Selection of sparse multifractional model

Pierre, Raphael Bertrand, Marie-Eliette Dury, Nabiha Haouas

Selection of sparse multifractional model

Pierre Raphaël BERTRAND*,† , Marie-Eliette DURY‡,
and Nabiha HAOUAS†,§

September 4, 2015

Abstract

The aim of this paper is to provide a simple model with a time-varying Hurst index. Such models should be both the simplest possible and fit well the real Hurst index. Moreover this would avoid a numerical artefact pointed out in this article. For this, after a recall on fBm, mBm and statistical estimation of the Hurst index, including a time-varying one, we propose a fitting test for a model with a time-varying Hurst index. Then an approach is given to select the most simple model.

Keywords: fractional Brownian motion (fBm), multifractional Brownian motion (mBm), Hurst index, Model selection, Sparse model, Big data, Portemanteau test, Quantitative finance, Efficient Market Hypothesis, Behavioural finance.

Introduction

The most famous centered Gaussian process is the Brownian motion. One of its generalisations is the fractional Brownian motion (fBm) introduced in 1940 by Kolmogorov [23] as *"Gaussian spirals in Hilbert space"* and popularised since 1968 by Mandelbrot and Van Ness. The fBm is the unique H-self-similar Gaussian process with stationary increments up to a multiplicative constant, where $0 < H < 1$ denotes the Hurst index. Case $H = 1/2$ corresponds to the Brownian motion. The multifractional Brownian motion (mBm) is hence defined from the fBm but with a time-varying Hurst index, which can be encountered in many different kinds of applications:

- In turbulence, Papanicolaou and Sølna (2002) denote that *"the power law itself* [i.e. the Hurst index, ...] *and the multiplicative constant are not constants but vary slowly"* in [30], whereas Lee (2003) uses mBm with a regularly time-varying Hurst index for the air velocity, see [25, Fig. 5, p. 103].

*This research is partially supported by grant ANR-12-BS01-0016-01 entitled *"Do Well B"*. The third author acknowledges also support from Grant CNRS/DGRS 10/R15-01 entitled *"Statistique des processus de type fractal"*.

†Laboratoire de Mathématiques, UMR CNRS 6620 and Université Clermont-Ferrand 2, France.

‡School of Economics, UDA, Université Clermont-Ferrand 1, France.

§Computational Mathematics Laboratory UR11ES51, Sousse University, Tunisia

- In a statistical study on magnetospheric dynamics, Wanliss and Dobias (2007) point out that an abrupt change in Hurst index can be observed a few hours before a space storm in solar wind [35].
- In systems biology, [28] uses mBm to simulate molecular crowding that matches the statistical properties of sample data, whereas Lim and Teo (2009) use mBm with piecewise constant Hurst index to model single file diffusion that is the motion of chemical, physical or biological particules in quasi-one-dimensional channel [27].

On the other hand, quantitative finance has been the most important field of application of both time series and stochastic processes for the last fifty years. Actually, fBm was revived during the 1960's by Mandelbrot to also serve as a model for speculative prices, as we read in its posthumous autobiography [26]. However, the efficient market hypothesis has lead to reject fBm as an admissible model for stock price. So, since the 1970's, the use of martingale models has become mainstream in quantitative finance. Unfortunately enough, from time to time, financial crises highlight the fact that the martingale model is just a good approximation for financial assets, but presents some drawbacks mainly during such crises. Each crisis reinforces the investigation of new or alternative models [31]. But, the main objection to fBm, as an admissible model for stock prices, is the existence of an arbitrage opportunity for such a fBm with a constant and known Hurst index. To put it into a nutshell, an arbitrage opportunity means the possibility of producing a positive return from zero investment by clever trading. For a fBm with known and constant Hurst index, it is possible to make an arbitrage, with a strategy based on infinitely small meshes of times and without transaction cost [32, 34, 17], which turn to be quite nonrealistic conditions. Moreover this objection is not applicable for generalisations of fBm that allows a Hurst index varying with time or frequency, see e.g. [8, 9, 4, 10] and the references therein.

An economic complementary point of view is developed in [11, 13, 14, 15, 20]: Firstly, by analysing different financial time series (Standard & Poor's 500 between 1982 and 2002, and Japanese Nikkei Index N225 between 1984 and 2004) [11, Fig.9 and Fig.10, p.275] Bianchi (2005) pointed out that the Hurst index estimated on sliding windows is varying with time between 0.45 and 0.65. Then Bianchi and Pianese (2008) [12, Fig.6 p. 583 and Fig.7, p.584] checked that the same empirical evidence is verified for the US Dow Jones Index (daily observed from 1928 to 2004) and for the UK FTSE 100 Index (daily observed from 1984 to 2005) with a Hurst index $H(t)$ varying between 0.3 and 0.6 in both cases.

Theoretical explanations are then developed by economists Bianchi, Pianese, Pantanella and Frezza [13, 14, 15, 20]. To sum up, arbitrage opportunities for fBm are possible when the Hurst index is constant and known in advance, but not when it is time-varying and random. Moreover, periods with a Hurst index that significantly differs from 1/2 can be explained by behavioural economics. For periods where $H(t) < 1/2$ the market overreacts, which means in probabilistic term, that the increments of the (log)price process are negatively correlated (antipersistence), whereas for periods where $H(t) > 1/2$ the market underreacts, which means that the increments of the (log)price process are positively correlated (persistence). In behavioural finance, underreaction is due to overconfidence of investitors, see e.g. [15, Table 1, p. 13]. Recall that the case $H = 1/2$ corresponds to independence of the increments and to efficiency of the market.

The next logical step is to assume that the Hurst index is itself a stochastic process, that is to say with irregular paths as the multifractional process with random exponent (MPRE) [2]. However, we will show in this paper that this choice is counterproductive as this complex model contains a statistical artefact. On the contrary, we here look for a model as simple as possible with a time-varying Hurst index, which can still be random. By using a fitting test we describe a way of model selection.

In the rest of the paper our plan will be the following: In a first section, we set the framework and explain the underlying ideas. Next in a second section, we recall the definition and main properties of fBm, mBm and statistical estimation of the Hurst index. Then in a third section we present the fitting test, we apply it to reject a stochastic Hurst index, then we provide the application to select the simplest model with a time-varying Hurst index. All technical proofs are postponed in appendices.

1 Framework and motivation of our study

In this paper, we aim at giving a method for the selection of a good probabilistic model with a time-varying Hurst index. So, to begin with, we provide an overview of the context and the process we work with. This process is the so-called multifractional Brownian motion (mBm) that can be viewed as a generalisation of the fractional Brownian motion (fBm). Let $\{B_H(t),\ t \in \mathbb{R}\}$ be a fBm with Hurst index $H \in (0, 1]$. It can be defined as follows:

Definition 1.1 *The fBm $\{B_H(t),\ t \in \mathbb{R}\}$ is the zero mean Gaussian process which covariance function is*

$$cov\big[B_H(s),\, B_H(t)\big] \quad = \quad \frac{\sigma^2}{2}\left\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right\} \quad \text{for all} \quad (s,t) \in \mathbb{R}^2. \tag{1}$$

The fBm has stationary increments, it admits different representations and can be also viewed as a Gaussian field depending both on the time t and the Hurst index H. For instance, the harmonisable representation of the fBm considered as a Gaussian field is given by

$$B(t, H) \quad = \quad \int_{\mathbb{R}} \frac{(e^{it\xi} - 1)}{|\xi|^{H+1/2}}\, dW(\xi), \qquad \text{for all} \quad t \in \mathbb{R}, \tag{2}$$

where $(t, H) \mapsto B(t, H) := B_H(t)$ is the fBm, and W is a complex valued Wiener measure such that $B(t, H)$ is real valued. As said above the mBm can be defined as a generalisation of the fBm where the Hurst index H is replaced by a time-varying function $H(t)$. The mBm is hence defined by

$$X(t) \quad = \quad B\big(t, H(t)\big), \qquad \text{for all} \quad t \in \mathbb{R}.$$

The use of such a process is motivated by statistical studies, which have shown that a model with constant Hurst index such as the fBm does not fit real life applications [6, 24]. Indeed fBm is a very rich model where a unique parameter, namely the Hurst index H, drives many properties: the correlation structure of the increments, the long range dependency, the self-similarity, and the roughness of the paths. Due to the time-varying Hurst index, stationarity of the increments does not hold anymore, therefore both long range dependency and structure of the increments are meaningless notions. For the mBm, only the roughness of the paths corresponds to the Hurst index $H(t)$. However, this property is satisfied under an extra-condition insuring that the Hölder regularity of the time-varying Hurst index $t \mapsto H(t)$ is greater than the maximum value of $H(t)$, see e.g. [2, 6]. In order to allow very general probabilistic models, new generalisations of fBm or mBm have been introduced with a Hurst index which can be very irregular and even be itself a stochastic process, namely multifractional process with random exponent (MPRE) or generalised multifractional process (GMP) [2, 3].

