k+\delta} \sigma_s^2 ds \quad (3.6)$$

where  $\delta$  is the aggregating time given by the estimation method.

To estimate the integrated variance, we used the following procedure:

1. Generate the theoretical - unobserved - price path  $S$ ;
2. Generate observed prices with the model with *Uncertainty Zones (UZ) for high-frequency data* for tick size  $\alpha = 5.10^{-4}$  and uncertainty band width  $\eta = 0.25$  designed by Robert and Rosenbaum [59] (see Appendix C);
3. Using those observed prices, estimate:
  - daily UZ integrated variance with one hour of data ( $\delta = 1/24$ ; from 10AM to 11AM) (see again Appendix C);
  - daily standard realized variance (RV) for 8 hours of data ( $\delta = 1/3$ ; from 8AM to 4PM);

To simulate the fractional Ornstein-Uhlenbeck process, we used the R package **yuima** [13] based on the YUIMA project which is very impressive in efficiency and content. More precisely, we relied on the celebrated Gaussian process simulation method by Wood and Chan [67] which is itself based on the characterizing Gaussian covariance structure and the formation of a so-called *circulant* matrix, a specific case of a Toeplitz matrix. This allows a discrete Fourier transform implementation as used in the original paper [29].

Other alternatives include the **fArma** [68] or the **somebm** [36] packages. The former proposes various implementation of fBM generation whereas the latter uses a spectral scheme (Fast Fourier Transform) that is very convenient to use for large path simulations.

### 3.4.2 Testing the simulation paths via autocovariance function

The idea here is to take the real market data for each asset and compare with the simulated paths to see if the RFSV model capture the autocovariance structure. Indeed, this is the main reasoning behind the original paper [29]. We study their ACF up to lag 100 as follows:

- **On SPY, BTCUSD, U.S. 2-Year Bond and XAUUSD (Figure 3.6):**

We note that, as stated in [29], SPY market data and simulation look very much alike. This has driven the authors to the conclusion that they had found a model based mostly on a few parameters including the Hurst exponent that could reproduce statistical properties of a whole price path. However, we observe here the limit of such model: indeed, we have that XAUUSD and US 2-Year Bond have Hurst parameters that are very close to each other. Because  $\nu$  depends solely on  $H$  (as in (3.1)), the whole path is mostly determined by  $H$  in the RFSV model. When simulated using the same seed, we see that the simulated path share the same autocorrelation properties and do not showcase any sign of seasonality. However, with real market data, they

are quite different from each other from the way the ACFs decrease to how strong the weekly seasonality is. Indeed, real market US 2Y Bond ACF is almost constant at  $1/2$  whereas XAUUSD ACF shrinks rapidly to zero. In addition, real market BTCUSD ACF plummets to zero whereas its simulation is very close to SPY - decaying slowly, again this is due to the fact that their Hurst parameters are very close to each other! This is one of the limit of the RFSV model. We explain this fast decay by the trading-wise immaturity of the BitCoin market: it is not a central as an asset as the U.S. Bond for example. Indeed, the latter has an almost constant ACF of  $1/2$  since those bonds are essentials in financial markets (many people actively manage their portfolio of bonds every day).

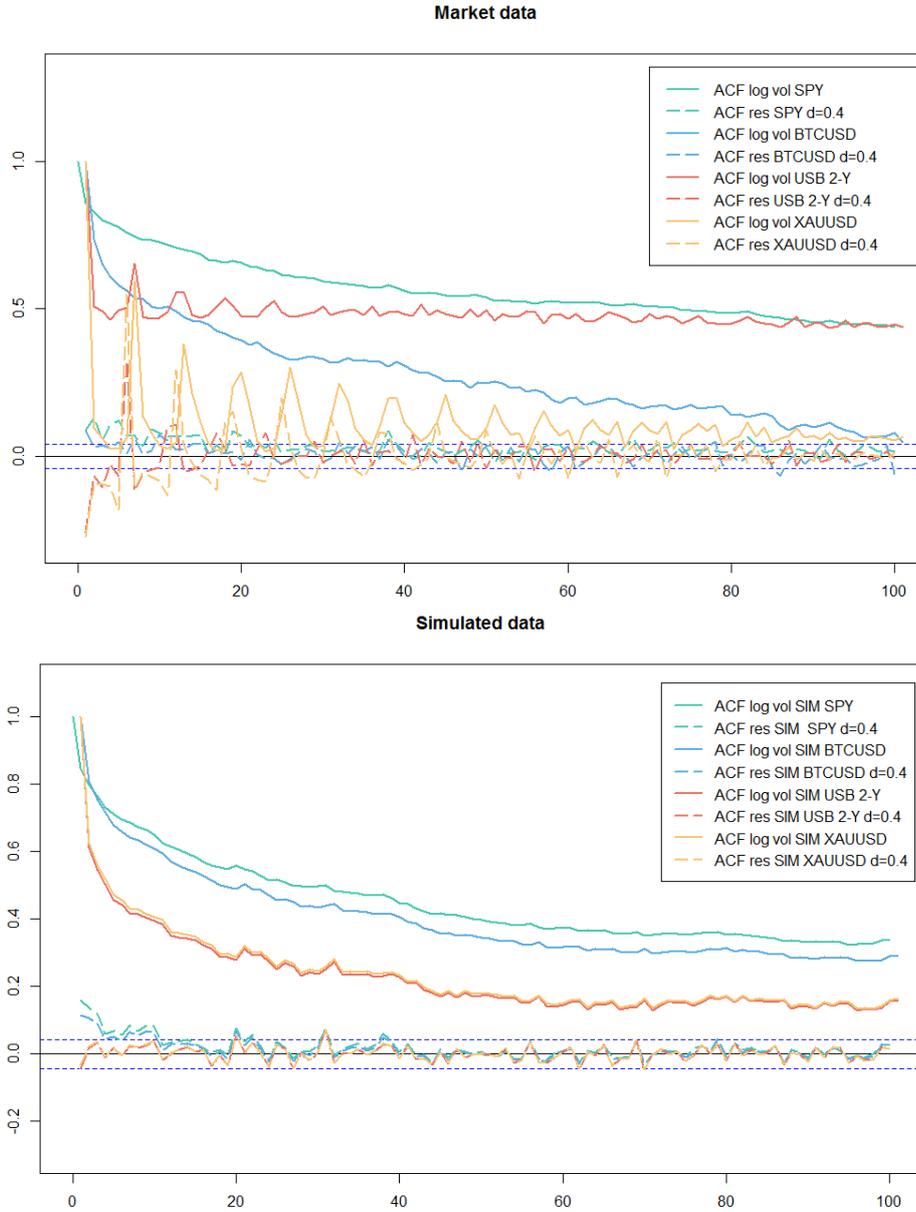


Figure 3.6: ACF up to lag 100 for the log-volatility (solid) and the differenced log-volatility (dashed) for both market data (TOP) and simulated data (BOTTOM).

This said, we did not expect seasonality to be replicated by the RFSV since it has never been considered in its design in [29]. However, an extension using external regressors could be implemented.

- **The special case of EURUSD (Figure 3.7):**

The RFSV model can also be *very effective* for other assets than equity: we observe that both real and simulated EURUSD paths resemble each other, much more than any other couple of paths.

Indeed, in Figure 3.7, we notice the weekly seasonality which may be explained by

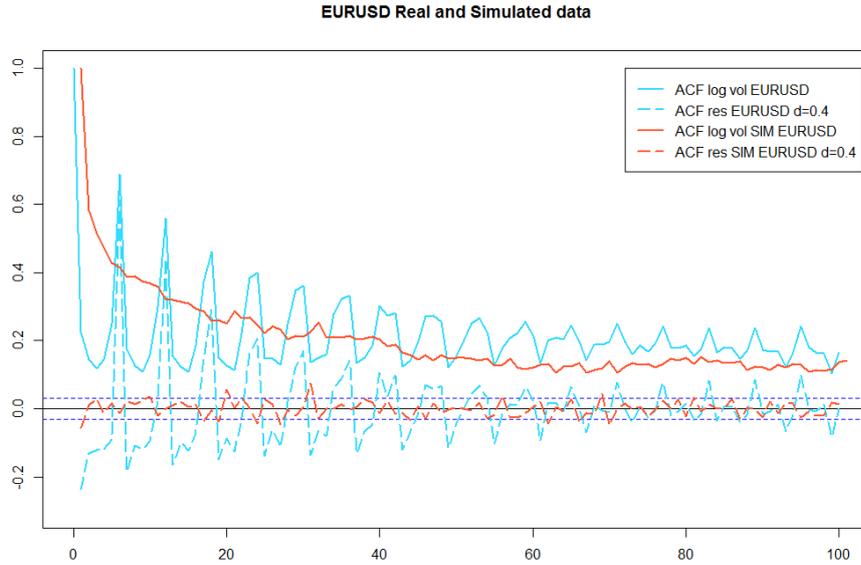


Figure 3.7: ACF up to lag 100 for real market (blue) and simulated (red) EURUSD rate: an example of an asset which is well-described by the RFSV model.

the adjustment pre and post-weekends which are the only times when FX markets are closed. However, unlike XAUUSD, this does not prevent the ACFs to present the same speed of decay. This indicates that the asset class may not be the most relevant way to classify asset in the context of the RFSV model. Indeed, the model may fail due to inherent characteristics of the assets but could be features that can be shared among different classes.

### 3.4.3 Testing the $\xi$ -estimation procedure

We can now check whether the simulation yields a correct Hurst parameter when we apply the  $\xi$ -estimation procedure. We obtain relatively good estimates depending on which measure of volatility we use. As explored by Gatheral et al. [29] and explained by their Appendix C on scaling of volatility, we observe that the UZ variance has some positive bias and the RV variance has an even larger positive bias. Indeed, as they explain: the longer we wait, the larger the bias is. Explicitly, we compute an approximate to the *spot* volatility using the mean value as in (3.6): the integral plays its smoothing role on the integrand as we *wait* longer, making the path *seen* by  $m(q, \Delta)$  *smoother* than in reality. This would

lead to over-estimating  $H$ . This initiates a trade-off: we want to wait sufficiently long to have estimates close to the true IV values, however, we do not want to wait too long since those IV values would be positively biased. For instance, with SPY simulated path, we should be getting  $H \approx 0.16$  but we obtain  $H_{UZ} \approx 0.185$  and  $H_{RV} \approx 0.21$ . See Figure 3.8 where the slopes  $H$  of the RV estimates (dashed lines) are larger than the slopes of the UZ estimates (solid lines) for each value  $q$ . We use  $\delta = 1/24$  for the UZ method and  $\delta = 1/3$  for the standard RV method hence the difference in bias. However, we do obtain parallel lines showing that the estimated Hurst exponents is approximately constant across values of  $q$  as desired.

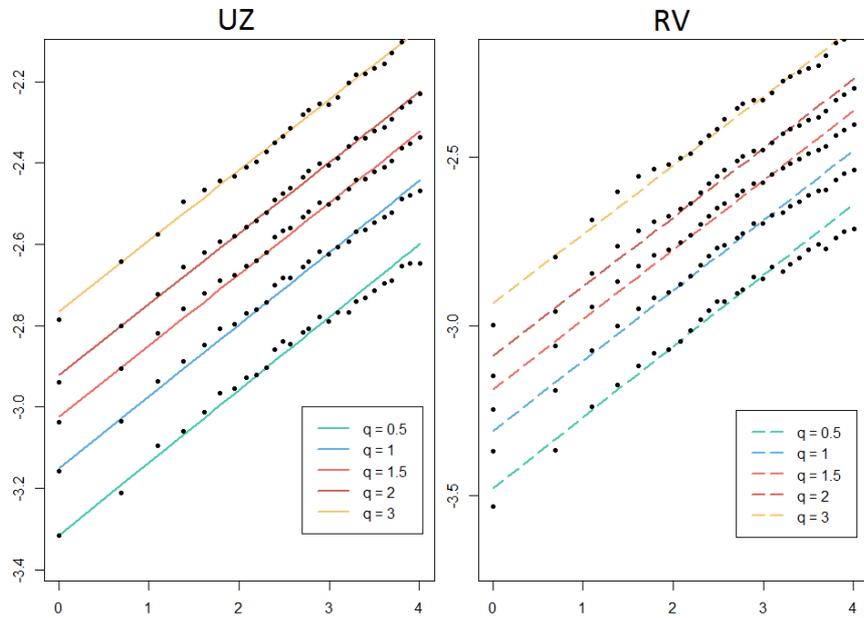


Figure 3.8:  $\log(m(q, \Delta))/q$  vs  $\log(\Delta)$ :  $\xi$ -estimation procedure on simulated SPY data. LHS: UZ integrated variance estimates; RHS: RV integrated variance estimates. Slope of each line is the estimated Hurst exponent  $H$ : we find  $H_{UZ} \approx 0.185$ ,  $H_{RV} \approx 0.21$ .

Concerning the other assets, all computations are consistent with what we would expect from the above-mentioned explanation on bias and Table 3.4.1 hence we omit the details here.

*Remark.* Note that the  $\xi$ -estimation procedure gives a Hurst parameter that is essential in the scaling of increments of log-volatility. However, it remains uncertain if a linear link between  $H$  and  $\xi$  should suffice to model properly the data. To test this hypothesis, Gatheral et al. [29] detail this relationship by computing  $m(q, \Delta)$  for many values of  $\Delta$  and  $q$  and simply plot  $\xi_q$  against  $q$ . For SPY and BTCUSD data, we obtain quite different results as we can see in Figure 3.9 below:

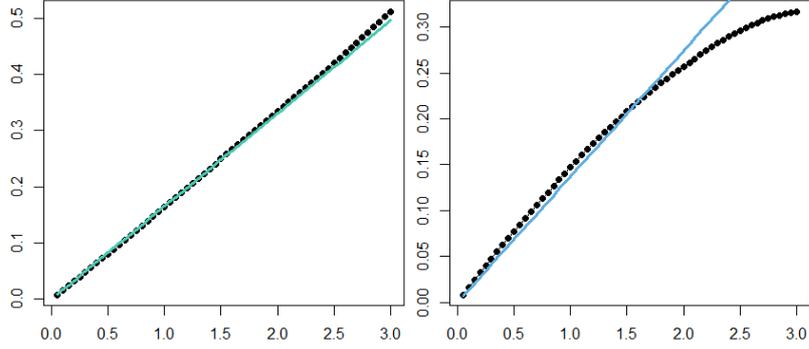


Figure 3.9:  $\xi_q$  against  $q$ : SPY data (LHS), BTCUSD data (RHS). Fitted straight lines in colour,  $\xi_q$  in dotted lines.

For the BTCUSD data, we see a convex curvature which seem to indicate that  $\xi_q \approx qH$  is inaccurate even as a first order assumption. After testing on different data sets, we made the same observation as the authors did: The smaller the dataset the stronger the curvature. Here, the SPY data contains 2242 data points against 2050 for BTCUSD. Also, when shrinking down the size of the data set of the EURUSD data set from 3700 data points down to 2000, we obtain roughly the same curvature as for BTCUSD. However, it is worth noticing that the curvature need not be convex (for example, the EURUSD data showcases concave curvature).

