Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates

Article *in* Physica A Statistical Mechanics and its Applications · November 2016

DOI: 10.1016/j.physa.2017.04.122

CITATIONS
75

READS
1,989

1 author:

Matthieu Garcin
Pôle Universitaire Léonard de Vinci
44 PUBLICATIONS **305** CITATIONS

SEE PROFILE

Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates

Matthieu Garcin

Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates

Matthieu Garcin*

November 16, 2016

Abstract

Hurst exponents depict the long memory of a time series. For human-dependent phenomena, as in finance, this feature may vary in the time. It justifies modelling dynamics by multifractional Brownian motions, which are consistent with time-varying Hurst exponents. We improve the existing literature on estimating time-dependent Hurst exponents by proposing a smooth estimate obtained by variational calculus. This method is very general and not restricted to the sole Hurst framework. It is globally more accurate and easier than other existing non-parametric estimation techniques. Besides, in the field of Hurst exponents, it makes it possible to make forecasts based on the estimated multifractional Brownian motion. The application to high-frequency foreign exchange markets (GBP, CHF, SEK, USD, CAD, AUD, JPY, CNY and SGD, all against EUR) shows significantly good forecasts. When the Hurst exponent is higher than 0.5, what depicts a long-memory feature, the accuracy is higher.

Keywords – multifractional Brownian motion, Hurst exponent, Euler-Lagrange equation, non-parametric smoothing, foreign exchange forecast

1 Introduction

Many empirical works demonstrate that Gaussian processes do not fit well the log-price evolution of various financial assets [33, 35]. Among all the ways explored to get past the standard Brownian motion, the fractional Brownian motion (fBm) is quite popular. A fBm is characterized by a parameter $H \in (0,1)$, the Hurst exponent, whose value indicates different properties of the process:

- ▷ if $H = 1/2$, the fBm is identical to a standard Brownian motion,
- ▷ if $H > 1/2$, the increments are persistent,
- ▷ if $H < 1/2$, the increments are anti-persistent.

*Natixis Asset Management-LabEx ReFi, 21 quai d'Austerlitz, 75013 Paris, France. This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR10LABX0095. It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (investments for the future Paris New Worlds) under the reference ANR11IDEX000602.

As a consequence, if we estimate a Hurst exponent $H \neq 1/2$, then the evolution of the log-prices is in part predictable and arbitrages may occur [36, 22]. This is in contradiction with most of the financial literature, based on the efficient-market hypothesis, which precludes any arbitrage. However, introducing uncertainty in the Hurst exponent enables a higher consistency with the no-arbitrage theory. The Hurst exponent should therefore be time-dependent. The resulting process is then called a multifractional Brownian motion (mBm). The accurate estimation of the time-dependent Hurst exponent is challenging. Indeed, the most widespread estimation methods, based on a rolling window, have some limits. In particular, one often circumvents the high variance of estimators by expanding the size of the window or by smoothing the erratic estimates with a moving average as a post-processing. Both approaches are philosophically similar since they assign a higher weight to distant times. Whatever the approach, it introduces inaccuracy. For example, if the true time-dependent Hurst exponent reaches a maximum in time t, the important weight assigned to times far from t leads to an underestimate of $H(t)$. The purpose of this paper consists in presenting a proper way of smoothing a series of estimates so as to gain in accuracy, in comparison to a simple moving average, when estimating time-dependent Hurst exponents.

The smoothing method we propose can be applied to any raw series of a time-dependent statistic. It is based on variational calculus. More precisely, we obtain a non-parametric smooth estimate which minimizes the quadratic difference with the raw estimate as well as it minimizes the quadratic variations of the new estimate. This penalized estimation method is simple to implement with few methodological choices and leads to satisfactory results and properties. In particular, a simulation study shows a higher accuracy than the simple moving average with exponential weights, in the framework of mBm's.

The method we propose is close to some existing smoothing methods. In fact, the formulation of the problem as a trade-off between a fidelity term and a smoothness penalty is very common. For example, in the actuarial literature, it is used to smooth raw mortality data [24, 44]. This is known as the Whittaker-Henderson problem and was introduced to go beyond the limitations of the moving average near the boundaries of the sample. The actuarial solution provided to this problem is however quite different since it is based on numerical schemes with finite differences. The numerical resolution can therefore be iterative or based on a matrix inversion. Whatever the technique chosen, it is very different from our Euler-Lagrange framework and the resolution we propose, which seems computationally faster. However, our paper is not the first to handle this kind of penalized estimation method with the help of the Euler-Lagrange equation. Indeed, in the image processing literature, such a method is used to smooth images disrupted by some measurement noise [12, 43, 48]. But, in these papers, the quite more general framework imposes an iterative resolution. By contrast, we derive an explicit solution of a second-order ordinary differential equation, reducing the high-dimension problem to a standard linear problem of dimension two, which can be solved by a closed formula. It seems therefore that our variational smoothing method is new, in particular when applied to financial econometrics.