Actually, we cannot know whether fluctuations reflect reality or are just artefacts byproducts of statistics. This phenomenon is brought to light by the estimation of a time-varying Hurst index for a process X being a fBm with a constant Hurst index $H = 0.7$. Indeed Fig. 1 gives the

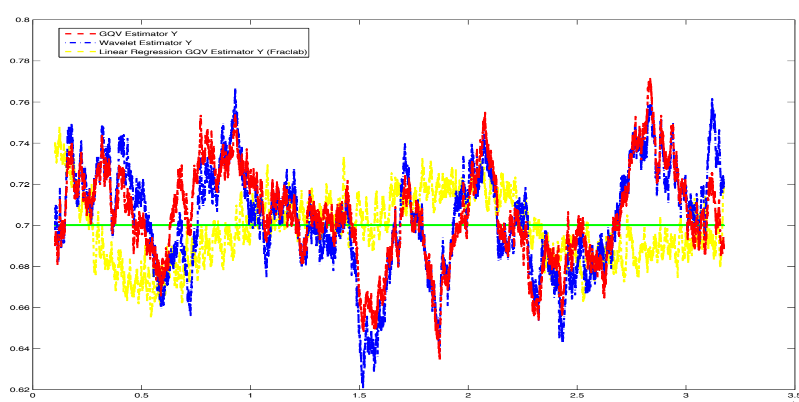

Figure 1: Estimation of a time-varying Hurst index $\widehat{H}(t)$ for a fBm with constant Hurst index $H = 0.7$.

feeling that the Hurst index is itself a stochastic process. In fact, the theoretical Hurst index is constant. But if we assume that this theoretical Hurst index is a time-varying function, namely $t \mapsto H(t)$, then at each time t the Hurst index is estimated on a small vicinity around the time t. Consequently, the sampling fluctuation induces that the time-varying estimator $\widehat{H}(t)$ becomes a stochastic process. The same statistical artefact, providing the feeling that the estimated Hurst index behaves as a stochastic process, would occur for any time-varying Hurst index $H(t)$ which is a $\mathcal{C}^1$ function or a piecewise $\mathcal{C}^1$ function. To sum up, the estimated Hurst index is a stochastic process, while the theoretical Hurst index is a deterministic function regularly varying with time. The same phenomenon appears in the article of Bardet-Surgailis [5, Fig. 2 and Fig. 3, pages 1023-1024]. Similarly, simulations presented in [13, Fig. 3, p.6] and [20, Fig. 1, p.1514] show a path of a mBm with a sine functional Hurst index $H(t)$ with mean 1/2 and the corresponding time-varying Hurst index $\widehat{H}(t)$. Clearly, the theoretical Hurst index $H(t)$ is a $\mathcal{C}^\infty$ function [20, Fig. 1 (c), p.1514], whereas the estimated Hurst index $\widehat{H}(t)$ looks like a continuous Hölder function with regularity $\alpha < 1$, see [20, Fig. 1 (d), p.1514].

This remark led us to introduce a sparse mBm in [10] for application to financial processes. The guiding idea is to choose a simple function $H(t)$ which describes the real dataset as well as a more complicated one.

Let us stress that in this section, we have chosen to provide the underlying ideas, avoiding any technicality.

2 Recalls on fBM, mBm, and statistical estimation of Hurst index

2.1 Recalls on fBm and mBm

One of the most famous Gaussian random processes is the Brownian motion. At the beginning of the 20th century, this process was developed by Louis Bachelier for stock options in finance and next by Albert Einstein in order to describe successive movements of atomic particules independent one from another. Then the mathematical theory is mainly due to Robert Wiener in the 1920's; he proved results on the non differentiability of the paths and the one-dimentional version is kwown as the Wiener process. The fractional Brownian motion (fBm) can hence appear as a generalisation of the Brownian motion.

After the paper of Mandelbrot and Van Ness (1968), modeling by a fBm became more and more widespread, and the statistical study of fBm was developed during the decades 1970's and 1980's. Nevertheless, in many applications the real data do not perfectly fit with fBm. More precisely, statistical tests reject the null hypothesis $H = 1/2$ as it should be for Brownian motion or diffusion processes, but any alternative hypothesis would also be rejected when the Hurst index is varying with time.

In fact, access to larger and larger datasets has shown that real time series look locally like a fBm, but with a time-varying Hurst index $t \mapsto H(t)$ rather than a constant one. This intuition was translated in mathematical modelling, by the introduction of multifractional Brownian motion (mBm) by Peltier, Lévy-Véhel (1995), and Benassi et al. (1997). Indeed, mBm is a continuous Gaussian process whose pointwise Hölder exponent evolves with time t. Recall that for the fBm, the pointwise Hölder exponent and Hurst index are equal. Therefore, a natural idea is to replace the Hurst index H by a function of time $t \mapsto H(t)$ in one of the representations of the fBm. Simultaneously, Peltier, Lévy-Véhel (1995) proposed to replace the Hurst index H by a time-varying one in the moving average representation, whereas Benassi et al.(1997) replaced it by a time-varying one in the harmonisable representation. Actually, both constructions correspond to the same process. Then, to be self-contained, we rely on the work of Ayache and Taqqu in [2] so we define the multifractional Brownian motion (mBm) as follows:

Definition 2.1 *Let $(t,H) \longmapsto B(t,H)$ be the Gaussian field defined by (2). The multi-fractional Brownian motion is defined by*

$$X(t) = B(t, H(t)). \tag{3}$$

2.2 Estimation of the Hurst index for fBm and mBm

Let X be a fBm or a mBm. We observe one path of size n of the process X with mesh h_n, namely $\big(X(0), X(h_n), \dots, X(nh_n)\big)$. For simplicity and without real restriction, we can assume that $h_n = 1/n$. We use quadratic variations to estimate the Hurst index. Let us first give the underlying idea: for a fBm with Hurst index H, we have

$$\mathbb{E}\left(|X(t+h_n) - X(t)|^2\right) \quad = \quad |h_n|^{2H}. \tag{4}$$

On the one hand, the stationarity of the increments of fBm allows us to estimate the variance by the empirical variance and to get a central limit theorem (CLT). On the other hand, we can estimate the variance at M different meshes of time, that is $h_n, 2h_n, \dots, Mh_n$; then linear regression of the logarithm of the empirical variance at those different meshes provides us an estimator of the Hurst index H. Moreover, a CLT is in force. Eventually, by a freezing argument, we can shift the technique from fBm to mBm.

More precisely, let $a = (a_0, \dots, a_\ell)$ be a filter of order p, $(t_k)_{k=1,\dots,n}$ a family of observation times, and X a fBm or a mBm. We define the associated increment by

$$\Delta_a X(t_k) \quad = \quad \sum_{q=0}^{\ell} a_q\, X(t_{k-q}). \tag{5}$$

Saying that a is a filter of order $p \geq 1$ means that

$$\sum_{q=0}^{\ell} a_q\, q^k = 0 \quad \text{for all} \quad k < p \quad \text{and} \quad \sum_{q=0}^{\ell} a_q\, q^p \neq 0. \tag{6}$$

For example, $a = (1, -1)$ is of order 1, whereas $a = (1, -2, 1)$ is of order 2. Next, for a filter $a = (a_0, \dots, a_\ell)$ and any integer $j \in \mathbb{N}$, we define its j^{th} dilatation $a^{(j)} = (a_0^{(j)}, \dots, a_{j\ell}^{(j)})$ by

$$a_{ij}^{(j)} = a_i \quad \text{and} \quad a_k^{(j)} = 0 \quad \text{if } k \notin j\mathbb{N}.$$

Since X is a zero mean Gaussian process, $\Delta_{a^{(j)}} X(t_k)$ is also a zero mean Gaussian variable for any time t_k and any dilatation j. For a fBm, that is when $X = B_H$, its variance is

$$\mathrm{Var}\left[\Delta_{a^{(j)}} B_H(t_k)\right] = C_a \times \left|\frac{j}{n}\right|^{2H}.$$

This variance can be estimated by the empirical variance. However, our aim is the estimation of the Hurst index for a mBm. The guiding idea is that a mBm behaves locally as a fBm. Therefore, we localise the estimation and we compute the empirical variance on a small vicinity of each time t, namely on

$$\mathcal{V}(t, \varepsilon_n) = \big\{t_k \text{ such that } |t_k - t| \le \varepsilon_n\big\},$$

where $\varepsilon_n \to 0$ and $\varepsilon_n / h_n \to \infty$ as $n \to \infty$. To sum up, given a filter a and a real number $t \in (0, 1)$, we set

$$V_n(t, a) \;=\; \frac{1}{v_n} \sum_{t_k \in \mathcal{V}(t, \varepsilon_n)} \left|\Delta_a X(t_k)\right|^2 \tag{7}$$

where $v_n = 2\varepsilon_n / h_n = 2\varepsilon_n \times n$ is asymptotically equivalent to the number of times t_k belonging to $\mathcal{V}(t, \varepsilon_n)$. Eventually, we calculate the empirical variance at M different scales j/n for $j = 1, \dots, M$. Then we set

$$\widehat{H}_n(t) \;=\; \frac{A^t}{2AA^t} \Big(\ln(V_n(t, a^{(j)}))\Big)_{j=1,\dots,M} \tag{8}$$

where A is the row vector defined by

$$A_j \;=\; \ln(j) - \frac{1}{M} \sum_{\nu=1}^{M} \ln(\nu) \quad \text{for } j = 1, \dots, M \tag{9}$$

and A^t the transpose vector (column vector).