### 3.5 Long-memory tests

We shall revisit the arguments and tests of [29] to see that if the model is misspecified, the data can be shown to verify properties that we know are not true or leading to contradictory conclusions. Those tests are mostly based on the work of Andersen-Bollerslev-Diebold-Ebens/Labys ([1], [2], [3]). However, when exploring other asset classes, some of the results obtained were not as clear as it appears in the original paper and sometimes in a surprising way. First, we are interested in power-law decay in the autocovariance function.

#### 3.5.1 Failure of the Power-law method

Investigating the power-law decay in the autocovariance function is the most classical way to determine if the (log-)volatility process exhibits long-range dependence. As noted in Remark 3.2.2, if the autocovariance function of the spot (log-)volatility decays as a power law since it would ensure we verify the long-memory property. More precisely, we look for a behaviour of the following functional form:

$$Cov(Y_{t+\Delta}, Y_t) = A\Delta^{-\gamma} + B, \quad A, B \in \mathbb{R} \quad (3.7)$$

with  $0 < \gamma < 1$  and  $Y_t = \log(\sigma_t)$  or  $Y_t = \sigma_t$ . It is a way to formalise that we would like the autocovariance function to *slowly* decrease to zero as  $\Delta \rightarrow +\infty$ , slowly enough so that Definition 3.2.1 is verified.

When looking at our market data, we observe that such observation cannot be made for SPY, BTCUSD, XAUUSD, U.S. 2-Year Bond as opposed to the what is widely taken as

granted. Indeed, in Figure 3.3, we obtain straight lines consistent with the RFSV model (as in (3.2)) - that is, log-volatility autocovariance as an affine function of  $\Delta^{2H}$ :

$$Cov(\log(\sigma_{t+\Delta}, \sigma_t)) \approx A'\Delta^{2H} + B'$$

However, it clearly shows that (3.7) is not suitable here since the *exponent is positive*. In Figure 3.4 where we observe the desired *volatility* autocovariance function up to an additive constant, we note again that we do obtain the necessary straight lines given by the RFSV model in (3.3) (since we are in semi-log scale). This also rejects (3.7) and the inverse relationship between  $Cov(\sigma_{t+\Delta}, \sigma_t)$  and  $\Delta$  we would have needed for long-memory. To conclude, neither the log-volatility or the volatility have a power-law with negative power in the acf. If there is a power-law decay, it is with positive power.

We conclude that we do not have a power-law decay in the autocovariance function capable of yielding long-memory property. This associated with the fact that  $H < 1/2$  for all assets, we conclude that we do have short-memory property for all of them. Note that, as a necessary but not sufficient property, not having the right power-law decay is not sufficient to prove short-memory in itself.

### 3.5.2 Variance of Realized Variance method

In [1] Section 4.1 / Figure 6, Andersen et al. propose to model volatility as a *multivariate continuous-time stochastic volatility diffusion* defined as follows:

$$d\sigma_t = \mu_t dt + \theta_t dW_t, \quad \text{with } W_t \text{ a standard BM}$$

where  $\mu$  is predictable and integrable and  $\theta$  square-integrable process. Therefore,  $\sigma$  is directly taken to be a *continuous local martingale* which is not the case in our study due to fBM itself which is not even a semi-martingale.

They design the following long-memory test:

1. Compute the variance of the integrated variance over  $[0; t]$  using intermediary points to a sequence  $(Var(IV, t_i))_{i=1}^N$ , with  $0 \leq t_1 < \dots < t_N = t$ ;
2. Regress this sequence on  $t \mapsto t^{2-\gamma}$ ;
3. If  $0 < \gamma < 1$  then we conclude the integrated variance has long memory property - otherwise, short-memory.

This is based on the authors' result that if the variance of the integrated variance over  $[0; t]$  behaves approximately as  $t^{2-\gamma}$  for  $\gamma < 1$ , then the log-volatility autocorrelation function is proportional to  $t^{-\gamma}$ . Formally, we define the *variance* of the integrated variance as follows:

$$Var(IV, t) := Var\left(\int_0^t \sigma_s^2 ds\right)$$

The results for our dataset are as expected for the most part, however we may make two observations given Figure 3.10 and Table 3.5.2.

More precisely, with real data, we largely verify the above-mentioned long-memory test for SPY, XAUUSD, U.S. 2-Year Bond and all other assets considered (except BTCUSD). This

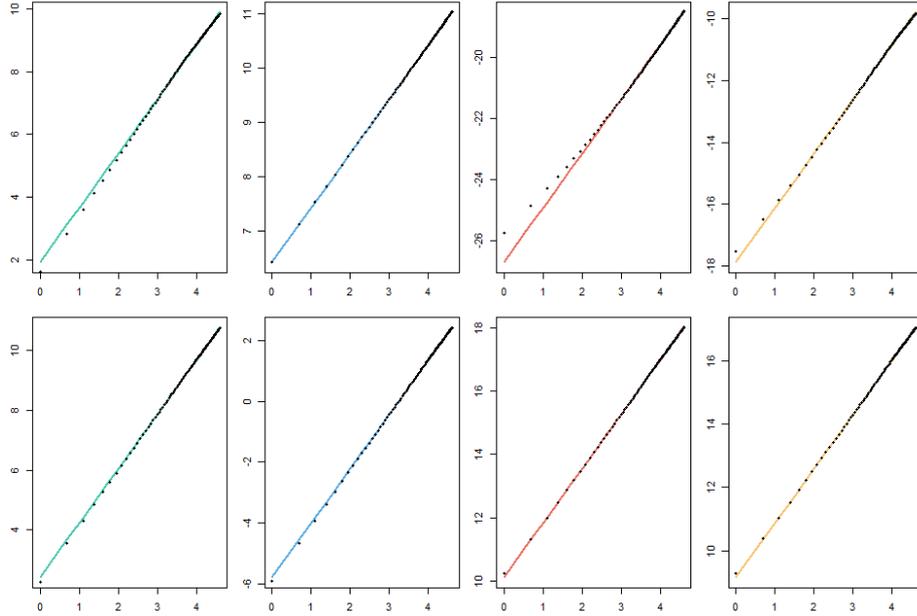


Figure 3.10: Variance of Integrated variance  $Var(IV, t)$  against  $t$  on a log-log scale for real and simulated data (resp. top and bottom plots). From left to right, SPY, BTCUSD , US 2-Year Bond, XAUUSD.

would imply wrongfully that our data verifies long-memory property which is in contradiction with the estimated Hurst exponent being below  $1/2$ .

However, for BTCUSD, we barely achieve it as  $\gamma$  is found to be 0.997 ! We can see that

	SPY	BTCUSD	U.S. 2-Y Bond	XAUUSD
<b>Market data</b>	0.265	0.997	0.212	0.238
<b>Simulated data</b>	0.177	0.213	0.283	0.281

Table 3.2: Estimation of  $\gamma$  in the Variance of integrated variance method

such power-law method on the scale of variance of integrated variance is not suitable to all datasets as we are clearly at the limit where we should keep in mind that there is some uncertainty around the estimation of  $\gamma$ . Also, proving that BTCUSD log-volatility autocovariance function does not decay as a power-law is not sufficient to show that we have the short-memory property. It is worth noticing that for the U.S. 2-Year Bond, the real data shows a slight convex shape inconsistent with the straight line we should obtain: this method is then not well-suited for all log-volatility datasets. Indeed, we observe more than simply the value of  $\gamma$  which is not covered by the test itself and might lead to inconsistent result if not property addressed.

On the simulation side, we see that our paths all largely verify  $\gamma < 1$  which is in total contradiction with the fact that  $H \ll 1/2$  for all simulations.

*Remark.* Notice how resemblant the real and simulated data plots are for all assets (except U.S. 2-Year Bond and its convex shape). Also, we can appreciate how contradictory the

conclusions are compared to Section 3.5.1. This emphasizes how fragile those tests are regarding model misspecification.

### 3.5.3 Fractional differentiation method

In another publication from Andersen et al, [3], they model in Proposition 3 the logarithmic price process  $P(t)$  as a square-integrable process such that:

$$dP_t = \mu_t dt + \sigma_t dW_t, \quad \text{with } W \text{ a standard BM}$$

where  $\mu$  is a predictable and integrable process. For  $\sigma$ , they do not define it explicitly but assume it is a square-integrable process on  $\mathbb{R}_+$ . As measure of volatility, they take the quadratic return variation on  $[t; t + h]$  which is equal to  $\int_0^h \sigma_{t+s}^2 ds$  at time  $t$ . Recall that in our RFSV model, the dynamics of volatility  $\sigma$  yield that it is not square-integrable for semi-infinite interval.

Let  $L$  be the standard backward operator:  $LX_n := X_{n-1}$ . They make the observation that *long-run dynamics of realized logarithmic volatilities are well approximated by a fractionally-integrated long-memory process* (page 3, [3]). They design a test for the long-range dependence in the volatility based on the following result:

if the fractionally differenced log-volatility  $\epsilon_t := (1 - L)^d \log(\sigma_t)$  with  $d = 0.4$  behaves as white noise, then the volatility exhibits the long memory property.

*Remark.* We take as definition for White Noise the following from Walden and Percival [55], p. 268:  $\{\eta_t\}$  is a sequence of pairwise uncorrelated RVs, each with the same mean (usually taken to be zero) and the same variance.

They conduct numerical experiments to find the optimal value for  $d$  using Geweke and Porter-Hudak GPH estimator [30], formalized and extended by Robinson [60]. Andersen et al. found that  $d \approx 0.4$  was a reasonable value for their data and the authors of [29] used it in their study under the assumption that such data was the same. However, we notice that the original paper was written more than twenty years ago and markets have undergone severe changes in trading timescale which might have affected the structure of the data. Also, using the R package **fracdiff** [52] and their implementation of the GPH estimator, we came to the conclusion that  $d = 0.55$  was probably more suitable. Although we did try  $d = 0.55$  on our data and saw that the fractionally differenced time series were closer to the behaviour of a white noise, we only report the case  $d = 0.4$ . Indeed, the argument of Andersen et al. does not hold if  $d > 1/2$  and since we would like to compare with the results of Gatheral et al. [29], we decided to continue to use  $d = 0.4$ .

	Lag	SPY	BTCUSD	US 2Y Bond	XAUUSD	EURUSD
Market data	1	2e-5	1e-4	0	0	0
	10	0	2e-6	0	0	0
Simulation data	1	0	2e-7	0.05	0.14	0
	10	0	0	0.16	0.17	0

Table 3.3: Ljung-Box test statistics on differenced log-volatility with  $d = 0.4$ ,  $lag$  as in Definition 3.2.1

Therefore, we observe a couple of features in Table 3.5.3: to begin with, none of the observed datasets exclude non-zero autocorrelation in the differenced log-volatility time series. This said, SPY and BTCUSD are the closest to fail to reject the null hypothesis of uncorrelation. This would normally exclude that  $\epsilon_t$  behaves as white noise. Moreover, we notice that with the simulated datasets do not have the same behaviour: we do not have enough evidence to show that  $\epsilon_t$  is not autocorrelated - a necessary condition to be white noise - for SPY, BTCUSD and EURUSD. On the other hand, the simulated U.S. 2-year Bond and XAUUSD fail to reject non-autocorrelation at lags 1 and 10 with 95% significance which suggest white noise behaviour happens sooner than with other assets. From a more qualitative point of view, plotting the differenced log-volatility showcases approximately constant mean and variance for all datasets, from market or simulated.

We can now complement this study with a more graphical approach: ACF plots for the first 100 lags of the differenced log-volatility  $\epsilon_t$ .

- **On SPY, BTCUSD, U.S. 2-Year Bond and XAUUSD (Figure 3.6):**

For differenced time series from real data, we observe a few spikes at smaller lags due probably to unaccounted seasonality (e.g. we retrieve 5-6 days between each spike of XAUUSD). This said,  $\epsilon$  falls rapidly within the 5% Bartlett error bands yielding the autocovariance function is statistically insignificant with 5% significance level across all 4 assets. We conclude that the real data passes this long-memory tests, although it could be confirmed by taking into account seasonality (via external regressors). Since they did not do such treatment in [29], we chose to let the time series as they were originally.

For the data generated by the RFSV, the results is even clearer: here, no sign of seasonality so no unexpected spikes. The differenced time series all fall within the Bartlett errors after lag 10. This is complemented by Table 3.5.3 and the nice results for U.S. 2-year Bond and XAUUSD. Therefore, our model let us create time series which are strongly short-range dependent ( $H$  ranging from 0.035 to 0.157) but still pass with confidence this long-memory test.

- **The special case of EURUSD (Figure 3.7):**

Concerning the asset which was the most satisfying in terms of ACF *fit* with real data, we see that we have large spikes in the real data differenced time series which shows that we *cannot* redeem this time series as insignificant. Therefore, we can see here that we fail to pass the test but it is more an artefact from an incomplete data preparation (no handling of seasonality) than any insightful result about the data. On the other hand, the model-generated data *still* proposes statistically insignificant differenced time-series  $\epsilon$  suggesting that there is long-range dependence in the log-volatility although  $H \approx 0.05$  and no power-law decay in the ACF.

To conclude, ignoring the bumps in the ACF of  $\epsilon_t$  due to the seasonality, we can see across all datasets (except real data for EURUSD) that the fractional differentiation made the ACF fall into the 5% Bartlett error bands quickly, revealing that they are no longer significant at least at large lags ( $> 20$ ). This means that  $(\epsilon_t)$  behaves like white noise in a weaker fashion - for larger time scales - suggesting long-memory behaviour in the log-volatility. We saw that the RFSV-generated data could pass without hesitation Andersen et al. long-memory test based on the differenced log-volatility. From the second-order stationarity of  $\epsilon_t$  to the pairwise uncorrelation at larger lags, we can say that it is plausible that a given time series appears to have long memory whilst being inherently short-range dependent and this even for extreme values of  $H \ll 1/2$ . We explain this through model misspecification.

## Chapter 4

# Forecasting Volatility

Forecasting volatility is an extremely popular topic and has raised many questions as it requires to understand how previous information from the market can help know better the future moves. This is in itself a problem: how to estimate volatility in a consistent manner? Across the millions of datapoints we used for this report, a question was consistently present: What scaling should we apply? What scaling has been applied?

Those questions are generally not linked to volatility forecasting (e.g. auto-regressive (AR) models are not sensitive to scaling). However, in the RFSV model, we will see the predicting volatility can be done with fewer data (compared to benchmark model like AR models) but, instead, the problem is transferred to the (mechanical) estimation of an integral. The problem is that that we would need intermediary data (which we do not have access to) to better approximate this integral via linear interpolation. From Proposition 2.4.1, as well as [29], Section 5, we have that in the RFSV model:

$$\log(\sigma_t^2) \approx 2\nu W_t^H + 2X_0^H, \quad \text{for } C \in \mathbb{R}$$

Using this and the prediction formula given in Theorem 2.6.1, for a forecast horizon  $\Delta > 0$ , we have:

$$\mathbb{E}[\log(\sigma_{t+\Delta}^2)|\mathcal{F}_t] = \frac{\cos(\pi H)}{\pi} \Delta^{H+1/2} \int_{-\infty}^t \frac{\log(\sigma_s^2)}{(t-s+\Delta)(t-s)^{H+1/2}} ds \quad (4.1)$$

The result (4.1) is a very nice way to express the fact that we need the infinite past to predict the future. However, practically speaking we cannot do that. Here, we have two choices: we could try to estimate increments that would prevent us from having the estimate the infinite past. In this particular case, we truncated the integral to control the accuracy and prevent infinite-horizon estimations. Recalling Theorem 2.6.1, Gatheral, Jaisson and Rosenbaum use a reformulation that unfolds as follows:

$$\mathbb{E}[\log(\sigma_{t+\Delta}^2)|\mathcal{F}_t] = \frac{\cos(\pi H)}{\pi} \int_0^{+\infty} \frac{\log(\sigma_{t-\Delta u}^2)}{(1+u)u^{H+1/2}} du \quad (4.2)$$

We then can use our natural intuition with positive numbers to find  $r$  such that for a given  $\delta > 0$ , we have:

$$\int_r^{+\infty} \frac{1}{(1+u)u^{H+1/2}} du \leq \delta$$

*Remark 4.0.1.* In the original paper [29], they advice to use  $r = 1$  so that  $\delta \approx 0.35$ . However, when trying to estimate the integral, we found that  $r = 10$  let us use  $\delta = 0.35$ . Finally, they judge 0.35 as being reasonable but we think it might be necessary to say that such error is reasonable given the scale of the volatility they are trying to predict. Indeed, we observe large change in accuracy of this method when the log-variance was large in the absolute sense. To prevent this, we always subtracted the mean to the log-volatility to center it better around zero. Also, we used  $r = 40$  as an attempt not to rely on the volatility scaling (see discussion of results below on why it matters).

We are going to consider three different prediction horizons: 1, 5 and 20 days. We will compare our model with three other models: AR(5), AR(10) and Corsi's HAR(1,5,20) model [20]. It is worth noting that, as written by Gatheral et al., GARCH model is not considered since Andersen et al. [3] proved that high-frequency realized volatility forecasts outperform daily returns-based forecast methods like GARCH. The AR and HAR models take the following functional form:

$$\begin{aligned} \text{AR}(p): \quad \widehat{\log(\sigma_{t+\Delta}^2)} &= K^\Delta + \sum_{i=0}^p A_i^\Delta \log(\sigma_{t-i}^2) \\ \text{HAR}(n,m,l): \quad \widehat{\log(\sigma_{t+\Delta}^2)} &= K^\Delta + A_n^\Delta RM(\log(\sigma_t^2), n) + A_m^\Delta RM(\log(\sigma_t^2), m) \\ &\quad + A_l^\Delta RM(\log(\sigma_t^2), l) \end{aligned}$$

where  $RM(\log(\sigma_t^2), m)$ ,  $m \in \mathbb{N}$  are the  $m$ -rolling mean from  $t$  backwards:

$$RM(\log(\sigma_t^2), m) := \frac{1}{m} \sum_{i=0}^{m-1} \log(\sigma_{t-i}^2)$$

To estimate the coefficients of the AR models, we used 500-day rolling windows along with the **stats** R package [58] and the Yule-Walker numerical procedure. For the HAR model, we wrote our own fitting method based on Corsi's least squares estimation in [20] for each forecast horizon. Also, the rolling means take intermediate predicted values as true to compute the next steps. Finally, for the RFSV predictions, we used a linear interpolation between variance estimates to approximate the integral via its Riemann sum. We also split this approximation into two chunks found empirically. The first one approximates the singularity at zero by having its first point at  $x = 2^{-17}$  and then  $2^{17}$  points between  $2^{-11}$  and  $2^{-4}$ . After that, we used  $2^{11}$  for every unit of length.

This procedure could be improved using the works of Bennedsen et al. [10] or simply using interpolation schemes for fBM given in the original paper by Mandelbrot and Van Ness [44]. This said, we wanted to compare with the results in [29] and that they just mention a simple Riemann-sum approximation, we did not want to add additional features.

*Remark.* The HAR(n,m,l) model attempts to capture three different structures: (n,m,l) are usually taken to represent a day, a week and a month, hence (1,5,20). In our experiments, it has shown to be very powerful to adapt to seasonality without having the fit external regressors. In addition, we chose  $p = 5$  and  $p = 10$  because they were respectively global and local AIC minima for SPY data (as they chose in [29]).

We take the comparing framework proposed in [29] based evaluating  $P$  defined as:

$$P := \frac{\sum_{k=500}^{N-\Delta} \left( \log(\sigma_{t+\Delta}^2) - \widehat{\log(\sigma_{t+\Delta}^2)} \right)^2}{\sum_{k=500}^{N-\Delta} \left( \log(\sigma_{t+\Delta}^2) - \mathbb{E}[\log(\sigma_{t+\Delta}^2)] \right)^2}$$

where  $\mathbb{E}[\log(\sigma_{t+\Delta}^2)]$  is the empirical mean on the entire dataset. This is quick way to estimate the *remaining/unexplained* variance in the tested model.

In the original paper [29], they demonstrate the predicting power of their RFSV model with very successful cases where RFSV forecasts outperform almost every time the three other models. Indeed, in the Live Data Library, for the originally tested `SPX2.rv`, `FTSE2.rv`, `N2252.rv`, `GDAXI2.rv` and `FCHI2.rv` datasets, RFSV is the best predicting model. However, when testing on 15 assets of this library, on both realized variance and realized kernels (on three horizons leading to 90 individual tests), we found out that the battle between HAR and RFSV models is more even. 44% of the cases were dominated by the HAR model. Equally, the RFSV model was the best performing model in 44% of the cases. This leaves only 2% of the cases to the AR models (one for AR(5), one for AR(10)). Also, RFSV forecasting has a slight advantage when dealing with realized variance against HAR as opposed to realized kernels. However, we could not spot any difference concerning the impact of the forecast horizon  $\Delta$ . This said, the best performances of the HAR model are for  $\Delta = 5$  which is the intermediate structure it tries to capture (as an (1, 5, 20) model).

As shown in Table D.7, it is also possible to find data (here prepared using HEAVY volatility model) where RFSV performs consistently better than AR and HAR models - although the difference is quite small. It is worth noticing that one of the types of assets for which RFSV performs better are assets that do not showcase seasonality or fast decay in the acf (like BTCUSD).

Finally, on our new dataset containing other assets classes (see Table D.6), the HAR model outperforms our RFSV model by far. This partially due to the construction of the HAR model: with the rolling means, it is easier to capture a sense of *seasonality*, particularly as weekly movement are directly represented by the second rolling mean ( $m = 5$ ). Also, its simple 3-element structure limits over-fitting as well as presenting the three time horizons we are interested in.

We conclude that method used to estimate volatility is key to picking the right forecasting model. Additionally, having data centred around zero seems to be central in obtaining the best performances from the RFSV model and could lead to further investigations.

It is also worth noticing that the HAR model is simple and seems to benefit from flexibility regarding the data used. However, using from 10 to 40 datapoints only (via  $r$ ), our RFSV predictions performs as well as the HAR model that required more than 500 points to be fitted! Note that increasing even further  $r$  did not improve significantly the predictions. Also, for both the AR and the HAR models, coefficients should be changed anytime we would like to predict at another time horizon  $\Delta$  or with the arrival on some fresh data which both require a new regression. On the other hand, the RFSV model just requires us to move forward the rolling window in the integrand for the Riemann-sum approximation.

## Chapter 5

# Conclusion

In this report, we have presented Gatheral, Jaisson and Rosebaum's RFSV model including the proofs of its distributional properties as well as their *Hurst exponent* estimation procedure. We proposed some reformulated arguments in which we proved a new convergence in probability of the moment estimator (Appendix B, Section B.6) that was assumed in the original paper [29] and a corrected argument for the proof of Proposition 2.4.1 which previously relied on an incomplete statement for the case  $H < 1/2$ .

Additionally, we extended their numerical experiments, primarily based on equity, to other classes : Foreign Exchange rates, Fixed Income and Commodities. By comparing market and simulated data, we have noticed that the RFSV model may or may not capture entirely the autocovariance structure independently of the asset classes: either in terms of speed of decay in the autocovariance function (ACF) or handling of seasonality. For example, RFSV seems indicated for BTCUSD when looking graphically at the autocovariance function (Figure 3.3) however simulated data shows that the ACF does not decay fast enough to replicate market data. On the other hand, EURUSD presents signs of seasonality but it still very well described by the RFSV model. We explain this from the fact that the RFSV model is essentially governed by one parameter, the Hurst exponent  $H$ . Therefore, when considering two assets with two estimated  $H$  values fairly close to each other, the model will have the same behaviour for both assets although they might be very different intrinsic properties (e.g. XAUUSD showcases seasonality, U.S. 2-year bond has an almost constant ACF of  $1/2$  but both have Hurst parameters around  $\approx 0.05$ ). We also noticed that their Hurst parameter estimation procedure was sensitive to the variance estimation method and scaling - essentially, data quality and consistency. Indeed, a jump-robust variance estimator like the minRV estimator (see[4]) does not necessarily filter out all the insignificant roughness in the volatility path compared to the standard Realized variance estimator on a smoother path. One other aspect of this exposure to scaling is forecasting as opposed to usual stastical models like AR models or the very efficient HAR model ([20]) which are not sensitive to this. However, we concluded that, although RFSV forecasting needs rescaling to be as competitive, it performs well. Indeed, we centered the log-variance around zero by subtracting by its mean. We also tried to maintain the same order of magnitude across all log-variance datasets. All in all, we obtained good results as in [29] but not universally as convincing as stated in the original paper. More precisely, the fight between HAR and

RFSV models is closer than they thought: we tested on more than 20 assets, using when available both Realized Variance and Kernels (for OMI datasets), over three different horizons (1, 5 and 20 days) and found that both models were indistinguishable.

Our final words are that the RFSV model is simple, quite flexible and proposes good results on a selection of assets which does not show sign of seasonality in volatility for example. Extending their model using external regressors for example might be useful to generalize it to such seasonal assets. Their estimation procedure for  $H$  is convincing by its simplicity and well-known behaviour regarding the underlying volatility estimation. It is well-suited to study statistical properties of the assets but for further use of Hurst exponent in pricing for example (Jacquier et al. [38]) or asymptotic expansions (Bayer et al. [8]), one might be interested in having a better statistical understanding of their procedure to know precisely its limits. This said, it is still an open question whether or not such results can be obtained.

# Appendix A

## Mandelbrot and Van Ness representation of fBM

Suppose that  $W^H$  verifies Definition 2.3.1. We will assume that  $W^H$  is well-defined first and provide an argument in Lemma A.1.8. The purpose of this part is to show that the definition from Mandelbrot and Van Ness yields to same properties as in Definition 2.2.1.

### A.1 General results using the representation

Mandelbrot and Van Ness made the insightful observation that the Hurst exponent must be in  $[0; 1[$  by definition. Indeed, the increments of the reduced fBM verify the *self-similarity* property (see Definition 2.3.2). We will go over the details of this self-similarity feature for the specific case of fBM in Lemma A.1.3.

**Lemma A.1.1.**  $W^H$  is  $H$ -self-similar.

*Proof.* Let  $t \in \mathbb{R}, a > 0, \omega \in \Omega$ . Using the change of variable  $s \mapsto au$

$$\begin{aligned} W^H(at) &\stackrel{d}{=} \frac{1}{\Gamma(H+1/2)} a^{H-1/2} \left\{ \int_{-\infty}^0 [(t-u)^{H-1/2} - (-u)^{H-1/2}] dW(au, \omega) \right. \\ &\quad \left. + \int_0^t (t-u)^{H-1/2} dW(au, \omega) \right\} \\ &\stackrel{d}{=} \frac{a^{H-1/2} \sqrt{a}}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^0 [(t-u)^{H-1/2} - (-u)^{H-1/2}] dW(u, \omega) \right. \\ &\quad \left. + \int_0^t (t-u)^{H-1/2} dW(u, \omega) \right\} \quad \text{using } dW(au, -) \stackrel{d}{=} \sqrt{a} dW(u, -) \\ &\stackrel{d}{=} a^H W^H(t) \end{aligned}$$

□

*Remark A.1.2.* Using  $H$ -self-similarity, observe that for any  $\omega \in \Omega, t \geq 0, q \geq 0$ , we have the *pointwise* equality:

$$W^H(t, \omega) = t^H W^H(1, \omega)$$

Following [44], Theorem 3.3:

**Lemma A.1.3.** *The increments of  $W^H$  are stationary.*

*Proof.* Without loss of generality, let  $t_2 > t_1 \geq 0$ . Then, from Definition 2.3.1 and stationarity of standard Brownian Motion increments:

$$\begin{aligned} W^H(t_2, \omega) - W^H(t_1, \omega) &= \frac{1}{\Gamma(H + 1/2)} \left[ \int_{-\infty}^{t_1} (t_2 - s)^{H-1/2} - (t_1 - s)^{H-1/2} dW(s, \omega) \right. \\ &\quad \left. + \int_{t_1}^{t_2} (t_2 - s)^{H-1/2} dW(s, \omega) \right] \\ &\stackrel{d}{=} \frac{1}{\Gamma(H + 1/2)} \left[ \int_{-\infty}^0 (t_2 - t_1 - u)^{H-1/2} - (-u)^{H-1/2} dW(u, \omega) \right. \\ &\quad \left. + \int_0^{t_2-t_1} (-u)^{H-1/2} dW(u, \omega) \right] \\ &\stackrel{d}{=} W^H(t_2 - t_1, \omega) \end{aligned}$$

So,  $W^H$  has stationary increments. □

Here again, equality is in distribution. Finally, we will look into the following proposition regarding the Hurst exponent possible values directly from Definition 2.3.1 of the reduced fBM:

**Lemma A.1.4.** *If  $X(t, \omega)$  has  $H$ -self-similar and stationary increments and is mean-square continuous, then  $0 \leq H < 1$ .*

*Proof.* The intuition behind this proof is the application of Minkowski's inequality for the upper bound as well as using the mean-square continuity of  $X$  for the lower bound. Formally, for any  $\tau \geq 0$ , we have by self-similarity of increments:

$$\mathbb{E}[(X(t + \tau) - X(t))^2] = \tau^{2H} \mathbb{E}[(X(t + 1) - X(t))^2] = \tau^{2H} \mathbb{E}[X(1)^2] \quad (\text{A.1})$$

where the last step comes from stationarity of increments. Define  $V := \mathbb{E}[X(1)^2]$ . Hence, for  $t_1 \geq 0$ ,  $t_2 > 0$ , using the trivial telescopic sum and Minkowski's inequality:

$$\begin{aligned} \mathbb{E}[(X(t + t_1 + t_2) - X(t))^2]^{1/2} &\leq \mathbb{E}[(X(t + t_1 + t_2) - X(t + t_2))^2]^{1/2} \\ &\quad + \mathbb{E}[(X(t + t_1 + t_2) - X(t + t_2))^2]^{1/2}; \end{aligned}$$

Therefore, applying (A.1) for, respectively,  $\tau = t_1 + t_2$ ,  $t_1$  and  $t_2$ , we obtain:

$$V^{1/2}(t_1 + t_2)^{2H/2} \leq V^{1/2}(t_1^{2H/2} + t_2^{2H/2})$$

Under the assumption that  $X(1) \in L^2$  so that  $V$  is finite, we can conclude that:

$$[t_1 + t_2]^H \leq t_1^H + t_2^H$$

and that, by the theory of Minkowski's inequality, equality is attained (by equivalence, meaning that  $H = 1$ ) if and only if  $X$  is a trivial process since the times  $\tau, t_1, t_2$  are arbitrary.

fBM is obviously not trivial therefore  $H < 1$ . Mean-square (m.-s.) continuity gives that  $H \geq 0$ . Otherwise, using (A.1) and m.-s. continuity on the LHS:

$$0 \xleftarrow{0^+ \leftarrow \tau} \mathbb{E}[(X(t + \tau) - X(t))^2] = \tau^{2H} \mathbb{E}[X(1)^2] \xrightarrow{\tau \rightarrow 0^+} +\infty$$

which is a contradiction, implying that  $H \geq 0$ . This said the case  $H = 0$  is trivial and will not be included in our general discussion.  $\square$

*Remark A.1.5.* Regarding the application of Lemma A.1.4 to the reduced fBM, it is clear the mean-square continuity is ensured by the continuity of integrals for  $t > 0$  after using Itô's isometry and the dominated convergence theorem to have the variance of an increment shrink to zero.