Actually the number v_n of terms in sum (7) converges to infinity when $n \to \infty$, thus a CLT is in force with a Gaussian limit. Then, the estimator of the Hurst parameter is also asymptotically Gaussian. More precisely, we can state the following proposition:

Proposition 2.1 (Coeurjolly, 2005–2006) *Let $a = (1, -2, 1)$ be a filter of order 2 as defined by (6), $(t_k = k/n)_{k=1,\dots,n}$ a family of observation times, $X = B_{H(t)}$ a mBm with Hurst index $H(t)$ and $\Delta_a X$ the associate increments defined by (5). Then $\widehat{H}_n(t) \xrightarrow[n\to\infty]{a.s.} H$ and*

$$\sqrt{2\varepsilon_n \cdot n} \times \Big(\widehat{H}_n(t) - H(t)\Big) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}'(t) \tag{10}$$

where $\mathbb{G}'(t)$ *is a zero mean Gaussian process with covariance structure given by*

$$Var(\mathbb{G}'(t)) = \left(\frac{1}{2\|A\|^4} \frac{1}{\pi^a_{H(t)}(0)^2} \sum_{k\in\mathbb{Z}} \pi^a_{H(t)}(k)^2 \right) \times A^t (UU^t) A \tag{11}$$

for all $t \in (0,1)$, *and*

$$cov(\mathbb{G}'(t_1), \mathbb{G}'(t_2)) = 0 \quad \textit{for all } (t_1, t_2) \in (0,1)^2 \quad \textit{with } t_1 \neq t_2 \tag{12}$$

where the row vector A *is defined by (9) and* $U = (1, \ldots, 1)$. *Moreover, for a filter* a, *and an integer* k, *the quantity* $\pi^a_H(k)$ *is defined by*

$$\pi^a_H(k) := -\frac{1}{2} \sum_{q=0}^{\ell} \sum_{q'=0}^{\ell} a_q \, a_{q'} \, |q - q' + k|^{2H}. \tag{13}$$

To sum up, we set

$$\gamma_{H(t)} := Var(\mathbb{G}'(t)) = \Lambda_H(t) \times (B.U.U^t.B^t)$$

with

$$\Lambda_H(t) = \frac{2}{\pi^a_{H(t)}(0)^2} \sum_{k\in\mathbb{Z}} \pi^a_{H(t)}(k)^2 \tag{14}$$

and

$$B = \frac{A^t}{2\|A\|^2}.$$

Proof. The proof can be obtained by combining [18, 19]. However, a more direct and natural proof is provided in Appendix A. □

3 Statement of our main results

We propose a fitting test for a time-varying Hurst index and apply it to a model selection approach, leading to the simplest model.

3.1 Fitting test

As the selection of a good probabilistic model is the guideline of this article, the idea is now to give an adequacy test to select admissible estimators and reject others. For this, we use the previous convergence result. Actually, the CLT given in Proposition 2.1 leads to the following convergence in law:

$$\sqrt{2\varepsilon_n \cdot n} \times \left(\widehat{H}_n(t) - H(t)\right) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}'(t)$$

for all $t \in (0;1)$ where $(\mathbb{G}'(t), t \in]0;1[)$ is a zero mean Gaussian process which covariance structure is known. If $H(.)$ is the theoretical index, then this means that we can explain the L^2 risk function, namely the MISE (Mean Integrated Squared Error) by $\mathbb{E}\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)}$, where

$$\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)} := \frac{1}{n}\sum_{k=1}^{n} |\widehat{H}_n(t_k) - H(t_k)|^2$$

with $(t_k = \frac{k}{n})_{k=1,\ldots,n}$ a family of observation times. Applying the previous CLT, we get the convergence in law

$$2n\varepsilon_n\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)} \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \frac{1}{n}\left[\sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2\right]. \tag{15}$$

Set

$$V_n := \frac{1}{n}\sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2. \tag{16}$$

We can deduce a CLT on V_n as stated in the following proposition

Proposition 3.1 *Under the same assumptions than in Proposition 2.1. Let V_n be defined by (16). We can rewrite V_n as follows*

$$V_n \quad = \quad \mu_n + S_n \times \xi_n \tag{17}$$

with $\mu_n = \mathbb{E}(V_n)$ its mean, $S_n = \sqrt{Var(V_n)}$ its standard deviation. Then we get the convergence in distribution to a standard normal deviate

$$\xi_n \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathcal{N}(0;1). \tag{18}$$

Proof. The proof is given in Appendix B. □

As n converges to infinity, we have

$$\mathbb{E}(V_n) \longrightarrow \int_0^1 \gamma_{H(t)} dt$$

and

$$\left(\frac{n}{2}\right) \times \mathrm{Var}(V_n) \longrightarrow \int_0^1 (\gamma_{H(t)})^2 dt.$$

By replacing these quantities by their limits, we can formulate the fitting test:

Theorem 3.1 *Under the same assumptions as in Proposition 2.1 and Proposition 3.1, we can test the eligibility of a function $\widetilde{H}(t)$ with the theoretical Hurst index. Namely set :*

$$(H_0) : \widetilde{H}(t) = H(t) \quad (19)$$
$$versus\ (H_1) : \widetilde{H}(t) \neq H(t).$$

Then $\widetilde{H}(t)$ is an eligible model if, for a given risk α,

$$|T_n(\widetilde{H}(t))| \leq u_\alpha$$

where $T_n(\widetilde{H}(t))$ is defined by

$$T_n(\widehat{H}_n(t)) = \frac{2n\varepsilon_n \|\widehat{H}_n(t) - \widetilde{H}(t)\|^2_{L^2(]0;1[)} - \int_0^1 \gamma_{\widetilde{H}(t)} dt}{\left(\frac{2}{n}\int_0^1 (\gamma_{\widetilde{H}(t)})^2 dt\right)^{1/2}} \quad (20)$$

where $\gamma_{H(t)} := Var\left(\mathbb{G}'(t)\right)$ is given by Formula (11) and u_α denotes the fractile of order $(1-\frac{\alpha}{2})$ of the standard normal law.

Proof. The proof is given in Appendix B. □

For instance, given a risk $\alpha = 0.05$, we accept the null hypothesis if $T_n(\widetilde{H}(t)) \in [-1.96, 1.96]$.

3.2 Application to model selection

As a by-product, the naive time-varying estimator $\widehat{H}(t)$ of the Hurst index could not be chosen as a valid model. Namely, it is not an admissible one. Nevertheless, the assumption $\widetilde{H}(t) = \widehat{H}(t)$ in the null hypothesis (19) is asymptotically rejected, as stated in the following corollary

Corollary 3.1 *if $\widetilde{H}(t) = \widehat{H}_n(t)$ we get*

$$T_n(\widetilde{H}(t)) = \frac{-\int_0^1 \gamma_{\widetilde{H}(t)} dt}{\left(\frac{2}{n}\int_0^1 (\gamma_{\widetilde{H}(t)})^2 dt\right)^{1/2}} = -\sqrt{\frac{n}{2}} \times \frac{\|\gamma_{\widetilde{H}(t)}\|_{L^1(]0;1[)}}{\|\gamma_{\widetilde{H}(t)}\|_{L^2(]0;1[)}} \longrightarrow \infty \ as\ n \to \infty$$

and then, as we are in the critical region, the null hypothesis (H_0) is rejected.