In addition, as stated by Mandelbrot and Van Ness in [44], fBM verifies the assumptions of Kolmogorov's continuity theorem that ensures the existence of a continuous substitute to fBM that is continuous and a.s. equal to the fBM itself for a time  $t$  in a compact set (Proposition 4.1 of the original paper). This yields that almost all sample paths of  $t \mapsto W^H(t, \omega)$  are a.s. continuous for a time  $t$  in a compact set. Kolmogorov-Čentsov's continuity theorem was stated in Karatzas and Shreve [41] (Theorem 2.8, page 53) as follows:

**Theorem A.1.6.** *Kolmogorov-Čentsov's continuity theorem*

Suppose that a process  $(X_t)_{t \in [0; T]}$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  satisfies that there exist some positive constants  $\alpha, \beta$  and  $C$ :

$$\forall s, t \in [0; T], \quad \mathbb{E}[|X_t - X_s|^\alpha] \leq C|t - s|^{1+\beta} \quad (\text{A.2})$$

Then, there exists a continuous modification  $\tilde{X} := (\tilde{X}_t)_{t \in [0; T]}$  of  $X$  which is locally Hölder-continuous with exponent  $\gamma$  for any  $\gamma \in ]0; \beta/\alpha[$ , i.e. there exists  $\delta > 0$  such that:

$$\mathbb{P} \left[ \sup_{\substack{0 < t-s < h(\omega) \\ t, s \in [0; T], \omega \in \Omega}} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t - s|^\gamma} \leq \delta \right] = 1$$

where  $h$  is an a.s. positive random variable.

*Remark.* The result can be generalised to any compact positive set in  $(\mathbb{R}, \mathcal{T}_{std})$  where  $\mathcal{T}_{std}$  is the standard topology generated by intervals  $\{]a; b[ : a < b \in \mathbb{R}\}$ .

Therefore, we can formulate the following result:

**Lemma A.1.7.** *Almost all path  $t \mapsto W^H(t, \omega)$  are continuous.*

*Proof.* To verify that we can apply Theorem A.1.6 to fBM with  $H < 1/2$ , we can replicate the proof from Mandelbrot and Van Ness [44] (Proposition 4.1). Let  $0 < k < H$  (so that  $H/k > 1$ ) and  $s, t \in [0; T]$  (w.l.o.g.  $s < t$ ), then from H-self-similarity and stationary increments properties, we have:

$$\mathbb{E}[|W^H(t) - W^H(s)|^{1/k}] = \mathbb{E}[|t - s|^{H/k} |W^H(t - s)|^{1/k}] \leq C|t - s|^{H/k}$$

where the last step is obtained using Hölder's inequality with  $q = 2/k > 1$  and the fact that  $X = 1 * X$ . By Itô's isometry on the integral representation of  $W^H$  and since  $T < \infty$  (or from the fact that any compact set in  $(\mathbb{R}, \mathcal{T}_{std})$  is bounded), we know that  $C < \infty$ . Explicitly,  $\alpha = H/k$ ,  $\beta = H/k - 1$ . □

*Remark.* Very interestingly, Decreusefond and Üstünel (Theorem 3.1, [23]) derived that in general, the sample paths of fBM could only be a.s. Hölder continuous of order less than  $H$ .

**Lemma A.1.8.**  $W^H$  is a well-defined centered Gaussian random variable.

*Proof.* Under the assumption that the process is well-defined, it is centered as sum of integrals w.r.t Brownian Motion.

Many references in the literature (e.g. [51], [47]) take this result as given but one can notice that Mandelbrot and Van Ness themselves in their classic paper [44] do not compute the variance explicitly and leave it as an integral. It is quite technical and requires to use a reformulation of the beta function on the semi-infinite interval  $[0; +\infty[$  after using Itô's isometry on the independent increments of the Brownian Motion. However, we only need to have the variance at time 1 by self-similarity of the increments (Remark A.1.2). From Norros et al. [47], page 2, we obtain:

$$\mathbb{E}[(W_1^H)^2] = \frac{\Gamma(H + 1/2)\Gamma(2 - 2H)}{2H\Gamma(3/2 - H)} \quad (\text{A.3})$$

We take the square-root of this to normalise the variance at time 1. By H-s.s, the variance at time  $t$  is proportional to  $t^{2H}$ .

Therefore, we have that the reduced fBM has finite variance for  $T < \infty$  and is a sum of stochastic integrals with respect to standard Brownian Motion. From the fact that the space of Gaussian random variables are closed under countable sum (closed linear space), it has to be the case that  $W^H$  is a (centered) Gaussian. □

Finally, we need to find the covariance structure verified by  $W^H$  to complete its characterisation as a Gaussian process. Suppose we have normalised the variance of  $W^H$  at time 1. Therefore, for  $s, t \in \mathbb{R}$ , (w.l.o.g  $s < t$ ):

$$\mathbb{E}[W^H(t)W^H(s)] = \frac{1}{2} [\mathbb{E}[(W^H(t))^2] + \mathbb{E}[(W^H(s))^2] - \mathbb{E}[(W^H(t) - W^H(s))^2]]$$

Then, we have by H-s.s. and stationarity of increments:

$$\mathbb{E}[W^H(t)W^H(s)] = \frac{1}{2} [t^{2H} + s^{2H} - |t - s|^{2H}] \quad (\text{A.4})$$

## A.2 Equivalence between fBM and its representation

The equivalence between two stochastic processes is usually defined using the equality of finite-dimensional distributions: Let  $X := (X_t)_{t \in [0; T]}$  and  $Y := (Y_t)_{t \in [0; T]}$  be two stochastic processes. Then  $X$  is equivalent to  $Y$  if and only if, for all  $n \in \mathbb{N}^*$  and for all  $0 \leq t_1 < \dots < t_n \leq T$ :

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (Y_{t_1}, \dots, Y_{t_n})$$

*Remark A.2.1.* For a Gaussian process, it is then equivalent to look at the mean and covariance structure (as hinted by Theorem 2.1.3).

If  $W^H$  verifies the representation given in Definition 2.3.1 (reduced fBM), then we have proved that almost all its paths are continuous using Kolmogorov-Čentsov's continuity theorem (Theorem A.1.6). We also proved in Lemma A.1.8 that it is almost surely a centered Gaussian process. Therefore, we know from the continuity that the density representing  $W^H(t)$  for each  $t \in [0; T]$  is unique (up to  $\mathbb{P}$ -null sets). From Theorem 2.1.3 and (A.4), we obtain that  $W^H$  must verify Definition 2.2.1.

On the other hand, if  $W^H$  verifies Definition 2.2.1 and take another copy of a reduced fBM  $V^H$  (that is - verifying Definition 2.3.1), we know from above that both are Gaussians and a.s. continuous. They have the same mean (zero), variance at any time. Hence, from the a.s. continuity of  $V_t^H$  for every  $t$  outside a  $\mathbb{P}$ -null set  $E$ , we have

$$W_t^H \stackrel{d}{=} V_t^H$$

Using the fact that  $W^H$  and  $V^H$  have the same covariance for any  $s, t \in \mathbb{R} \setminus E$  and from the characterisation of Gaussians using their covariance structure (Theorem 2.1.3), we have that  $W^H$  is equivalent to  $V^H$  which is ensured by the uniqueness of the probability density under the continuity assumption. As indicated in Baudoin and Nualart [7], Section 5, this is the only equivalence we have between the two definitions of fBM we have seen so far.

# Appendix B

## Technical proofs

### B.1 Proof of Definition 2.2.1

#### B.1.1 Ossiander and Waymire's proof

*Proof.* Let  $n \in \mathbb{N}$ ,  $0 \leq t_1 < \dots < t_n$  and  $u \in \mathbb{R}^n$ . Consider first two special cases, namely for  $H = 0$  we have:

$$\sum_{1 \leq i, j \leq n} u_i u_j \Gamma_0(t_i, t_j) = \frac{1}{2} \left( \sum_{i=1}^n u_i \right)^2 \geq 0$$

Also, for  $H = 1$ :

$$\sum_{1 \leq i, j \leq n} u_i u_j \Gamma_1(t_i, t_j) = \sum_{1 \leq i, j \leq n} u_i u_j t_i t_j = \langle ut, ut \rangle = \|ut\|^2 \geq 0$$

where  $\langle -, - \rangle$  is the usual dot product and  $ut := (u_i t_i)_{i=1}^n$ .

To prove that the property still holds for  $0 < H < 1$ , we will rely on the proof given by Ossiander and Waymire [53], Lemma 2.1 and Theorem 2.1 in the specific case of  $d = 1$ . They define an intermediate function  $\gamma_H \in L^2(\mathbb{R})$  as follows:

$$\gamma_H(s, r) := |s - r|^{H-1/2} \text{sign}(s - r) + |r|^{H-1/2} \text{sign}(r)$$

Their argument is quite standard: using the standard inner product  $\langle -, - \rangle$  on  $L^2(\mathbb{R})$  as stated below:

$$\forall f, g \in L^2(\mathbb{R}) \quad \int_{-\infty}^{+\infty} f(r)g(r)dr$$

they show that the fractional Brownian Motion covariance kernel  $\Gamma_H$  is obtained (up to a positive multiplicative constant) by taking the inner product of  $\gamma_H(t, -)$  and  $\gamma_H(s, -)$ . We start by proving that  $t \mapsto \gamma_H(t, -) \in L^2(\mathbb{R})$ , observe that  $\gamma_H(0, r) = 0, \forall r \in \mathbb{R}$  and for  $r \neq 0$ :

$$\begin{aligned} \gamma_H^2(s, r) &= [|s - r|^{H-1/2} \text{sign}(s - r) + |r|^{H-1/2} \text{sign}(r)]^2 \\ &= |s|^{2H-1} [|1 - r/s|^{H-1/2} \text{sign}(s - r) + |r/s|^{H-1/2} \text{sign}(r)]^2 \\ &= |s|^{2H-1} [|1 - r/s|^{H-1/2} \text{sign}(1 - r/s) + |r/s|^{H-1/2} \text{sign}(r/s)]^2 \quad \text{using the square} \\ &= |s|^{2H-1} \gamma_H^2(1, r/s) \end{aligned}$$

To show that  $u \mapsto \gamma_H(1, u)$  is square integrable, we first need to show that  $u \mapsto |u|^{H-1/2} \text{sign}(u)$  and  $u \mapsto |1-u|^{H-1/2} \text{sign}(1-u)$  are, respectively, locally square integrable in the neighbourhoods of  $u = 0$  and  $u = 1$  especially since we are interested in the case  $H < 1/2$ . Let  $a, b > 0$ , we have

$$\begin{aligned} \int_{1-a}^{1+b} |1-u|^{2H-1} du &= \int_0^b |v|^{2H-1} dv - \int_0^a |v|^{2H-1} dv \\ &= \frac{1}{2H} \left[ [|x|^{2H} \text{sign}(x)]_0^b - [|x|^{2H} \text{sign}(x)]_0^a \right] < \infty \end{aligned}$$

This case verifies the properties for both functions stated above; the usual trick  $(a+b)^2 \leq 2(a^2+b^2)$  let us conclude that we have settled the square integrability of  $u \mapsto \gamma_H(1, u)$  by symmetry on  $] -2; 2[$ . We still have to prove square integrability on  $\mathbb{R} \setminus ] -2; 2[$ . Again, by symmetry, consider:

$$\begin{aligned} \int_2^{+\infty} \gamma_H^2(1, u) du &= \int_2^{+\infty} \left[ u^{H-1/2} - (1-u)^{H-1/2} \right]^2 du \\ &= \int_2^{+\infty} \left( \int_{u-1}^u \left( H - \frac{1}{2} \right) x^{H-3/2} dx \right)^2 du \\ &\leq \int_2^{+\infty} \left( \int_{u-1}^u \left( H - \frac{1}{2} \right)^2 x^{2H-3} dx \int_{u-1}^u dy \right) du \quad \text{by Hölder's inequality} \\ &\leq \left( H - \frac{1}{2} \right)^2 \left( \int_2^{+\infty} u^{2H-3} dx + \int_2^{+\infty} (u-1)^{2H-3} du \right) \\ &< \infty \quad \text{since } 2H - 3 < -2 \end{aligned}$$

So, we have for  $s \neq 0$ :

$$0 \leq \|\gamma_H^2(s, -)\|_2^2 = |s|^{2H} \|\gamma_H^2(1, -)\|_2^2 < \infty$$

Hence,  $s \mapsto \gamma_H(s, -) \in L^2(\mathbb{R})$ . We also need to find a way to express the *cross term* in the covariance kernel in terms of  $\gamma_H$ . Observe that

$$\begin{aligned} \|\gamma_H(t, -) - \gamma_H(s, -)\|_2^2 &= \int_{-\infty}^{+\infty} \left[ |t-u|^{H-1/2} \text{sign}(t-u) - |s-u|^{H-1/2} \text{sign}(s-u) \right]^2 du \\ &\stackrel{v:=s-u}{=} \int_{-\infty}^{+\infty} \left[ |t-s+v|^{H-1/2} \text{sign}(t-s+v) - |v|^{H-1/2} \text{sign}(v) \right]^2 du \\ &\stackrel{w:=-v}{=} \int_{-\infty}^{+\infty} \left[ |t-s-w|^{H-1/2} \text{sign}(t-s-w) - |w|^{H-1/2} \text{sign}(w) \right]^2 du \\ &= \|\gamma_H(t-s, -)\|_2^2 = |t-s|^{2H} \|\gamma_H(1, -)\|_2^2 \end{aligned}$$

From the bilinearity of  $\langle -, - \rangle$ , we have the usual trick:

$$\begin{aligned} \langle \gamma_H(t, -), \gamma_H(s, -) \rangle &= \frac{1}{2} \left[ \|\gamma_H(t, -)\|_2^2 + \|\gamma_H(s, -)\|_2^2 - \|\gamma_H(t-s, -)\|_2^2 \right] \\ &= \frac{\|\gamma_H(1, -)\|_2^2}{2} \left[ |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right] \end{aligned}$$

Defining our intermediary function  $\widehat{\gamma}_H$  using  $\widehat{\gamma}_H := \gamma_H / \sqrt{\|\gamma_H(1, -)\|_2^2}$  let us verify the positive semi-definiteness for our kernel: let  $0 \leq t_1 < \dots < t_n \leq T$  for  $n \in \mathbb{N}^*$  and  $(c_i)_{i=1}^n \in \mathbb{R}^n$ , we have:

$$\begin{aligned} \sum_{1 \leq i, j \leq n} c_i c_j \Gamma_H(t_i, t_j) &= \sum_{1 \leq i, j \leq n} c_i c_j \langle \widehat{\gamma}_H(t_i, -), \widehat{\gamma}_H(t_j, -) \rangle \\ &= \langle c \widehat{\gamma}_H(t, -), c \widehat{\gamma}_H(t, -) \rangle \geq 0 \end{aligned}$$

where  $c \widehat{\gamma}_H(t, -) := (c_i \widehat{\gamma}_H(t_i, -))_{i=1}^n$ .