Proof. The proof is deduced from Theorem 3.1. □

The next idea is to determine the simplest possible function $\widetilde{H}(t)$ that will describe the theoretical Hurst index $H(t)$. Note that such a model is in the same time simpler and fits better the theoretical value of the Hurst index as it does not contain the statistical artefact. We are hence able to look for a suitable model; the aim is to determine the most simple model that is eligible for test (19). This model selection is a kind of Portemanteau test. Thus, for this, set

- $\mathcal{M}_0$ the family of constant models $\widetilde{H}(t) = H$
- $\mathcal{M}_1$ the family of affine models $\widetilde{H}(t)$
- $\mathcal{M}_2$ the family of piecewise affine models $\widetilde{H}(t)$
- $\mathcal{M}_3$ the family of quadratic models $\widetilde{H}(t)$
- $\mathcal{M}_4$ the family of piecewise quadratic models $\widetilde{H}(t)$.

We successively test models extracted from the previous families. Those families of models are classified by order of complexity of function $\widetilde{H}(t)$. We stop and use the first eligible model, namely for family $\mathcal{M}_i$ with the lowest i. By construction of these families, the selected model is thus the simplest one.

Conclusion

To sum up, the naive multifractional estimator $\widehat{H}_n(t)$ is too complicated and has too many fluctuations that appear as a statistical artefact as shown in Fig. 1. Moreover, we have built a fitting test which asymptotically rejects $\widehat{H}_n(t)$ as an appropriate estimator of the time-varying theoretical Hurst index $H(t)$. Next, this fitting test is used to select the simplest time-varying Hurst index $\widetilde{H}(t)$ from a given families of models, by a Portemanteau procedure. We have proposed in Sect. 3.2 a family of piecewise polynomial functions. However different choices are possible such as logistic functions, see e.g. [25, Fig. 2, p.101].

In a certain way, our work confirms and enhances the multifractional process with random Hurst exponent (MPRE) introduced by Ayache and Taqqu (2005) [2]. Indeed, the Hurst exponent could be random without being itself a stochastic process. For instance a piecewise affine (or quadratic function) with change of slope at random times is still a random exponent, without having to oscillate roughly, see e.g. [10, Fig. 6, p. 15]. So, we have disentangled a random time-varying Hurst exponent from a roughly oscillating exponent resulting from a statistical artefact. This result also better fits the interpretations proposed by scholars from applied fields: it is simpler to interpret a slowly varying function taking values larger or smaller than the nominal value [30, 25, 35, 14, 20].

Let us add that the selected model is both simpler and fits better the theoretical value of the Hurst index $H(t)$. Consequently, this study opens the way to further research like online detection of change of slope of the Hurst index or study of the different kinds of families of model.

References

[1] Ayache, A, Bertrand, PR, Lévy-Véhel J (2007), A Central Limit Theorem for the Generalized Quadratic Variation of the Step Fractional Brownian Motion. *Statistical Inference for Stochastic Processes* **10**, 1–27.

[2] Ayache, A, Taqqu, MS (2005), Multifractional processes with random exponent. *Publ. Mat.* **49**, 459–486.

[3] Ayache, A, Jaffard, S, Taqqu, MS (2007) Wavelet construction of Generalized Multifractional processes. *Revista Matematica Iberoamericana*, **23**, No 1, 327–370.

[4] Bardet, JM, Bertrand, PR (2007). Definition, properties and wavelet analysis of multiscale fractional Brownian motion. *Fractals* 15, 73–87.

[5] Bardet, JM, Surgailis, D (2013). Nonparametric estimation of the local Hurst function of multifractional processes. *Stochastic Processes and Applications*, 123, 1004–1045.

[6] Benassi A, Jaffard S, Roux D (1997), Elliptic Gaussian random processes. *Revista Matematica Iberoamericana*, 13(1):19-90.

[7] Benassi A, Bertrand PR, Cohen S, Istas J (2001) Identification of the Hurst index of a step fractional Brownian motion. *Statistical Inference for Stochastic Processes* 3(1/2): 101–111.

[8] Bertrand, PR, Bardet, JM (2001) " Some generalization of fractional Brownian motion and Control", in *Optimal Control and Partial Differential Equations*, J.L. Menaldi, E. Rofman and A. Sulem editors, p.221-230, IOS Press.

[9] Bertrand, PR (2005) "Financial Modelling by Multiscale Fractional Brownian Motion.", in *Fractals in Engineering. New trends in theory and applications*, edited by Lutton E. and Levy Vehel J., p.159-175, Springer Verlag.

[10] Bertrand, PR, Hamdouni, A, Khadhraoui S (2012), Modelling NASDAQ series by sparse multifractional Brownian motion, *Methodology and Computing in Applied Probability*. Vol. 14, No 1: 107–124.

[11] Bianchi, S (2005) Pathwise Identification of the memory function of multifractional Brownian motion with application to finance. *International Journal of Theoretical and Applied Finance* Vol. 8, No. 2, 255–281.

[12] Bianchi, S, Pianese, A (2008) Modeling Stock Prices by Multifractional Brownian Motion: An Improved Estimation of the Pointwise Regularity *International Journal of Theoretical and Applied Finance* Vol.11, No 6, 567–595.

[13] Bianchi, S, Pantanella, A, Pianese, A (2011) Modeling Stock Prices by Multifractional Brownian Motion: An Improved Estimation of the Pointwise Regularity *Quantitative Finance* 13(8)

[14] Bianchi, S, Pantanella, A, Pianese, A (2015) Efficient Markets and Behavioral Finance: a comprehensive multifractal model. *Advances in Complex Systems* 18.

[15] Bianchi, S, Pianese, A (2014) Multifractional Processes in Finance. SSRN Electronic Journal 5(1):1-22

[16] Breuer, P, Major, P (1983) Central Limit Theorems for Non-Linear Functionals of Gaussian Fields. *J.of Multivariate Analysis* **13**, 425–441.

[17] Cheridito P(2003) Arbitrage in Fractional Brownian Motion Models. *Finance Stoch.* 7(4): 533–553.

[18] Coeurjolly, JF (2005), Identification of multifractional Brownian motion, *Bernoulli* **11** (6), 987–1008.

[19] Coeurjolly, JF (2006), Erratum: Identification of Multifractional Brownian Motion, *Bernoulli* **12** (2), 381–382.

[20] Frezza, M (2012) Modeling the time-changing dependence in stock markets. *Chaos Solitons & Fractals* 45(12):1510–1520

[21] Guyon, X, Leon, JR (1989), Convergence en loi des H-variations d'un processus gaussien fractionnaire. *Ann. Inst. H. Poincaré* **25** (3), 265–282.

[22] Istas, J, Lang, G (1997), Quadratic variation and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré* **33** (4), 407–437.

[23] Kolmogorov AN (1940) Wienersche Spiralen und einige andere interessante Kurven in Hilbertchen Raume. Doklady 26: 115–118.

[24] Peltier, RF, Lévy-Véhel, J (1995), Multifractional Brownian motion: definition and preliminary results, Research Report RR-2645, INRIA, Rocquencourt, http://www.inria.fr/rrrt/rr-2645. html.

[25] Lee KC (2013) Characterization of turbulence stability through the identification of multifractional Brownian motions *Nonlinear Processes in Geophysics* 20(1):97-106.

[26] Mandlebrot, BB (2012) *The fractalist. Memoir of a Scientific Maverick.* Pantheon.

[27] Lim SC , Teo LP (2009) Modeling Single-File Diffusion by Step Fractional Brownian Motion and Generalized Fractional Langevin Equation. *Journal of Statistical Mechanics Theory and Experiment*

[28] Marquez-Lago TT, Leier A, Burrage K. Anomalous diffusion and multifractional Brownian motion: simulating molecular crowding and physical obstacles in systems biology *IET Systems Biology* 6(4):134-42.

[29] Ryvkina, J (2013), Fractional Brownian Motion with variable Hurst parameter: Definition and properties. *Journal of Theoretical Probability* DOI 10.1007/s10959-013-0502-3 Print ISSN 0894-9840

[30] Papanicolaou, G, Sølna, K (2002) Wavelet based estimation of local Kolmogorov turbulence. In: P. Doukhan, G. Oppenheim and M.S. Taqqu (eds), *Long-range Dependence: Theory and Applications*, Birkhäuser, Basel, (2002), 473-506.

[31] Rásonyi M (2009) Rehabilitating Fractal Models in Finance. *ERCIM News*: 19–20.

[32] Rogers LCG (1997) Arbitrage with fractional Brownian motion. *Mathematical Finance* 7: 95–105.

[33] Samorodnitsky, G, Taqqu MS (1994), *Stable non-Gaussian Random Processes*, Chapman and Hall.

[34] Shiryaev AN (1998) On arbitrage and replication for fractal models. Technical report. MaPhySto.

[35] Wanliss JA, Dobias, P (2007) Space storm as a phase transition *Journal of Atmospheric and Solar-Terrestrial Physics* 69 (2007) 675-684.

A Proof of Proposition 2.1

The proof is divided in three steps. Both Step 1 and Step 2 are concerned with fBm. In Step 1, we prove a CLT for the localised quadratic variations of order 2 of the fBm. In Step 2, we deduce a CLT for the estimator of Hurst index obtained by linear regression of the logarithm of quadratic variation at different meshes of times for fBm. Next, Step 3 explains how to shift from fBm to mBm. Before going further, let us state a technical lemma used in Step 1.

Lemma A.1 *Let a be a filter of order $p \geq 1$ as defined by (6), $(t_k = k/n)_{k=1,\dots,n}$ a family of observation times, $X = B_H$ a fBm with covariance given by (1), and $\Delta_a X$ the associate increments defined by (5). Then*

1. *$\Delta_a X(.)$ is a zero mean Gaussian vector, with covariance structure given by*

$$cov\left[\Delta_a X(t_k), \Delta_a X(t_{k'})\right] \;=\; \sigma^2 n^{-2H} \pi_H^a(k-k') \tag{21}$$

   for all pair (k,k'), where $\pi_H^a(k)$ is defined by (13).

2. *As a by-product, for all k*

$$Var\left[\Delta_a X(t_k)\right] \;=\; \mathbb{E}\left|\Delta_a X(t_k)\right|^2 = \sigma^2 n^{-2H} \pi_H^a(0). \tag{22}$$

3. *Moreover, for all $k \in \mathbb{Z}$*

$$\left|\pi_H^a(k)\right| \;\leq\; Ctte \times |k|^{2H-2p}. \tag{23}$$

Remark A.1 *Formula (22) implies that* $Var[\Delta_a X(t_k)] = Ctte \times h_n^{2H}$, *with* $h_n = 1/n$. *This proves and generalises formula (4).*

Proof. 1) and 2) Since $X = B_H$ is a zero mean Gaussian process, we deduce that $\Delta_a X(t_k)$ is a zero mean Gaussian vector, with covariance structure given by

$$\begin{aligned}
\text{cov}\left[\Delta_a X(t_k), \Delta_a X(t_{k'})\right] &= \sum_{q=0}^{\ell}\sum_{q'=0}^{\ell} a_q\, a_{q'} \,\text{cov}\left[B_H(t_{k-q}), B_H(t_{k'-q'})\right] \\
&= \frac{\sigma^2}{2}\sum_{q=0}^{\ell}\sum_{q'=0}^{\ell} a_q\, a_{q'}\, \left\{|t_{k-q}|^{2H} + |t_{k'-q'}|^{2H} - |t_{k-q} - t_{k'-q'}|^{2H}\right\} \\
&= -\frac{\sigma^2}{2}\sum_{q=0}^{\ell}\sum_{q'=0}^{\ell} a_q\, a_{q'}\, |t_{k-q} - t_{k'-q'}|^{2H},
\end{aligned}$$

where the last equality follows from Eq.(6). Next, by setting $t_k = k/n$, we can deduce Formula (21). As pointed in Lemma A.1, Formula (22) follows from Formula (21).

3) See Coeurjolly (2001, lemma 1). □

Step 1: CLT for quadratic variations of fBm.

For a fBm $X = B_H$, the variance of the increments does not depend on the time t_k, see Lemma A.1 Formula (22). Thus from (7) we get

$$\begin{aligned}
V_n(t,a) &= \text{Var}\left[\Delta_a X(t_k)\right] \times \left\{1 + \frac{1}{v_n}\sum_{t_k \in \mathcal{V}(t,\varepsilon_n))}\left[\frac{\left|\Delta_a X(t_k)\right|^2}{\mathbb{E}\left|\Delta_a X(t_k)\right|^2} - 1\right]\right\} \\
&= \sigma^2 \left(\frac{1}{n}\right)^{2H} \pi_H^a(0) \times \left\{1 + \widetilde{V}_n(t,a)\right\} \qquad (24)
\end{aligned}$$

where

$$\widetilde{V}_n(t,a) := \frac{1}{v_n}\sum_{t_k \in \mathcal{V}(t,\varepsilon_n))}\left[\left(Z_k^{(a)}\right)^2 - 1\right]$$

and

$$Z_k^{(a)} := \frac{\Delta_a X(t_k)}{\sqrt{\mathbb{E}\left|\Delta_a X(t_k)\right|^2}}. \qquad (25)$$

For notational convenience, we drop the index (a) in the sequel, and we note that for $k = 1, \ldots, n$, Z_k forms a stationary family of zero mean standard Gaussian variables with correlation

$$r(k) = \text{corr}(Z_j, Z_{j+k}) = \frac{\pi_H^a(k)}{\pi_H^a(0)} \qquad (26)$$

where π_H^a is defined by (13). Actually, $Z_k^2 - 1 = H_2(Z_k)$ where $H_2(x) = x^2 - 1$ is the Hermite polynomial of order 2. By using Breuer-Major Theorem [16], a CLT with a Gaussian limit is in force as soon as $\sum_{k\in\mathbb{Z}} r(k)^2 < \infty$. Combining Formula (26) and Bound (23) in Lemma A.1, we deduce that $r(k)^2 = \mathcal{O}(k^{4H-4p})$. The series $\sum_{k\in\mathbb{Z}} k^{4H-4p}$ converges if and only if $4H - 4p < -1$ or equivalently iff $H < p - 1/4$. So, for $p = 1$ which corresponds to the case $a = (1, -1)$ and quadratic variations, we get a CLT iff $H < 3/4$; whereas for $p = 2$ with $a = (1, -2, 1)$, and the so-called generalised quadratic variations (GQV), the CLT is in force for all Hurst index $H \in (0, 1)$. For these reasons we do prefer the use of GQV rather than simple quadratic variations, see also Istas-Lang or Guyon-Leon.

Note that for $a = (1, -2, 1)$, we get $\pi_H^a(0) = 4 - 2^{2H}$.

The rate of convergence is $\sqrt{v_n} = \sqrt{2n\varepsilon_n}$. Moreover,

$$\mathbb{E}\Big(H_2(Z_k)\cdot H_2(Z_{k'})\Big) = 2\left[\mathbb{E}\big(Z_k\cdot Z_{k'}\big)\right]^2.$$

This relation combined with (26) involves the following calculation of the variance

$$\begin{aligned}
\mathbb{E}\left\{\left[\sqrt{v_n}\cdot \widetilde{V}_n(t,a)\right]^2\right\} &= \frac{1}{v_n}\times \sum_{k,t_k\in\mathcal{V}(t,\varepsilon_n)}\ \sum_{k',t_{k'}\in\mathcal{V}(t,\varepsilon_n)} \mathbb{E}(H_2(Z_k)\cdot H_2(Z_{k'}))\\
&= \frac{2}{v_n}\times \sum_{k,t_k\in\mathcal{V}(t,\varepsilon_n)}\ \sum_{k',t_{k'}\in\mathcal{V}(t,\varepsilon_n)} \Big[\mathbb{E}(Z_k\cdot Z_{k'})\Big]^2\\
&= \frac{2}{v_n}\times \sum_{|k|<v_n} \big(v_n - |k|\big)\times \frac{\pi_H^a(k)^2}{\pi_H^a(0)^2}\\
&= \frac{2}{\pi_H^a(0)^2}\times \sum_{|k|<v_n}\left(1 - \frac{|k|}{v_n}\right)\times \pi_H^a(k)^2
\end{aligned}$$

But, $\sum_{k\in\mathbb{Z}} \pi_H^a(k)^2 < \infty$ since $p = 2$, therefore

$$\lim_{n\to\infty} \mathbb{E}\left\{\left[\sqrt{v_n}\cdot \widetilde{V}_n(t,a)\right]^2\right\} = \frac{2}{\pi_H^a(0)^2}\times \sum_{k\in\mathbb{Z}} \pi_H^a(k)^2.$$

To sum up, Breuer-Major Theorem induces that

$$\sqrt{v_n}\cdot \widetilde{V}_n(t,a) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}(t)$$

where $\mathbb{G}(t)$ is a zero mean Gaussian process with variance (see formula (14)) $\mathrm{Var}(\mathbb{G}(t)) = \dfrac{2}{\pi_H^a(0)^2}\times \sum_{k\in\mathbb{Z}} \pi_H^a(k)^2 := \Lambda_H(t)$ for all $t \in (0,1)$ and covariance $\mathrm{cov}(\mathbb{G}(t_1), \mathbb{G}(t_2)) = 0$ for all pair $(t_1, t_2) \in (0,1)^2$ with $t_1 \neq t_2$.

Step 2: CLT for estimation of the Hurst index of fBm.