We conclude that the process  $W^H$  is well-defined.  $\square$

### B.1.2 Gangolli's proof:

The following is based on a case study by Gangolli himself, [28], page 134, Case II on how fractional Brownian Motion can be defined on multidimensional parameter space  $\mathbb{R}^d$ . Here, we study in the 2-dimensional case to give an overview of the multidimensional parameter case. For the algebraic definitions used below, see Liao's book on Lévy processes in Lie Groups [43], Appendix A.

*Proof.* As pinpointed in [53], historically the works of Gangolli [28] found its first application to prove the positive semi-definiteness of (2.1) using so-called *Lévy-Schönberg* covariance kernels. In this context, a kernel on a topological space  $S$  is any complex valued function defined on  $S \times S$ . Gangolli defines those kernels as follows (Definition 2.3, [28]):

*Given a separable topological group  $G$  and a closed subgroup  $K$  of  $G$  and  $G/K$  denoting the space of cosets  $xK$ ,  $x \in G$  which we equip with the quotient topology, a kernel  $f$  on  $G/K$  is a Lévy-Schönberg kernel if:*

- $f(a, b) = f(b, a) \quad \forall a, b \in G/K$
- $\exists e \quad f(a, e) = 0 \quad \forall a \in G/K$
- *The characteristic kernel on  $G/K$  given by  $r(a, b) = f(a, a) + f(b, b) - 2f(a, b)$  is invariant under  $G$ , i.e.  $r(xa, xb) = r(a, b) \quad \forall x \in G, a, b \in G/K$*
- *$f$  is positive definite*

Here positive semi-definiteness of a function  $f$  is taken with coefficients in  $\mathbb{C}$ : for  $(u_i)_{i=1}^n \in \mathbb{C}$ , we want to verify that:

$$\sum_{1 \leq i, j \leq n} u_i \bar{u}_j f(u_i, u_j) \geq 0$$

In particular, we consider  $G$  the group (by composition  $\circ$ ) of all proper rigid motions of  $\mathbb{R}^2$  (i.e. rotations then a translation) and  $K$  being the subgroup of proper rotations around  $0_{\mathbb{R}^2}$  which is obviously closed by composition (i.e. the angles are just summed up).  $K$  is normal in  $G$ : the left group action  $g \mapsto g \circ K$  gives the same element for  $g$  as the right group action  $g \mapsto K \circ g$  since the rotations are around the origin  $0$ .  $\mathbb{R}^2$  acts transitively on  $G$  by

composition: indeed, for  $g := e^{i\theta}(-) + \nu \in G$  and  $y \in \mathbb{R}^2$ , define  $x := (y - \nu)e^{-i\theta} \in \mathbb{R}^2$ . Then, we can check

$$g \circ x = y$$

Therefore, from the fact that  $K \cong SO(2)$  is closed and normal in  $G$ , we have that  $(\mathbb{R}^2, \mathcal{T}_{std}, \times)$  (equivalent to the space whose topology is generated by the Euclidean norm) is characterised as a topological group by  $G/K$  equipped with the quotient topology (Gangolli, [28], page 134, Case II). Recall that the quotient topology is the smallest topology such that the canonical projection  $G \rightarrow G/K$ ,  $g \mapsto gK$  is continuous. We say that the topological space  $(\mathbb{R}^2, \mathcal{T}_{std})$  is one of the homogeneous spaces of  $G$ . This is what justifies studying  $G/K$  instead of  $\mathbb{R}^2$ .

First, we check that the kernel:

$$f(a, b) := \frac{1}{2}[\|a\|^{2H} + \|b\|^{2H} - \|a - b\|^{2H}] \quad (\text{B.1})$$

verifies the afore-mentioned definition except the last bulletpoint. Here,  $\|\cdot\|$  is the Euclidean norm in  $\mathbb{R}^2$ . The symmetry is trivial and given  $K$ , we know that our *origin*  $e$  will be the coset  $0_{\mathbb{R}^2}K$ . Note that  $r(a, b) = \|a - b\|^{2H}$  is invariant with respect to rigid motions as required. The true insight comes from Gangolli' Theorem 3.6 which we reformulate to fit our needs (see original paper [28] for full list of assumptions implicitly verified here):

*The kernel defined in (B.1) is a Lévy-Schönberg kernel if and only if, essentially,*

$$f(a, b) = \frac{1}{2}(r(a, 0) + r(b, 0) - r(a, b)), \quad a, b \in \mathbb{R}^2$$

where  $r(a, b) = \Psi(a - b)$  and if we denote the Bessel function of the first kind of order 0, namely  $J_0$ , and  $L$  a non-negative measure, we ask that  $\Psi$  verifies:

$$\Psi(a) = \int_0^{+\infty} (1 - Y_0(\lambda\|a\|))dL(\lambda) \quad \forall a \in \mathbb{R}$$

Recall that:

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m + \alpha + 1)} \left(\frac{x}{2}\right)^{2m+\alpha}$$

Gangolli makes the observation that the choice  $dL(\lambda) := d\lambda/\lambda^{1+2H}$  let us obtain our previous radial function  $r(a, b) = \|a - b\|^{2H}$ . Indeed, let  $R > 0$ , then we have by definition:

$$\begin{aligned} \int_0^R (1 - J_0(\lambda\|a\|)) \frac{d\lambda}{\lambda^{1+2H}} &= \int_0^R \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m!\Gamma(m + 1)} \left(\frac{\lambda\|a\|}{2}\right)^{2m} \frac{d\lambda}{\lambda^{1+2H}} \\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m!\Gamma(m + 1)} \frac{\|a\|^{2m}}{2^{2m}} \int_0^R \left(\frac{\lambda\|a\|}{2}\right)^{2m} \frac{d\lambda}{\lambda^{1+2H}} \\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m!\Gamma(m + 1)} \frac{\|a\|^{2m}}{2^{2m}} \int_0^R \lambda^{2m-2H-1} d\lambda \\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m!\Gamma(m + 1)} \frac{\|a\|^{2m}}{2^{2m}} \frac{R^{2m-2H}}{2m-2H} \\ &= \|a\|^{2H} 2^{2H} \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m!\Gamma(m + 1)} \frac{1}{2m-2H} \left(\frac{R\|a\|}{2}\right)^{2m-2H} \end{aligned}$$

So, there exists  $M^-, M^+ > 0$  such that:

$$\|a\|^{2H}(1 - M^- J_{-2H}(R\|a\|)) \leq \int_0^R (1 - J_0(\lambda\|a\|)) \frac{d\lambda}{\lambda^{1+2H}} \leq \|a\|^{2H}(1 - M^+ J_{-2H}(R\|a\|))$$

Recall that  $J_{-2H}(x) \xrightarrow{x \rightarrow +\infty} 0$ , so taking  $R \rightarrow +\infty$ , we obtain:

$$\Psi(a) = \|a\|^{2H}$$

which concludes the proof.  $\square$

## B.2 Proof of Equation (2.2)

*Proof.* The proof of (2.2) resides in a careful application of the fBM covariance kernel. Let  $s > 1$ ,  $H \in ]0; 1[$ . First, observe that

$$\frac{(1+s)^{2H} - s^{2H}}{s^{2H}} = \left(1 + \frac{1}{s}\right)^{2H} - 1 = \frac{2H}{s} + \frac{H(2H-1)}{s^2} + o\left(\frac{1}{s^2}\right)$$

and:

$$\frac{(s-1)^{2H} - s^{2H}}{s^{2H}} = \left(1 - \frac{1}{s}\right)^{2H} - 1 = -\frac{2H}{s} + \frac{H(2H-1)}{s^2} + o\left(\frac{1}{s^2}\right)$$

Hence, we obtain:

$$\begin{aligned} \mathbb{E}[W^H(1)(W^H(1+s) - W^H(s))] &= \frac{1}{2} [1 + (1+s)^{2H} - s^{2H} - 1 - s^{2H} + (s-1)^{2H}] \\ &= \frac{1}{2} [(1+s)^{2H} - s^{2H} + (s-1)^{2H} - s^{2H}] \\ &= H(2H-1)s^{2H-2} + o(1) \end{aligned}$$

which concludes the proof.  $\square$

## B.3 Proof of Proposition 2.4.1

*Proof.* Lemma 2.4.2 will be crucial in the proof of (2.11). Indeed, we will separate the integral into two parts: one for the history of fBM before time 0 and one from time 0 to time  $t$ . Using (2.10) and *integration by parts* (as in (2.5)) on the integral:

$$X_t^\alpha = \nu \left[ W_t^H - W_0^H + \int_{-\infty}^t -\alpha e^{-\alpha(t-s)} W_s^H ds \right] + m$$

After subtracting  $X_0^\alpha$  and since  $W_0^H = 0$ :

$$X_t^\alpha - X_0^\alpha - \nu W_t^H = - \int_0^t \alpha \nu e^{-\alpha(t-s)} W_s^H ds + \int_{-\infty}^0 \alpha \nu (e^{\alpha s} - e^{-\alpha(t-s)}) W_s^H ds$$

Applying the triangle inequality on both the sum and the integrals and by Lemma A.1.7, we have:

$$|X_t^\alpha - X_0^\alpha - \nu W_t^H| \leq \int_0^t \alpha \nu e^{-\alpha(t-s)} |W_s^H| ds + \int_{-\infty}^0 \alpha \nu (e^{\alpha s} - e^{-\alpha(t-s)}) |W_s^H| ds$$

Hence, from the usual rules concerning suprema of a.e. continuity of  $t \mapsto W_t^H(\omega)$  and integrals, we obtain:

$$\sup_{t \in [0; T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H| \leq \alpha \nu T \sup_{t \in [0; T]} |W_t^H| + \int_{-\infty}^0 \alpha \nu (e^{\alpha s} - e^{-\alpha(t-s)}) |W_s^H| ds$$

For the first term, we use Lemma 2.4.2 with  $p = 1$  directly. For the second one, observe that  $s \mapsto \alpha \nu (e^{\alpha s} - e^{-\alpha(t-s)}) |W_s^H|$  is non-negative: we can apply Fubini's theorem to interchange the integral and the expectation. Finally, by Taylor's expansion on the exponential and the fact that  $W^H$  is Gaussian:

$$\mathbb{E} \left[ \sup_{t \in [0; T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H| \right] \leq C(1, H) \nu \alpha T T^H + \nu \alpha \sqrt{\frac{2}{\pi}} \int_{-\infty}^0 e^{\alpha s} (T \alpha) |s|^H ds$$

Hence, there exists a constant  $C$  (that could depend on  $H$ ) such that:

$$\mathbb{E} \left[ \sup_{t \in [0; T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H| \right] \leq C \left( \alpha T^H + \alpha^2 \int_{-\infty}^0 e^{\alpha s} |s|^H ds \right)$$

By a simple change of variable and from the strength of convergence of the exponential compared to linear functions, we have that:

$$\int_{-\infty}^0 e^{\alpha s} |s|^H ds = \frac{1}{\alpha^{1+H}} \underbrace{\int_{-\infty}^0 e^u |u|^H du}_{< \infty}$$

Since  $0 < H < 1$ , we have:

$$\frac{\alpha^2}{\alpha^{1+H}} \int_{-\infty}^0 e^u |u|^H du \xrightarrow{\alpha \rightarrow 0} 0$$

Therefore, by taking the limit  $\alpha \rightarrow 0$ :

$$\mathbb{E} \left[ \sup_{t \in [0; T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H| \right] \xrightarrow{\alpha \rightarrow 0} 0$$

□

## B.4 Proof of Lemma 2.4.2

*Remark.* This proof in [49] is valid only for  $1/2 < H < 1$ . The only bottleneck for it to be extended to the case  $0 < H < 1/2$  is fact that fBM has infinite quadratic variation for  $H < 1/2$  (see [57]). We propose an extension using Jost [39] that formalise changes in Hurst parameter, namely here changing from  $H$  to  $1 - H$  for  $1 \leq p \leq 2$ . Note that the case  $p \geq 2$  was proved in the more general context of fractional Lévy processes by Bender et al. [9].

*Proof.* The proof makes use of so-called Volterra-like representations of fBM: for a more general theory, see [7] with the special case of fBM treated in Section 5.

Explicitly, the argument presented here is a reformulation to fit our needs from Novikov and Valkeila, [49]. Let  $Y_t^H = \int_0^t s^{1/2-H} dW_s^H$ . Using integration by parts (Remark 2.3.4, (2.5)), we find that

$$W_t^H = \int_0^t s^{H-1/2} dY_s^H \quad (\text{B.2})$$

Let  $M^H$  be the so-called *fundamental integral* defined by

$$M_t^H := \int_0^t w(t, s) dW_s^H$$

where  $w(t, s)$  is a scaled beta kernel such that  $w(t, s) := c(H) s^{-\alpha} (t-s)^{-\alpha}$  for  $s \in ]0; t[$  and  $w(t, s) := 0$  for  $s > t$  with  $\alpha := H - 1/2$ ,  $c(H)$  is a function of  $H$  and the standard Beta function, see [49]. In addition, combining these definitions yields:

$$M_t^H = c(H) \int_0^t (t-s)^{1/2-H} dY_s^H$$

We are interested in finding an integral representation of  $Y^H$  with respect to  $M^H$  as we did for  $W^H$  with  $Y^H$ . In Norros et al. [47], they provided a formal proof (Theorem 3.4) that  $M^H$  is a Gaussian (hence continuous) martingale for  $H \in ]0; 1[$  with respect to the filtration generated by  $W^H$ . This is very surprising since  $W^H$  is not even a semi-martingale! Their argument relies on a reformulation of the fBM covariance kernel as given by  $(t, s) \mapsto w(t, s)$ .