The estimator of Hurst index as defined in (8) is obtained by linear regression of the log-variance with the j-dilated filters $a^{(j)}$. Stress that the j-dilated filters $a^{(j)}$ behave like the a filter with mesh $h_n^{(j)} = j/n$ instead of $1/n$. Then, Eq.(24) is replaced by

$$V_n(t, a^{(j)}) \quad = \quad \sigma^2 \left(\frac{j}{n}\right)^{2H} \pi_H^a(0) \times \left\{ 1 + \widetilde{V}_n(t, a^{(j)}) \right\} \tag{27}$$

with

$$\sqrt{2\varepsilon_n \times n} \cdot \widetilde{V}_n(t, a^{(j)}) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}_j(t). \tag{28}$$

By taking the logarithm of Eq. (27), we have

$$\ln V_n(t, a^{(j)}) \quad = \quad 2H \ln(j/n) + \ln\left(\sigma^2\, \pi_H^a(0)\right) + \ln\left\{1 + \widetilde{V}_n(t, a^{(j)})\right\}.$$

Next, by Eq. (28), we get

$$\ln V_n(t, a^{(j)}) \quad \simeq \quad 2H \ln(j/n) + \ln\left(\sigma^2\, \pi_H^a(0)\right) + \frac{1}{\sqrt{2\varepsilon_n \times n}} \ln \mathbb{G}_j(t).$$

Therefore the Hurst index H can be estimated as the slope by linear regression of the family $\{\ln V_n(t, a^{(j)}), j = 1, \dots, M\}$ onto the predictor $\big(\ln(j/n)\big)_{j=1,\dots,M}$. Thus

$$\widehat{H}_n(t) \quad = \quad \frac{A^t}{2AA^t} \Big(\ln V_n(t, a^{(j)}) \Big)_{j=1,\dots,M}. \tag{29}$$

The right Hurst index is obtained by canceling the stochastic part. By doing so, it comes

$$H \quad = \quad \frac{A^t}{2AA^t} \Big(\ln\left(\sigma^2 (j/n)^{2H}\, \pi_H^a(0)\right) \Big)_{j=1,\dots,M}$$

then

$$\begin{aligned} \widehat{H}_n(t) - H \quad &= \quad \frac{A^t}{2AA^t} \Big(\ln V_n(t, a^{(j)}) - \ln\left(\sigma^2 (j/n)^{2H}\, \pi_H^a(0)\right) \Big)_{j=1,\dots,M} \\ &= \quad \frac{A^t}{2AA^t} \Big(\ln P_n^j(t) \Big)_{j=1,\dots,M} \end{aligned}$$

where we have set

$$P_n^j(t) \quad = \quad \left(\frac{n}{j}\right)^{2H} \times \frac{V_n(t, a^{(j)})}{\sigma^2 \pi_H^a(0)}.$$

On the other hand, Eq. (27) implies

$$P_n^j(t) \quad = \quad \left\{ 1 + \widetilde{V}_n(t, a^{(j)}) \right\}.$$

Moreover Eq. (28) implies that $\widetilde{V}_n(t, a^{(j)})$ converges to 0 as $n \to \infty$. Therefore

$$\ln P_n^j(t) \simeq \widetilde{V}_n(t, a^{(j)}). \tag{30}$$

Next, in order to get the covariance of $\mathbb{G}'$ stated in Prop. 2.1, we look at the following covariance structure:

$$\begin{aligned}
& \operatorname{cov}\left(\sqrt{2\varepsilon_n n}\left(\widehat{H}_n(t_1) - H\right), \sqrt{2\varepsilon_n n}\left(\widehat{H}_n(t_2) - H\right)\right) \\
= \ & \operatorname{cov}\left(\sqrt{2\varepsilon_n n}\frac{A^t}{2AA^t}\left(\ln P_n^j(t_1)\right)_{j=1,\dots,M}, \sqrt{2\varepsilon_n n}\frac{A^t}{2AA^t}\left(\ln P_n^j(t_2)\right)_{j=1,\dots,M}\right) \\
\simeq \ & \operatorname{cov}\left(\sqrt{2\varepsilon_n n}\frac{A^t}{2AA^t}\left(\widetilde{V}_n(t_1, a^{(j)})\right)_{j=1,\dots,M}, \sqrt{2\varepsilon_n n}\frac{A^t}{2AA^t}\left(\widetilde{V}_n(t_2, a^{(j)})\right)_{j=1,\dots,M}\right) \\
= \ & \operatorname{cov}\left(\frac{A^t}{2AA^t}\left(\sqrt{2\varepsilon_n n}\,\widetilde{V}_n(t_1, a^{(j)})\right)_{j=1,\dots,M}, \frac{A^t}{2AA^t}\left(\sqrt{2\varepsilon_n n}\,\widetilde{V}_n(t_2, a^{(j)})\right)_{j=1,\dots,M}\right) \\
\simeq \ & \operatorname{cov}\left(\frac{A^t}{2AA^t}\left(\mathbb{G}_j(t_1)\right)_{j=1,\dots,M}, \frac{A^t}{2AA^t}\left(\mathbb{G}_j(t_2)\right)_{j=1,\dots,M}\right)
\end{aligned}$$

where we have successively used Eq. (30) and Eq. (28). Since $\operatorname{cov}(\mathbb{G}(t_1), \mathbb{G}(t_2)) = 0$ for all pair $(t_1, t_2) \in (0,1)^2$ with $t_1 \neq t_2$ (see Step 1), we get

$$\lim_{n\to\infty} \operatorname{cov}\left(\sqrt{2\varepsilon_n n}\left(\widehat{H}_n(t_1) - H\right), \sqrt{2\varepsilon_n n}\left(\widehat{H}_n(t_2) - H\right)\right) = 0$$

for all pair $(t_1, t_2) \in (0,1)^2$ with $t_1 \neq t_2$, which induces (12). Similarly, when $t_1 = t_2$, we get (11).

Step 3: Freezing

The freezing technics insure that mBm behaves almost as a fBm of Hurst index $H(t_0)$ in a small enough vicinity of time t_0. Therefore our strategy is to show that the freezing error is negligible with respect to the rate of convergence of the estimator of the Hurst index for fBm. We recall the Ayache-Taqqu Theorem and its corollary

Theorem A.1 (Ayache, Taqqu (2005)) *Let $B(t, H)$ be the field defined by Eq. (3). There exists an event Ω^* with $\mathbb{P}(\Omega^*) = 1$ on which $B(t, H)$ is $\mathcal{C}^\infty$ with respect to the variable H, uniformly for all (t, H) in any compact subset $[-T, T] \times [a, b] \subset \mathbb{R} \times (0, 1)$.*

Proof. The proof relies on wavelet series expansion of fBm, see [2, Th 2.1 and Prop. 2.2, item c), p.467]. □

Corollary A.1 *Let X be a mBm as defined by Eq. (3). Assume that $t \mapsto H(t)$ is an η-Hölder continuous function, then there exists a random variable $C_1(\omega)$ with finite moment of every order such that for all $t_0 \in (0,1)$ and $\varepsilon > 0$, we have*

$$\left|X(t) - B_{H(t_0)}(t)\right| \ \le \ C_1(\omega) \times \varepsilon^\eta, \quad \textit{for all } t \in \mathcal{V}(t_0, \varepsilon).$$

Proof. Indeed, from Th. A.1 and Hölder continuity, we get:

$$\begin{aligned} \left|X(t) - B_{H(t_0)}(t)\right| &\leq C_2(\omega) \times |H(t) - H(t_0)| \\ &\leq C_2(\omega) \times M_1 |t - t_0|^\eta \\ &\leq M_1 \cdot C_2(\omega) \times \varepsilon^\eta, \end{aligned}$$

for all t such that $|t - t_0| \leq \varepsilon$. By setting $C_1(\omega) = M_1 \cdot C_2(\omega)$, this finishes the proof of Cor. A.1. □

From Cor. A.1, we then get

$$X(t) = B_{H_0}(t) + \xi(t) \tag{31}$$

for all $t \in \mathcal{V}(t_0, \varepsilon_n) =]t_0 - \varepsilon_n, t_0 + \varepsilon_n[$,

- where $H_0 = H(t_0)$,
- η is the Holder regularity of map $t \mapsto H(t)$, for t in the vicinity of t_0,
- and $|\ \xi(t)| \leq C_1(\omega)\, \varepsilon_n^\eta$ where the random variable C_1 has finite moment of every order.

We deduce from (31) that

$$\Delta_a X(t) = \Delta_a B_{H_0}(t) + \Delta_a \xi(t),$$

for all $t \in \mathcal{V}(t_0, \varepsilon_n) =]t_0 - \varepsilon_n, t_0 + \varepsilon_n[$. Our strategy, in the rest of the proof of Step 3, is to make an expansion in the vicinity of time t_0, then around B_{H_0}. Indeed, since $\varepsilon_n = n^{-\alpha}$, for all $t \in \mathcal{V}(t_0, \varepsilon)$, we have

$$\Delta_a B_{H_0}(t) \sim n^{-H_0} \quad \text{and} \quad \left|\Delta_a \xi(t)\right| \leq 2\, C_1(\omega)\, n^{-\alpha\eta}. \tag{32}$$

The condition

$$\alpha \cdot \eta > H_0 \tag{33}$$

insures that $\Delta_a \xi(t)$ is infinitely smaller than $\Delta_a B_{H_0}(t)$, uniformly for all $t \in \mathcal{V}(t_0, \varepsilon)$. Then V_n, as defined by (7), becomes

$$\begin{aligned} V_n(X, t_0, a^{(j)}) &= \frac{1}{v_n} \sum_{t_k \in \mathcal{V}(t_0, \varepsilon_n)} \left\{ \left|\Delta_a B_{H_0}(t_k)\right|^2 + 2\Delta_a B_{H_0}(t_k) \cdot \Delta_a \xi(t_k) + \left|\Delta_a \xi(t_k)\right|^2 \right\} \\ &= V_n(B_{H_0}, t_0, a^{(j)}) + \left(\frac{2}{v_n} \sum_{t_k \in \mathcal{V}(t_0, \varepsilon_n)} \Delta_a B_{H_0}(t_k) \cdot \Delta_a \xi(t_k) \right) + V_n(\xi, t_0, a^{(j)}) \end{aligned}$$

We can easily deduce from (32) and condition (33), that

$$\left(\frac{2}{v_n}\sum_{t_k\in\mathcal{V}(t_0,\varepsilon_n)}\Delta_a B_{H_0}(t_k)\cdot\Delta_a\xi(t_k)\right)+V_n(\xi,t_0,a^{(j)})$$

is infinitely smaller than $V_n(B_{H_0},t_0,a^{(j)})$. Next Taylor expansion induces

$$\ln V_n(X,t,a^{(j)}) \quad = \quad \ln V_n(B_{H_0},t,a^{(j)})+\frac{\left(\frac{2}{v_n}\sum_{t_k\in\mathcal{V}(t_0,\varepsilon_n)}\Delta_a B_{H_0}(t_k)\cdot\Delta_a\xi(t_k)\right)+V_n(\xi,t_0,a^{(j)})}{V_n(B_{H_0},t_0,a^{(j)})}$$

By Cauchy-Schwarz inequality, we get

$$\Big|\frac{1}{v_n}\sum_{t_k\in\mathcal{V}(t_0,\varepsilon_n)}\Delta_a B_{H_0}(t_k)\cdot\Delta_a\xi(t_k)V_n(B_{H_0},t_0,a^{(j)})\Big| \quad \leq \quad V_n(B_{H_0},t,a^{(j)})^{1/2}\times V_n(\xi,t_0,a^{(j)})^{1/2}$$

which implies

$$\ln V_n(X,t,a^{(j)}) \quad = \quad \ln V_n(B_{H_0},t_0,a^{(j)})+2\mu\,\theta_n^{1/2}+\theta_n \tag{34}$$

where $\mu\in[-1,1]$ and $\theta_n=\dfrac{V_n(\xi,t_0,a^{(j)})}{V_n(B_{H_0},t,a^{(j)})}$.

Lemma A.2 *Under the same assumptions as previously,*

$$\big|\theta_n\big| \quad \leq \quad C_3(\omega)\times\varepsilon_n^{2\eta'} \tag{35}$$

where the variable C_3 has finite moment of every order and $\eta'=\eta-\frac{H_0}{\alpha}$.

Proof. Cor. A.1 implies the bound (32), which induces $V_n(\xi,t_0,a^{(j)})\leq 2\,C_1(\omega)\,n^{-2\alpha\eta}$. Indeed, $V_n(\xi,t_0,a^{(j)})$ is the average of the quantities $\big|\Delta_a\xi(t_k)\big|^2$ for $t_k\in\mathcal{V}(t_0,\varepsilon_n)$, which are uniformly bounded by $2\,C_1(\omega)\,n^{-2\alpha\eta}$. Next, by using formula (27) we get

$$\begin{aligned}\big|\theta_n\big| \quad &\leq \quad \frac{2\,C_1(\omega)\;j^{2H_0}}{\sigma^2\pi_H^a(0)\times\left\{1+\widetilde{V}_n(t_0,a^{(j)})\right\}}\times n^{-2(\alpha\eta-2H_0)}\\ &\leq \quad \frac{2\,j^{2H_0}}{\sigma^2\pi_H^a(0)}\times\frac{C_1(\omega)}{\left\{1+\widetilde{V}_n(t_0,a^{(j)})\right\}}\times\varepsilon_n^{2\eta'}\end{aligned}$$

This proves the bound (35) with

$$C_3(\omega)=\frac{2\,j^{2H_0}}{\sigma^2\pi_H^a(0)}\times\frac{C_1(\omega)}{\left\{1+\widetilde{V}_n(t_0,a^{(j)})\right\}}.$$

Then, it remains to prove that C_3 has finite moments, namely that

$$\mathbb{E}\left(\frac{1}{\left\{1+\widetilde{V}_n(t_0,a^{(j)})\right\}^l}\right)<\infty,$$

for all $l\in\mathbb{N}$. For this, by Hölder inequality, we get

$$\forall l\in\mathbb{N},\ \mathbb{E}\left(|C_3(\omega)|^l\right)\leq\mathbb{E}\left(C_1^{lq}(\omega)\right)^{\frac{1}{q}}\times\mathbb{E}\left(\frac{1}{(1+\widetilde{V}_n(t_0,a^{(j)}))^{lp}}\right)^{\frac{1}{p}}$$

where $\frac{1}{p}+\frac{1}{q}=1$. We deduce from Corollary A.1 that

$$\mathbb{E}\left(C_1^{lq}(\omega)\right)<+\infty.$$

Thus it remains to prove that the other part as finite moments of any order :

$$\mathbb{E}\left(\frac{1}{(1+\widetilde{V}_n(t_0,a^{(j)}))^{lp}}\right)<+\infty.$$

But from Formula (24) (see Step 1) we get

$$1+\widetilde{V}_n(t_0,a^{(j)})=\frac{1}{v_n}\sum_{t_k\in\mathcal{V}(t,\varepsilon_n))}\left[\frac{\left|\Delta_a B_{H_0}(t_k)\right|^2}{\mathrm{Var}\left(\Delta_a B_{H_0}\right)}\right]$$

where $\Delta_a B_{H_0}(t_k)$ is a centered Gaussian random variable. Therefore we get

$$1+\widetilde{V}_n(t_0,a^{(j)})=\frac{1}{v_n}\sum_{t_k\in\mathcal{V}(t,\varepsilon_n))}\left|Z_k\right|^2$$

where Z_k is a standard random variable defined by (25) as $Z_k:=\dfrac{\Delta B_{H_0}(t_k)}{\sqrt{\mathrm{Var}\left|\Delta B_{H_0}(t_k)\right|}}$. Moreover random variables Z_k are weakly dependent, which implies that $1+\widetilde{V}_n(t_0,a^{(j)})\ \xrightarrow[n\to\infty]{\mathcal{D}}\ \chi^2_{d_n}$, see e.g. Istas-Lang 1997 [22] or Ayache-Bertrand-Lévy-Vehel 2007 [1], with $d_n\to+\infty$ as $n\to+\infty$. We can deduce that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{(1+\widetilde{V}_n(t_0,a^{(j)}))^{lp}}\right) &\simeq \mathbb{E}\left(\frac{1}{(\chi^2_{d_n})^{lp}}\right)\\ &= \int_0^\infty\frac{1}{t^{lp}}t^{\frac{d_n}{2}-1}e^{-t/2}dt\\ &= \int_0^\infty t^{\frac{d_n}{2}-lp-1}e^{-t/2}dt<\infty.\end{aligned}$$

This finishes the proof of Lemma A.2. □

Next using (29) combined with (34) and Lemma A.2, we get

$$\begin{aligned}\widehat{H}_n(X,t) &= \frac{A^t}{2AA^t}\Big(\ln V_n(X,t_0,a^{(j)})\Big)_{j=1,\ldots,M} \\ &= \frac{A^t}{2AA^t}\Big(\ln V_n(B_{H_0},t_0,a^{(j)}) + 2\mu\,\theta_n^{1/2} + \theta_n\Big)_{j=1,\ldots,M} \\ &= \widehat{H}_n(B_{H_0},t) + \mathcal{O}(\varepsilon_n^{\eta'}).\end{aligned}$$

But for each fixed Hurst index H_0, the following CLT, given by (10), holds

$$\sqrt{2\varepsilon_n \cdot n} \times \Big(\widehat{H}_n(B_{H_0},t_0) - H_0\Big) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}'(t).$$

This CLT remains in force for mBm X as soon as the freezing error is negligible with respect to the rate of convergence of the estimator of the Hurst index for fBm, namely

$$\varepsilon_n^{\eta'} \ll \sqrt{2\varepsilon_n \cdot n}.$$

By taking $\varepsilon_n = n^{-\alpha}$ we get the following condition

$$n^{-\alpha}\varepsilon_n^{\eta'} \ll \sqrt{2n \times n^{-\alpha}}$$

namely

$$n^{-\alpha}\varepsilon_n^{\eta'} \ll \sqrt{2}n^{\frac{\alpha-1}{2}}$$

which means that the Necessary and Sufficient Condition is

$$2\alpha\eta' > 1-\alpha$$

which is equivalent to

$$\alpha > \frac{1}{1+2\eta'} := \phi(\eta').$$

This finishes the proof of Step 3 (freezing) and consequently the proof of Proposition 2.1.