They also showed that the filtrations generated by  $M^H$  and  $Y^H$  are the same - up to null sets - hence the name *fundamental* for  $M^H$ . It would be interesting to cover it but unfortunately it is out of the scope of this report and we are going to assume this result. In the same theorem, they use these facts to infer the following:

$$\mathbb{E}[Y_T | \mathcal{F}_t^Y] = \mathbb{E}[Y_T | \mathcal{F}_t^M] = 2H \int_0^t (T-s)^{H-1/2} dM_s^H$$

Since  $T$  is arbitrary, this implies that for any  $t \geq 0$ :

$$Y_t^H = \mathbb{E}[Y_t^H | \mathcal{F}_t^Y] = \mathbb{E}[Y_t^H | \mathcal{F}_t^M] = 2H \int_0^t (t-s)^{H-1/2} dM_s^H \quad (\text{B.3})$$

Finally, we can use (B.2) and (B.3) to find the upper bound by integration by parts (as in (2.5))

$$\begin{aligned} \sup |W_t^H| &\leq T^{H-1/2} \sup(|Y_t^H| - |Y_0^H|) && \text{by (B.2)} \\ &\leq 2T^{H-1/2} \sup(|Y_t^H|) \\ &\leq 4HT^{2H-1} \sup(|M_t^H| - |M_0^H|) && \text{by (B.3)} \\ &\leq 8HT^{2H-1} \sup(|M_t^H|) \end{aligned}$$

This leads to:

$$\mathbb{E}[(\sup |W_t^H|)^p] = (8HT^{2H-1})^p \mathbb{E}[(\sup(|M_t^H|))^p]$$

So, recall that  $M$  is a continuous martingale. We obtain from the celebrated Burkholder-Davis-Gundy (BDG) inequality applied on  $M^H$  for  $p \geq 1$  that there exists a constant  $A_p^H > 0$  such that:

$$\mathbb{E}[(\sup |W_t^H|)^p] \leq (8HT^{2H-1})^p A_p^H \mathbb{E}[(\langle M^H \rangle_t)^{p/2}] \quad (\text{B.4})$$

Then, observe that by definition of  $M^H$  and  $W^H$ , we can compute the quadratic variation of  $M^H$ . First, let  $K \in [1/2; 1]$  and  $W^K$  be a  $K$ -fBM. Recall the covariance kernel used in Section B.1.1 (Ossiander and Waymire's proof) which can be adapted in this situation as detailed in Proposition 2.1, [47]. Formally, since  $M^K$  is a continuous martingale with  $M_0^K = 0$ , their computation yields:

$$\text{Var}(M_t^K) = \langle M^K \rangle_t = \frac{c^2(K)}{(2K)^2(2-2K)} t^{2-2K} =: d(K)t^{2-2K}$$

where  $d(K)$  has an obvious definition. This result was also given in [49], Proposition 2.1. We will rely on a marvellous theory inspired from the Molchan-Golosov-type integral and proposed by Jost [39] that, essentially, let us *change* the Hurst parameter using a specified transformation. Given our  $H$ -fBM  $W^H$ , by Corollary 5.2, [39], we have that there exists a unique  $(1-H)$ -fBM,  $W^{1-H}$ , such that:

$$W_t^H = \left( \frac{2H}{\Gamma(2H)\Gamma(3-2H)} \right)^{1/2} \int_0^t (t-s)^{2H-1} dW_s^{1-H} \quad a.s., \quad t \in [0; T]$$

Given that, Jost herself gives the result for the fundamental integral  $M^H$  of  $W^H$  in Remark 5.6, [39]:

$$M_t^H = \sqrt{\frac{1-H}{H}} \int_0^t s^{1-2H} dM_s^{1-H}$$

Then, apply the previously computed quadratic variation for  $1-H > 1/2$ , we obtain that:

$$\begin{aligned} \langle M^H \rangle_t &= \frac{1-H}{H} \int_0^t s^{2-4H} d\langle M^{1-H} \rangle_s \\ &= \frac{1-H}{H} d(1-H) \int_0^t s^{2-4H} d(s^{2H}) \\ &= \frac{2H(1-H)}{H} d(1-H) \int_0^t s^{1-2H} ds \\ &= d(1-H)t^{2(1-H)} \end{aligned}$$

By applying a second time the BDG inequality, now on  $\langle M^H \rangle$ , it yields a constant  $B_p^{1-H}$  that depends on  $d(1-H)$  s.t.:

$$\mathbb{E} \left[ \sqrt{\langle M^H \rangle_t}^p \right] \leq B_p^{1-H} t^{p(1-H)}$$

We have using (B.4):

$$\mathbb{E}[(\sup |W_t^H|)^p] \leq (8H)^p A_p^H B_p^{1-H} T^{p(2H-1)} T^{p(1-H)}$$

Define  $C(p, H) := (8H)^p A_p^H B_p^{1-H}$ . Finally, we conclude:

$$\mathbb{E}[(\sup |W_t^H|)^p] \leq C(p, H) T^{pH}$$

□

*Remark B.4.1.* The Burkholder-Davis-Gundy inequality was originally developed for  $p > 1$  by Burkholder [14] and the case  $p = 1$  is due to the work of Davis [21].

## B.5 Proof of Proposition 2.4.4

*Proof.* Pipiras and Taqqu [56], page 289, defined an inner product space  $\Lambda_H$  for  $H < 1/2$  on  $\mathbb{R}$  as follows:

$$\Lambda_H := \{f : \mathbb{R} \rightarrow \mathbb{R} : \int_{\mathbb{R}} [G_H f(s)]^2 ds < \infty\}$$

where  $G_H f(s) := \frac{1}{\Gamma(H)} \int_{\mathbb{R}} f(u)(u-s)^{H-1} I_{\{u \geq s\}} ds$  and equipped with the inner product (with  $c(H)$  a constant that depends solely on H):

$$(f, g)_{\Lambda_H} := \frac{\Gamma(H+1)^2}{c(H)^2} \int_{\mathbb{R}} (G_H f)(s)(G_H g)(s) ds$$

Cheridito et al. [16], page 8, made the observation that  $f(x) := I_{\{x \leq 0\}} e^{\lambda x}$  and  $g(x) := I_{\{x \leq s\}} e^{\lambda x}$  for any  $s \in \mathbb{R}$  belong to  $\Lambda_H$ . Given this fact and (2.10) we have that for any  $s, t \in \mathbb{R}$

$$Cov(X_{t+s}^\alpha, X_t^\alpha) = \nu^2 e^{-\alpha s} (f, g)_{\Lambda_H} = \nu^2 \frac{\Gamma(2H+1) \sin(\pi H)}{2\pi} \int_{\mathbb{R}} e^{isx} \frac{|x|^{1-2H}}{\alpha^2 + x^2} dx$$

Finally, we take the specific case of  $s = \Delta$  to have the result that:

$$Cov(X_{t+\Delta}^\alpha, X_t^\alpha) = K \int_{\mathbb{R}} e^{i\Delta x} \frac{|x|^{1-2H}}{\alpha^2 + x^2} dx$$

where  $K := \nu^2 \frac{\Gamma(2H+1) \sin(\pi H)}{2\pi}$ . By stationarity, we have the following reformulation for the second moment of the increment:

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] = 2Var(X_t^\alpha) - 2Cov(X_{t+\Delta}^\alpha, X_t^\alpha) = 2K \int_{\mathbb{R}} (1 - e^{i\Delta x}) \frac{|x|^{1-2H}}{\alpha^2 + x^2} dx$$

We can find a uniform bound in  $\alpha$  by taking  $\alpha \rightarrow 0$ :

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] \leq 2K \int_{\mathbb{R}} (1 - e^{i\Delta x}) \frac{|x|^{1-2H}}{x^2} dx$$

The left-hand side is real because it is a sum of a real integral and the Fourier transform of an even function. This bound is crucial since we are dealing with Gaussian random variables

which are characterised by their mean and but more importantly their (co-)variance. We want to prove uniform integrability for all moments of the family of increments  $(X_{t+\Delta}^\alpha - X_t^\alpha)_{\alpha \in \mathbb{R}_+}$ . We achieve this by using Hölder's inequality on 1 and the increment itself. Recalling that  $Y_t^{\Delta, \alpha} := X_{t+\Delta}^\alpha - X_t^\alpha$  is a centered Gaussian, we can then go back to the studied case  $q = 2$  from the property that, for any even  $n$ , the  $n$ -th Gaussian moment is proportional to the second moment to the power of  $n/2$ . Let  $q > 0$  and define  $h$  to be the smallest even integer strictly greater than  $q$ . Since  $T < \infty$ , there exists a constant  $C(T) > 0$  such that:

$$\begin{aligned} \mathbb{E}[|X_{t+\Delta}^\alpha - X_t^\alpha|^q] &\leq C(T)\mathbb{E}[(|Y_t^{\Delta, \alpha}|^q)^{\frac{h}{q}}]^{\frac{q}{h}} \quad \text{by Hölder's inequality} \\ &\leq C(T)\mathbb{E}[(|Y_t^{\Delta, \alpha}|^h)^{\frac{q}{h}}] \end{aligned}$$

So, using the properties of Gaussians, we have that there exists  $K > 0$  such that:

$$\mathbb{E}[|X_{t+\Delta}^\alpha - X_t^\alpha|^q] \leq C(T)K\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2]^{\frac{q}{2}}$$

Using the above-mentioned uniform bound in  $\alpha$  for the second moment and the fact that finite-variance Gaussians belong to  $L^2$ , we can uniformly bound any moment of the increments which gives that the family (in  $\alpha$ )  $(|X_{t+\Delta}^\alpha - X_t^\alpha|^q)_{\alpha \geq 0}$  is uniformly integrable for a fixed  $t$  and a fixed  $q$ . Given that Gaussian variables are in  $L^1$  and using Proposition 2.4.1 yielding the convergence in probability for any  $t$ ,  $X_t^\alpha - X_0^\alpha \xrightarrow[\alpha \rightarrow 0]{p} \nu W_t^H$ , we conclude.  $\square$

## B.6 Proof of the application of Proposition 2.5.6

*Remark.* We prove the result when  $\alpha \rightarrow 0$  but the case  $\alpha > 0$  relies on the computation of variance for the fOU process as given in [29] and would be done in the *exact* same fashion as presented here.

*Proof.* Gut [32] uses the so-called *uniform integrability in the Cesàro sense* (Definition (f), Section 2, page 49, [32]) defined as follows:

Let  $\{X_{nk}, 1 \leq k \leq v_n, n \geq 1\}$  be an array of random variables. This array is said to be

- *uniformly integrable if*

$$\mathbb{E}[|X_{ni}|I_{|X_{ni}|>a}] \xrightarrow{a \rightarrow +\infty} 0 \quad \text{uniformly in } n \text{ and in } i$$

- *uniformly integrable in the Cesàro sense if*

$$\frac{1}{v_n} \sum_{k=1}^{v_n} \mathbb{E}[|X_{ni}|I_{|X_{ni}|>a}] \xrightarrow{a \rightarrow +\infty} 0 \quad \text{uniformly in } n$$

The main result we are going to use ([32], page 50) unfolds as follows:

**Theorem B.6.1.** *Gut's theorem, Section 3, [32]*

Let  $\{X_{nk}, 1 \leq k \leq v_n, n \geq 1\}$  be an array of non-negative random variables.

Let  $\mathcal{F}_{nk} := \sigma\{X_{ni}, 1 \leq i \leq k\}$ , for  $1 \leq k \leq v_n, n \geq 1$  and  $\mathcal{F}_{n0} = \{\emptyset, \Omega\}, n \geq 1$ .

Let  $1 \leq p < 2$  and suppose that the array  $\{(|X_{ni}|^p, 1 \leq i \leq v_n), n \geq 1\}$  is uniformly integrable in the Cesáro sense. Set

$$S_n^X := \begin{cases} \sum_{k=1}^{v_n} X_{nk}, & 0 < p < 1 \\ \sum_{k=1}^{v_n} (X_{nk} - \mathbb{E}[X_{nk} | \mathcal{F}_{n,i-1}]), & 1 \leq p < 2 \end{cases}$$

Then:

$$v_n^{-1/p} S_n^X \xrightarrow{L^p} 0 \quad \text{as } n \rightarrow +\infty$$

Therefore, we also have convergence in probability.

Let  $\Delta_n = T/n$  assuming  $T$  is a multiple of  $n$ . We add the additional observation: from the proof of Proposition 2.4.4 (Section B.5), we have that the family of increments  $(|X_{t+\Delta_n}^\alpha - X_t^\alpha|^q)_{\alpha \geq 0}$  is uniformly integrable for a given  $t$  and  $q$ .

Define  $v_n := n$  and define the triangular array constructed as follows:

$$Y_{nk}^\alpha := n^{qH} |X_{k\Delta_n}^\alpha - X_{(k-1)\Delta_n}^\alpha|^q, \quad 1 \leq k \leq n$$

Using Proposition 2.4.1 and its result in terms of convergence in law, we will set  $\nu := 1$  for simplify the notation. More precisely, define

$$Y_{nk}^0 := n^{qH} |W_{k\Delta_n}^H - W_{(k-1)\Delta_n}^H|^q$$

By stationarity of increments and H-self-similarity, we obtain:

$$Y_{nk}^0 \stackrel{d}{=} n^{qH} \Delta_n^{qH} |W_1^H|^q \stackrel{d}{=} T^{qH} |W_1^H|^q$$

So, our array  $Y^0$  is made of identically distributed yet dependent variables. This observation can be extended to  $Y^\alpha$  as well. Therefore, the convergence to 0 of

$$\mathbb{E} \left[ |Y_{ni}^\alpha| I_{|Y_{ni}^\alpha| > a} \right] \quad \text{as } a \rightarrow +\infty$$

is independent  $n$  and  $i$ . This concludes that both  $Y^0$  and  $Y^\alpha$  are uniformly integrable as arrays. Hence, again by identical distribution of the entries of the two arrays, we obtain that  $\{Y_{ni}^0\}_{n \in \mathbb{N}, 1 \leq i \leq n}$  and  $\{Y_{ni}^\alpha\}_{n \in \mathbb{N}, 1 \leq i \leq n}$  are uniformly integrable in the Cesáro sense.

*Remark.* We could then carry on with  $\alpha > 0$  but we are going to prove for  $\alpha \rightarrow 0$  using self-similarity of fBM.