B Proof of our main result - Theorem 3.1

Using CLT given in Proposition (2.1), by Formula (10) we get this convergence in law:

$$\sqrt{2\varepsilon_n \cdot n} \times \Big(\widehat{H}_n(t) - H(t)\Big) \quad \xrightarrow[n\to\infty]{\mathcal{D}} \quad \mathbb{G}'(t)$$

for all $t \in (0;1)$ where $(\mathbb{G}'(t), t \in]0;1[)$ is a zero mean Gaussian process which covariance structure is known. If $H(.)$ is the real index, then this means that we can explain the L^2 risk function, namely the MISE (Mean Integrated Squared Error) by $\mathbb{E}\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)}$, where

$$\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)} := \frac{1}{n}\sum_{k=1}^{n} |\widehat{H}_n(t_k) - H(t_k)|^2$$

with $(t_k = \frac{k}{n})_{k=1,\dots,n}$ a family of observation times. Applying the previous CLT, we get (15):

$$2n\varepsilon_n \|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)} \quad \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \quad \frac{1}{n}\left[\sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2\right].$$

We can hence use it under the following form

$$\|\widehat{H}_n(t) - H(t)\|^2_{L^2(]0;1[)} \simeq_{CTL} \frac{1}{2n\varepsilon_n}\left[\frac{1}{n}\sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2\right].$$

Set V_n defined by (16)

$$V_n := \frac{1}{n}\sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2.$$

We here aim at proving that V_n satisfies a CTL.

Proof. Up to a multiplicative factor, it sufficies to prove the CLT for $\widetilde{V}_n$ defined by

$$\widetilde{V}_n := n \times V_n = \sum_{k=1}^{n} |\mathbb{G}'(t_k)|^2.$$

For this, we can apply the CLT proved in [1, Th 3.1]

$$\widetilde{V}_n \quad = \quad \widetilde{\mu}_n + \widetilde{S}_n \times \xi_n$$

and consequently Formula (17)

$$V_n \quad = \quad \mu_n + S_n \times \xi_n$$

where $\mu_n = \mathbb{E}(V_n)$ is the expected value, $S_n = \sqrt{\mathrm{Var}(V_n)}$ is the standard deviation, and the following convergence in distribution to a standard normal deviate (18) holds

$$\xi_n \quad \underset{n\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \quad \mathcal{N}(0;1).$$

Actually, as $\mathbb{G}'(t_k)$ are Gaussian centred random variables, we can apply a result from Istas-Lang (2007) [22]. A sufficient condition to get result (18) is

$$\lim_{n\longrightarrow\infty} \frac{\max_{k\in\{1,\dots,n\}}\left[\sum_{j=1}^{n} \mathrm{cov}\Big(\mathbb{G}'(t_k);\mathbb{G}'(t_j)\Big)\right]}{\sqrt{\mathrm{Var}(\widetilde{V}_n)}} = 0. \tag{36}$$

For $j \neq k$, if t_j and t_k where adequately located, namely at a sufficient distance each other (N, with $N \longrightarrow \infty$), we would have

$$\mathrm{cov}\Big(\mathbb{G}'(t_k);\mathbb{G}'(t_j)\Big) = 0. \tag{37}$$

Then, as $\mathbb{G}'(t_k)$ are Gaussian random variables with zero mean, we just keep

$$\sum_{j=1}^{n} \mathrm{cov}\Bigg(\mathbb{G}'(t_k);\mathbb{G}'(t_j)\Bigg) = \mathrm{Var}\Bigg(\mathbb{G}'(t_k)\Bigg) = E\Bigg(\mathbb{G}'(t_k)^2\Bigg).$$

Next, it comes that

$$\max_{k\in\{1,\ldots,n\}} \left[\sum_{j=1}^{n} \mathrm{cov}\Bigg(\mathbb{G}'(t_k);\mathbb{G}'(t_j)\Bigg)\right] = \max_{k\in\{1,\ldots,n\}} \left[\mathrm{Var}\Bigg(\mathbb{G}'(t_k)\Bigg)\right].$$

From Proposition (2.1), formula (11) gives us the expression of $\mathrm{Var}\Bigg(\mathbb{G}'(t_k)\Bigg)$

$$\mathrm{Var}\Bigg(\mathbb{G}'(t_k)\Bigg) = \gamma_{H(t_k)}.$$

On the other hand, by definition of $\widetilde{V}_n$ and by assumption (37) of independence of Gaussian random variables $\mathbb{G}'(t_k)$, we can write that

$$\mathrm{Var}(\widetilde{V}_n) = \sum_{k=1}^{n} \mathrm{Var}\Bigg(\mathbb{G}'(t_k)^2\Bigg). \tag{38}$$

As $\mathbb{G}'(t_k)$ is Gaussian, the variance of $\mathbb{G}'(t_k)^2$ can be explained in the following way

$$\begin{aligned}
\mathrm{Var}\Bigg(\mathbb{G}'(t_k)^2\Bigg) &= \mathbb{E}\left[\Bigg(\mathbb{G}'(t_k)^2 - \mathbb{E}(\mathbb{G}'(t_k)^2)\Bigg)^2\right] \\
&= \mathbb{E}\left[\mathbb{G}'(t_k)^4\right] - (\mathbb{E}[\mathbb{G}'(t_k)^2])^2 \\
&= 2 \times \left[\mathrm{Var}(\mathbb{G}'(t_k))\right]^2
\end{aligned}$$

and it comes that

$$\mathrm{Var}\Bigg(\mathbb{G}'(t_k)^2\Bigg) = 2 \times \Bigg(\gamma_{H(t_k)}\Bigg)^2. \tag{39}$$

Then sufficient condition (36) becomes

$$\lim_{n\longrightarrow\infty} \frac{\max_{k\in\{1,\ldots,n\}}\Bigg(\gamma_{H(t_k)}\Bigg)}{\sqrt{\sum_{k=1}^{n}\Bigg(2\times(\gamma_{H(t_k)})^2\Bigg)}} = 0$$

since

$$\frac{\max_{k\in\{1,\dots,n\}}\Big(\gamma_{H(t_k)}\Big)}{\sqrt{\sum_{k=1}^{n}\Big(2\times(\gamma_{H(t_k)})^2\Big)}} \le \frac{\max_{k\in\{1,\dots,n\}}\Big(\gamma_{H(t_k)}\Big)}{\sqrt{2n\times\min_{k\in\{1,\dots,n\}}\Big[\Big(\gamma_{H(t_k)}\Big)^2\Big]}}$$

so this condition is satisfied and CLT (Th.3.1 in [1]) holds. We hence have proved that CLT (17) holds. Then, by definition of V_n, we have

$$\mathbb{E}(V_n) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{Var}\Big(\mathbb{G}'(t_k)\Big) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big(\mathbb{G}'(t_k)^2\Big) = \frac{1}{n}\sum_{k=1}^{n}\gamma_{H(t_k)} \tag{40}$$

and

$$\mathrm{Var}(V_n) = \frac{1}{n^2}\mathrm{Var}(\widetilde{V}_n) = \frac{2}{n}\left[\frac{1}{n}\sum_{k=1}^{n}(\gamma_{H(t_k)})^2\right] \tag{41}$$

using respectively (38) and (39). From (40) we get that $\mathbb{E}(V_n) \longrightarrow \int_0^1 \gamma_{H(t)}dt$, and from (41), we get that $\left(\frac{n}{2}\right)\times\mathrm{Var}(V_n) \longrightarrow \int_0^1 (\gamma_{H(t)})^2 dt$ as $n\to\infty$. Consequently, set as in Formula (20):

$$T_n(\widehat{H}_n(t)) \quad = \quad \frac{2n\varepsilon_n\|\widehat{H}_n(t)-H(t)\|^2_{L^2(]0;1[)} - \int_0^1\gamma_{H(t)}dt}{\left(\frac{2}{n}\int_0^1(\gamma_{H(t)})^2dt\right)^{1/2}}.$$

Combining (15) and (18) we get that $T_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$ as $n\to\infty$. □