We can apply the convergence in distribution (given in the proof of Proposition 2.4.4) to obtain :

$$N^{-1} S_N^{Y^\alpha} \xrightarrow{d} N^{-1} S_N^{Y^0} \quad \text{as } \alpha \rightarrow 0$$

In addition, for  $\Delta := T/N$ , applying the Gut's theorem for  $p = 1$  on  $Y^0$  yields:

$$N^{-1} S_N^{Y^0} = \frac{N^{qH}}{N} \sum_{k=1}^N |W_{k\Delta}^H - W_{(k-1)\Delta}^H|^q \xrightarrow{p} \frac{N^{qH}}{N} \sum_{k=1}^N \mathbb{E} \left[ |W_{k\Delta}^H - W_{(k-1)\Delta}^H|^q | \mathcal{F}_{n,i-1} \right]$$

as  $N \rightarrow +\infty$ . By direct computation (or see WinkelBauer, [66]), for  $Z \sim N(0, \sigma_Z^2)$  and  $q > -1$ :

$$\mathbb{E}[|Z|^q] = \sigma_Z^q \frac{2^{q/2}}{\sqrt{\pi}} \Gamma\left(\frac{q+1}{2}\right)$$

In particular, for  $Z := \mathbb{E}\left[W_{k\Delta}^H - W_{(k-1)\Delta}^H \mid \mathcal{F}_{N,k-1}\right]$ , we have  $\sigma_Z = O(\Delta^H) = O\left(\frac{T^H}{N^H}\right)$  (standard Landau  $O(-)$  notation) with  $T < \infty$ . This proves that:

$$\zeta := \frac{N^{qH}}{N} \sum_{k=1}^N \mathbb{E}\left[|W_{k\Delta}^H - W_{(k-1)\Delta}^H|^q \mid \mathcal{F}_{N,k-1}\right]$$

is such that

$$|\zeta| < \infty \quad a.s.$$

Therefore, we conclude that:

$$\frac{N^{qH}}{N} \sum_{k=1}^N |X_{k\Delta}^\alpha - X_{(k-1)\Delta}^\alpha|^q \xrightarrow{p} \zeta \quad \text{as } \alpha \rightarrow 0, N \rightarrow +\infty$$

with  $\zeta$  a.s. bounded which the result we were seeking. □

## Appendix C

# UZ model for high-frequency data

As two pioneers in the research field of microstructure noise, Robert and Rosenbaum developed a model for high-frequency data which comprises many features observed in the market: the model with uncertainty zones (UZ). For completeness, please refer to their original paper [59]. We shall expose herein the essential points we are going to use in our study.

They start from the idea that non-speculative market participants are reluctant to price changes as it may entail a reaction or adjustment on their behalves (e.g. hedging). This implies that, from the discreteness of market prices, we should not consider that the true market price is the closest value on the price grid  $\{k\alpha, k \in \mathbb{N}\}$ . Indeed, a naive idea would be that, given a tick of size  $\alpha$  and an - unobserved - efficient price  $P$ , the market price should be given by  $k\alpha$ ,  $k \in \mathbb{N}$ , where  $k$  is such that  $|P - k\alpha|$  is minimal.

They assume that the price changes whenever market participants have sufficient proof or belief that, despite the uncertainty surrounding the price itself from the last traded price, a change in price is necessary. In some sense, they require the efficient price to be *significantly* closer to the next price, as they write. To model this "significance" formally, they introduce a parameter  $\eta$  (which might be time-dependent) which define the width (relative to the tick  $\alpha$ ) of the uncertainty bands around the price. In this framework, a price change occurs whenever the price not only exit a band but also reaches the *next* one. By this, we mean that the price should enter the lower band of the tick price  $(k+1)\alpha$  from below or enter the upper band of  $(k-1)\alpha$  from above. Essentially, this implies a no-man's land of size  $2\eta\alpha$  around the mid-tick where we decide that we cannot know if the price has to move and, from the reluctance of market participants, we say the price does not move.

More generally, they also define stopping times that models the time at which the efficient price moves sufficiently to imply a market move. In addition, they also model the fact that depending on market conditions and/or price range we are situated at, uncertainty bands might be thicker or thinner by using a stochastic jump size parameter or with a variable parameter  $\eta$ . As they do in [59], we shall consider that a trade happens at every timestep (therefore a potential price change occurs at every timestep, which we should check) and that all *up* and *down* bands have the same width of size:

$$\alpha(1/2 + \eta)$$

In our simulation study, we are given a sample path  $X$  that we take to be the efficient - theoretical - price which is unobserved on real markets. Indeed, in practice we only see the observed price from the last trade. To generate this observed path, we are going to check at each timestep if the necessary price (to induce a move in the observed price) was achieved . This will generate a path  $P$  with changes in price at times  $\tau_i$  which are taken to be the same as the trade times  $t_i$  for simplicity. Following [59], we then define our realized volatility estimator as stated below: at time  $t$ , we have:

$$\widehat{RV}_t := \sum_{\tau_i \leq t} \left( \frac{\widehat{X}_{\tau_i} - \widehat{X}_{\tau_{i-1}}}{\widehat{X}_{\tau_{i-1}}} \right)^2$$

where  $\widehat{X}_{\tau_i} := P_{\tau_i} - \text{sign}(P_{\tau_i} - P_{\tau_{i-1}})(1/2 - \eta)\alpha$ .

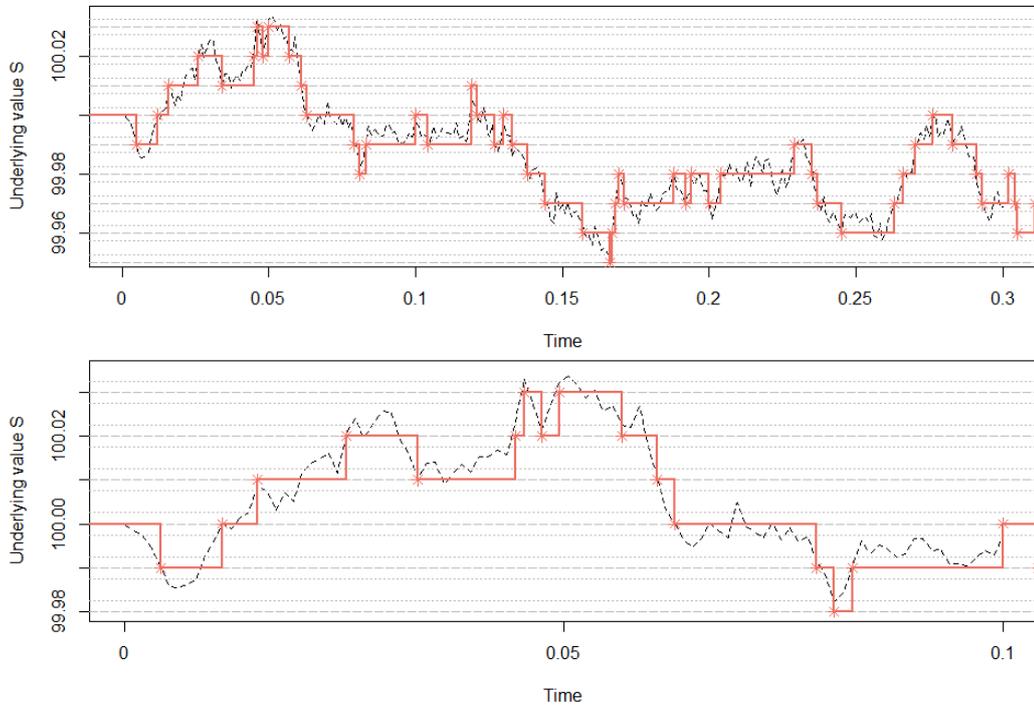


Figure C.1: Example of generation of observed prices (in orange) from theoretical prices (dashed black line) through the UZ model with  $\eta = 0.25$ ,  $\alpha = 0.01$ . Price process is a GBM with  $S_0 = 100$ ,  $\sigma = 0.2$ ,  $r = 0$ ,  $T = 1$ .

# Appendix D

## Real market dataset

### D.1 $\xi$ -estimation procedure

Table D.1:  $\xi$ -estimation procedure on Realized Variance from Live Data Library

	% NA	$\xi_{0.5/0.5}$	$\xi_1/1$	$\xi_{1.5/1.5}$	$\xi_2/2$	$\xi_3/3$
<b>SPX2.rv</b>	4.40	0.138	0.138	0.138	0.138	0.137
<b>FTSE2.rv</b>	3.90	0.134	0.131	0.129	0.127	0.123
<b>N2252.rv</b>	7.30	0.137	0.136	0.135	0.133	0.126
<b>GDAXI2.rv</b>	3.10	0.129	0.129	0.128	0.127	0.126
<b>DJI2.rv</b>	4.30	0.126	0.126	0.125	0.125	0.124
<b>FCHI2.rv</b>	2.50	0.135	0.134	0.133	0.132	0.130
<b>HSI2.rv</b>	11.60	0.082	0.083	0.084	0.084	0.084
<b>KS11.rv</b>	5.80	0.128	0.127	0.127	0.126	0.122
<b>AEX.rv</b>	2.60	0.136	0.139	0.141	0.142	0.142
<b>SSMI.rv</b>	4.20	0.149	0.153	0.157	0.159	0.159
<b>IBEX2.rv</b>	3.30	0.121	0.123	0.123	0.123	0.121
<b>NSEI.rv</b>	17.10	0.116	0.113	0.110	0.107	0.100
<b>MXX.rv</b>	4.30	0.081	0.081	0.081	0.081	0.080
<b>BVSP.rv</b>	6.40	0.114	0.115	0.116	0.116	0.115
<b>STOXX50E.rv</b>	3.10	0.127	0.124	0.119	0.111	0.084

Table D.2:  $\xi$ -estimation procedure on Realized Kernel from Live Data Library

	% NA	$\xi_{0.5}/0.5$	$\xi_1/1$	$\xi_{1.5}/1.5$	$\xi_2/2$	$\xi_3/3$
<b>SPX2.rk</b>	4.40	0.149	0.149	0.150	0.151	0.151
<b>FTSE2.rk</b>	3.90	0.148	0.146	0.144	0.142	0.137
<b>N2252.rk</b>	7.30	0.136	0.135	0.134	0.134	0.129
<b>GDAXI2.rk</b>	3.10	0.142	0.139	0.137	0.136	0.136
<b>DJI2.rk</b>	4.30	0.146	0.147	0.148	0.148	0.147
<b>FCHI2.rk</b>	2.50	0.127	0.127	0.127	0.126	0.124
<b>HSI2.rk</b>	11.60	0.097	0.096	0.095	0.095	0.092
<b>KS11.rk</b>	5.80	0.132	0.132	0.132	0.131	0.126
<b>AEX.rk</b>	2.60	0.136	0.139	0.140	0.141	0.140
<b>SSMI.rk</b>	4.20	0.167	0.170	0.173	0.175	0.174
<b>IBEX2.rk</b>	3.30	0.127	0.127	0.127	0.127	0.124
<b>NSEI.rk</b>	17.10	0.147	0.141	0.136	0.130	0.117
<b>MXX.rk</b>	4.30	0.091	0.092	0.090	0.088	0.082
<b>BVSP.rk</b>	6.40	0.109	0.112	0.114	0.114	0.112
<b>STOXX50E.rk</b>	3.10	0.128	0.125	0.120	0.112	0.081

Table D.3:  $\xi$ -estimation procedure on other asset classes

	$\xi_{0.5}/0.5$	$\xi_1/1$	$\xi_{1.5}/1.5$	$\xi_2/2$	$\xi_3/3$
<b>BTC/USD.rv</b>	0.159	0.150	0.139	0.125	0.091
<b>EUR/USD.minrv</b>	0.041	0.037	0.033	0.028	0.021
<b>GOLD/USD.minrv</b>	0.059	0.054	0.050	0.044	0.034
<b>WTI/USD.minrv</b>	0.090	0.085	0.080	0.074	0.062
<b>USB02Y.minrv</b>	0.054	0.050	0.042	0.035	0.025
<b>USB05Y.minrv</b>	0.056	0.050	0.047	0.045	0.036
<b>USB10Y.minrv</b>	0.052	0.047	0.045	0.042	0.028

Table D.4:  $\xi$ -estimation procedure on Realized Variance from Realized Library

	% NA	$\xi_{0.5}/0.5$	$\xi_1/1$	$\xi_{1.5}/1.5$	$\xi_2/2$	$\xi_3/3$
DJI_rv	17.1	0.135	0.135	0.133	0.131	0.125
FCHI_rv	16.1	0.114	0.115	0.116	0.117	0.118
FTSE_rv	27.7	0.115	0.116	0.116	0.114	0.109
IBEX_rv	16.9	0.111	0.113	0.116	0.117	0.117
IXIC_rv	17.1	0.113	0.113	0.113	0.112	0.110
MIBTEL_rv	44.7	0.136	0.136	0.137	0.137	0.138
MID_rv	17.2	0.123	0.121	0.121	0.120	0.118
N225_rv	19.7	0.109	0.111	0.113	0.114	0.115
RUA_rv	17.1	0.133	0.130	0.128	0.125	0.120
RUI_rv	17.1	0.133	0.131	0.128	0.125	0.120
RUT_rv	17.0	0.121	0.117	0.114	0.114	0.113
MIB30_rv	16.4	0.136	0.135	0.135	0.136	0.136
GDAXI_rv	16.2	0.118	0.117	0.116	0.116	0.115
SPTSE_rv	35.7	0.115	0.114	0.114	0.115	0.119
SPX_rv	17.0	0.136	0.133	0.130	0.127	0.121
MSCIAU_rv	41.2	0.100	0.101	0.101	0.102	0.105
MSCIBE_rv	38.1	0.123	0.128	0.133	0.138	0.144
MSCIBR_rv	59.9	0.147	0.144	0.141	0.140	0.140
MSCICA_rv	49.1	0.111	0.114	0.117	0.120	0.126
MSCICH_rv	38.3	0.145	0.149	0.153	0.157	0.160
MSCIDE_rv	37.9	0.126	0.127	0.128	0.130	0.131
MSCIES_rv	38.6	0.113	0.116	0.117	0.117	0.115
MSCIFR_rv	37.8	0.126	0.128	0.128	0.129	0.129
MSCIGB_rv	37.9	0.132	0.135	0.136	0.136	0.131
MSCIIT_rv	38.2	0.133	0.134	0.134	0.135	0.136
MSCIJP_rv	43.3	0.128	0.130	0.131	0.133	0.138
MSCIKR_rv	42.7	0.143	0.142	0.141	0.142	0.142
MSCIMX_rv	59.3	0.118	0.121	0.124	0.128	0.138
MSCINL_rv	37.8	0.131	0.136	0.140	0.143	0.146
MSCIWO_rv	46.8	0.133	0.121	0.106	0.091	0.069

Table D.5:  $\xi$ -estimation procedure on Realized Kernel from Realized Library

	% NA	$\xi_{0.5}/0.5$	$\xi_1/1$	$\xi_{1.5}/1.5$	$\xi_2/2$	$\xi_3/3$
DJI_rk	17.1	0.143	0.142	0.14	0.137	0.131
FCHI_rk	16.1	0.106	0.105	0.106	0.107	0.108
FTSE_rk	27.7	0.112	0.112	0.112	0.111	0.108
IBEX_rk	16.9	0.110	0.114	0.116	0.117	0.116
IXIC_rk	17.1	0.111	0.109	0.107	0.105	0.102
MIBTEL_rk	44.7	0.120	0.121	0.122	0.123	0.126
MID_rk	17.2	0.123	0.121	0.119	0.118	0.116
N225_rk	19.7	0.104	0.105	0.107	0.108	0.111
RUA_rk	17.1	0.133	0.131	0.128	0.125	0.120
RUI_rk	17.1	0.133	0.131	0.128	0.125	0.120
RUT_rk	17.0	0.120	0.117	0.115	0.115	0.113
MIB30_rk	16.4	0.119	0.122	0.123	0.125	0.128
GDAXI_rk	16.2	0.115	0.114	0.114	0.115	0.115
SPTSE_rk	35.7	0.114	0.112	0.111	0.112	0.114
SPX_rk	17.0	0.135	0.131	0.128	0.125	0.119
MSCIAU_rk	41.2	0.087	0.090	0.092	0.093	0.097
MSCIBE_rk	38.1	0.102	0.107	0.111	0.115	0.122
MSCIBR_rk	59.9	0.137	0.133	0.131	0.128	0.125
MSCICA_rk	49.1	0.099	0.101	0.103	0.106	0.111
MSCICH_rk	38.3	0.123	0.129	0.134	0.139	0.144
MSCIDE_rk	37.9	0.103	0.107	0.109	0.111	0.115
MSCIES_rk	38.6	0.095	0.098	0.100	0.100	0.098
MSCIFR_rk	37.8	0.107	0.108	0.108	0.108	0.107
MSCIGB_rk	37.9	0.114	0.117	0.118	0.119	0.120
MSCIIT_rk	38.2	0.112	0.115	0.117	0.120	0.123
MSCIJP_rk	43.3	0.130	0.129	0.130	0.130	0.133
MSCIKR_rk	42.7	0.127	0.129	0.130	0.131	0.131
MSCIMX_rk	59.3	0.114	0.113	0.113	0.114	0.117
MSCINL_rk	37.8	0.111	0.116	0.120	0.124	0.130
MSCIWO_rk	46.8	0.111	0.102	0.092	0.081	0.064

## D.2 Forecasting volatility: tables

Table D.6: Predicting log-variance at horizons 1, 5 and 20 for other asset classes

	AR(5)	AR(10)	HAR(1,5,20)	RFSV
EURUSD $\Delta = 1$	0.522	0.505	<b>0.496</b>	0.521
EURUSD $\Delta = 5$	0.578	0.562	<b>0.537</b>	0.574
EURUSD $\Delta = 20$	0.623	0.619	<b>0.610</b>	0.680
BTCUSD $\Delta = 1$	0.528	0.523	0.521	<b>0.520</b>
BTCUSD $\Delta = 5$	0.799	0.782	<b>0.734</b>	0.776
BTCUSD $\Delta = 20$	1.05	1.044	<b>0.918</b>	1.003
XAUUSD $\Delta = 1$	0.549	0.551	<b>0.526</b>	0.533
XAUUSD $\Delta = 5$	0.684	0.669	<b>0.630</b>	0.680
XAUUSD $\Delta = 20$	0.762	0.741	<b>0.721</b>	0.794
WTIUSD $\Delta = 1$	0.333	0.334	<b>0.319</b>	0.323
WTIUSD $\Delta = 5$	0.446	0.439	<b>0.415</b>	0.437
WTIUSD $\Delta = 20$	0.538	<b>0.528</b>	0.534	0.568
USB02Y $\Delta = 1$	0.282	0.279	<b>0.267</b>	0.274
USB02Y $\Delta = 5$	0.317	0.311	<b>0.308</b>	0.325
USB02Y $\Delta = 20$	0.364	<b>0.354</b>	0.368	0.360
USB05Y $\Delta = 1$	0.658	0.654	<b>0.644</b>	0.662
USB05Y $\Delta = 5$	0.746	0.737	<b>0.722</b>	0.775
USB05Y $\Delta = 20$	0.835	0.829	<b>0.818</b>	0.887
USB10Y $\Delta = 1$	0.499	0.497	<b>0.493</b>	0.507
USB10Y $\Delta = 5$	0.640	0.651	<b>0.612</b>	0.627
USB10Y $\Delta = 20$	<b>0.718</b>	<b>0.718</b>	0.720	0.756

Table D.7: Predicting log-variance at horizons 1, 5 and 20 for SPY and BAC:

	AR(5)	AR(10)	HAR	RFSV
BAC $\Delta = 1$	0.118	0.119	0.117	<b>0.115</b>
BAC $\Delta = 5$	<b>0.190</b>	0.190	0.191	<b>0.190</b>
BAC $\Delta = 10$	0.307	0.305	0.332	<b>0.301</b>
SPY $\Delta = 1$	0.218	0.218	<b>0.214</b>	<b>0.214</b>
SPY $\Delta = 5$	0.342	0.342	<b>0.341</b>	0.346
SPY $\Delta = 10$	0.532	0.532	0.543	<b>0.530</b>

Table D.8: OMI Live Library data: Percentage of cases where each of the models had the smallest P value.

	AR(5)	AR(10)	HAR	RFSV	TOTAL
$\Delta = 1$	0%	0%	33.3%	66.6%	100%
$\Delta = 5$	3.33%	3.33%	73.3%	20%	100%
$\Delta = 20$	0%	0%	40%	60%	100%
<b>Overall</b>	1%	1%	44%	44%	100%

# References

- [1] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (1999). The distribution of stock return volatility. *Preprint*, pages 1–48.
- [2] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55.
- [3] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.
- [4] Andersen, T. G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 169(1):75–93.
- [5] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.
- [6] Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied econometrics*, 17(5):457–477.
- [7] Baudoin, F. and Nualart, D. (2003). Equivalence of volterra processes. *Stochastic processes and their applications*, 107(2):327–350.
- [8] Bayer, C., Friz, P. K., Gulisashvili, A., Horvath, B., and Stemper, B. (2017). Short-time near-the-money skew in rough fractional volatility models. *arXiv preprint arXiv:1703.05132*.
- [9] Bender, C., Knobloch, R., and Oberacker, P. (2015). Maximal inequalities for fractional lévy and related processes. *Stochastic Analysis and Applications*, 33(4):701–714.
- [10] Bennedsen, M., Lunde, A., and Pakkanen, M. S. (2015). Hybrid scheme for brownian semistationary processes. *arXiv preprint arXiv:1507.03004*.
- [11] BitCoinCharts (2017). Btc-e data from bitcoin chart api, based on btc-e bitcoin trading activity (<https://btc-e.com/>) in us dollars. <http://api.bitcoincharts.com/v1/csv/btceUSD.csv.gz> Last accessed May 18 2017.

- [12] Boudt, K., Cornelissen, J., and Payseur, S. (2017). *highfrequency: Tools for Highfrequency Data Analysis*. R package version 0.5.1 <https://CRAN.R-project.org/package=highfrequency>.
- [13] Brouste, A., Fukasawa, M., Hino, H., Iacus, S. M., Kamatani, K., Koike, Y., Masuda, H., Nomura, R., Ogihara, T., Shimuzu, Y., Uchida, M., and Yoshida, N. (2014). The yuima project: A computational framework for simulation and inference of stochastic differential equations. *Journal of Statistical Software*, 57(4):1–51. <http://www.jstatsoft.org/v57/i04/>.
- [14] Burkholder, D. L. (1966). Martingale transforms. *The Annals of Mathematical Statistics*, 37(6):1494–1504.
- [15] Cabrera, M. O. and Volodin, A. I. (2005). Mean convergence theorems and weak laws of large numbers for weighted sums of random variables under a condition of weighted integrability. *Journal of Mathematical Analysis and Applications*, 305(2):644–658.
- [16] Cheridito, P., Kawaguchi, H., Maejima, M., et al. (2003). Fractional ornstein-uhlenbeck processes. *Electron. J. Probab*, 8(3):1–14.
- [17] Cohen, A. (2000). Wavelet methods in numerical analysis. *Handbook of numerical analysis*, 7:417–711.
- [18] Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323.
- [19] Corporation, O. (2017). Oanda rest-v20 api for financial data. <http://developer.oanda.com/rest-live/introduction/>.
- [20] Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, page nbp001.
- [21] Davis, B. (1970). On the integrability of the martingale square function. *Israel Journal of Mathematics*, 8(2):187–190.
- [22] Debicki, K. and Tomanek, A. (2009). Estimates for moments of supremum of reflected fractional brownian motion. *arXiv preprint arXiv:0912.3117*.
- [23] Decreusefond, L. et al. (1999). Stochastic analysis of the fractional brownian motion. *Potential analysis*, 10(2):177–214.
- [24] Dieker, T. (2004). Simulation of fractional brownian motion. *MSc theses, University of Twente, Amsterdam, The Netherlands*.
- [25] Ding, Z., Granger, C. W., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106.
- [26] Fouque, J.-P., Papanicolaou, G., and Sircar, K. R. (2000). Mean-reverting stochastic volatility. *International Journal of theoretical and applied finance*, 3(01):101–142.

- [27] Gallager, R. G. (2013). *Stochastic processes: theory for applications*. Cambridge University Press.
- [28] Gangolli, R. (1967). Positive definite kernels on homogeneous spaces and certain stochastic processes related to lévy's brownian motion of several parameters. *Annales de l'I.H.P. Probabilités et statistiques*, 3(2):121–226.
- [29] Gatheral, J., Jaisson, T., and Rosenbaum, M. (2014). Volatility is rough. *arXiv preprint arXiv:1410.3394*.
- [30] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238.
- [31] Gushchin, A. A. and Kùchler, U. (2005). On recovery of a measure from its symmetrization. *Theory of Probability & Its Applications*, 49(2):323–333.
- [32] Gut, A. (1992). The weak law of large numbers for arrays. *Statistics & probability letters*, 14(1):49–52.
- [33] Hand, D. J. (2010). Theory of stochastic processes: With applications to financial mathematics and risk theory by dmytro gusak, alexander kukush, alexey kulik, yuliya mishura, andrey pilipenko. *International Statistical Review*, 78(3):461–461.
- [34] Hu, Y., Nualart, D., Song, J., et al. (2009). Fractional martingales and characterization of the fractional brownian motion. *The Annals of Probability*, 37(6):2404–2430.
- [35] Hu, Y., Nualart, D., Xu, F., et al. (2014). Central limit theorem for an additive functional of the fractional brownian motion. *The Annals of Probability*, 42(1):168–203.
- [36] Huang, J. (2013). *somebm: some Brownian motions simulation functions*. R package version 0.1, <https://CRAN.R-project.org/package=somebm>.
- [37] Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808.
- [38] Jacquier, A., Martini, C., and Muguruza, A. (2017). On vix futures in the rough bergomi model.
- [39] Jost, C. (2006). Transformation formulas for fractional brownian motion. *Stochastic Processes and their Applications*, 116(10):1341–1357.
- [40] Kaarakka, T. and Salminen, P. (2007). On fractional ornstein-uhlenbeck processes. *arXiv preprint arXiv:0710.5024*.
- [41] Karatzas, I. and Shreve, S. (2012). *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media.
- [42] Lamperti, J. (1962). Semi-stable stochastic processes. *Transactions of the American mathematical Society*, 104(1):62–78.

- [43] Liao, M. (2004). *Lévy processes in Lie groups*, volume 162. Cambridge university press.
- [44] Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.
- [45] McGonegal, B. (2014). Fractional brownian motion and the fractional stochastic calculus. *arXiv preprint arXiv:1401.2752*.
- [46] McLeod, A. I. and Hipel, K. W. (1978). Preservation of the rescaled adjusted range: 1. a reassessment of the hurst phenomenon. *Water Resources Research*, 14(3):491–508.
- [47] Norros, I., Valkeila, E., Virtamo, J., et al. (1999). An elementary approach to a girsanov formula and other analytical results on fractional brownian motions. *Bernoulli*, 5(4):571–587.
- [48] Noureldin, D., Shephard, N., and Sheppard, K. (2012). Multivariate high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 27(6):907–933.
- [49] Novikov, A. and Valkeila, E. (1999). On some maximal inequalities for fractional brownian motions. *Statistics & Probability Letters*, 44(1):47–54.
- [50] Nualart, D., Xu, F., et al. (2013). Central limit theorem for an additive functional of the fractional brownian motion ii. *Electronic Communications in Probability*, 18:1–10.
- [51] Nuzman, C. J. and Poor, H. V. (2000). Linear estimation of self-similar processes via lamperti’s transformation. *Journal of Applied Probability*, 37(02):429–452.
- [52] original by Chris Fraley, S., U. Washington, port by Fritz Leisch at TU Wien; since 2003-12: Martin Maechler; fdGPH, S. R., fdSperio, etc by Valderio Reisen, and Lemonte., A. (2012). *fracdiff: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*. R package version 1.4-2 <https://CRAN.R-project.org/package=fracdiff>.
- [53] Ossiander, M. and Waymire, E. C. (1989). Certain positive-definite kernels. *Proceedings of the American Mathematical Society*, 107(2):487–492.
- [54] Oxford-Man Institute, o. Q. F. (2017). Realized library data download, last accessed on may 17 2017. <http://realized.oxford-man.ox.ac.uk/data/download>.
- [55] Percival, D. B. and Walden, A. T. (2006). *Wavelet methods for time series analysis*, volume 4. Cambridge university press.
- [56] Pipiras, V. and Taqqu, M. S. (2000). Integration questions related to fractional brownian motion. *Probability theory and related fields*, 118(2):251–291.
- [57] Pratelli, M. (2011). A remark on the  $1/h$ -variation of the fractional brownian motion. In *Séminaire de Probabilités XLIII*, pages 215–219. Springer.
- [58] R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

- [59] Robert, C. Y. and Rosenbaum, M. (2011). A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366.
- [60] Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.*, 23(3):1048–1072.
- [61] Rosenbaum, M. (2009). First order p-variations and besov spaces. *Statistics & Probability Letters*, 79(1):55–62.
- [62] Rosenbaum, M. (2011). A new microstructure noise index. *Quantitative Finance*, 11(6):883–899.
- [63] Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231.
- [64] Spodarev, E., Shmileva, E., and Roth, S. (2015). Extrapolation of stationary random fields. In *Stochastic Geometry, Spatial Statistics and Random Fields*, pages 321–368. Springer.
- [65] Vardar, C. (2009). Results on the supremum of fractional brownian motion. *arXiv preprint arXiv:0910.5193*.
- [66] Winkelbauer, A. (2012). Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*.
- [67] Wood, A. T. and Chan, G. (1994). Simulation of stationary gaussian processes in  $[0, 1]$  d. *Journal of computational and graphical statistics*, 3(4):409–432.
- [68] Wuertz, D., many others, and see the SOURCE file (2013). *fArma: ARMA Time Series Modelling*. R package version 3010.79 <https://CRAN.R-project.org/package=fArma>.