We also applied this technique to make predictions on foreign exchange rates. In the literature, the Hurst exponent is often only invoked as an indicator of long-memory. If authors find a long-memory feature, revealed by Hurst exponents higher than 0.5, they make some predictions with a model which does not specifically relate to the theory behind the Hurst exponent [36, 46]. On the contrary, the forecasts we make are directly derived from the mBm, in which the time-varying Hurst exponent plays a prominent role. By doing so, we obtain significantly good forecasts for all the nine exchange rates we give in example. The accuracy of the forecasts also increases when the Hurst exponent is higher than 0.5.

The rest of the paper is structured as follows. We first present the state of the art concerning Hurst exponents (section 2). We particularly focus on the possible ways to estimate them. Then, we present the post-processing smoothing step of the raw estimate (section 3): we confront the

variational smoothing to other standard smoothing techniques, such as the moving average or wavelet-based methods. Then, we give two examples in which we estimate time-varying Hurst exponents on simulated mBm's (section 4). This allows us comparing empirically the moving average and the variational smoothing. Finally, we present a forecasting procedure which we apply to the FX market (section 5).

2 Hurst exponents: state of the art

In this section, we recall some well-known facts about the fBm, the mBm and the identification of the Hurst exponent.

2.1 From the fractional to the multifractional Brownian motion

The fBm was introduced by Mandelbrot and van Ness [34], following works initiated by Kolmogorov [30]. A fBm B_H, of Hurst exponent H and volatility scale parameter σ^2, is defined as the only centred Gaussian process with zero at the origin and with the following covariance function, for $s, t \geq 0$:

$$\mathbb{E}\{B_H(t)B_H(s)\} = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}).$$

Increments of B_H of duration τ are thus stationary: they follow a centred Gaussian distribution of variance $\sigma^2|\tau|^{2H}$. If $H = 1/2$, it corresponds to the standard Brownian motion, whose increments have a variance $\sigma^2|\tau|$. It is worth noting that increments are positively correlated if $H > 1/2$ and negatively correlated if $H < 1/2$, while they are independent for $H = 1/2$. Indeed:

$$\begin{array}{rl} & Correl(B_H(n+1) - B_H(n), B_H(1) - B_H(0)) \\ = & [(n+1)^{2H} - 2n^{2H} + (n-1)^{2H}]/2 \\ = & \frac{H(2H-1)}{n^{2(1-H)}} + \underset{n\to\infty}{o}\left(\frac{1}{n^{2(1-H)}}\right). \end{array}$$

Moreover,

$$\sum_{n=1}^{N} Correl(B_H(n+1) - B_H(n), B_H(1) - B_H(0)) \overset{N\to+\infty}{\sim} -\frac{1}{2} + HN^{2H-1},$$

which diverges toward infinity when $N \to +\infty$, if and only if $H > 1/2$. This depicts the long-memory feature of the fBm with $H > 1/2$. Alternatively, the fBm can be seen as a fractional primitive function of a standard Brownian motion, if $H > 1/2$, or as a fractional derivative function of a standard Brownian motion, if $H < 1/2$.

The mBm is a generalization of the fBm to time-dependent Hurst exponents [39, 42]. In this model, the evolution of the Hurst exponent is smooth. In particular it is Hölderian of order $0 < \eta \leq 1$ [13]. However, some more general functions were introduced for H, for example with singularities, like the generalized multifractional Brownian motion [3], or even as a stochastic process [4]. In the present paper, we only focus on the smooth representation of the Hurst exponent function. For a mBm X, the covariance function becomes, for $s, t \in [0, 1]$ [13]:

$$\mathbb{E}\{X(t)X(s)\} = \frac{\sigma^2}{2} g(H(t), H(s))(|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)}),$$

where

$$\left\{ \begin{array}{l} g(H(t), H(s)) = K(H(t) + H(s))^{-1} \sqrt{K(2H(t))K(2H(s))} \\ K(\alpha) = \Gamma(1+\alpha)\sin(\alpha\pi/2)/\pi. \end{array} \right.$$

For s and t such that $H(t) = H(s)$, we have $g(H(t), H(s)) = 1$ and the covariance is equal to the one of a fBm.

The dependence of the increments makes the computation of a fBm or a mBm not as straightforward as the standard Brownian motion. For the mBm, two representations are commonly used [13]. The first one is based on weighted sums of the increments of a standard Brownian motion W [39]:

$$X(t) = \frac{\sigma\sqrt{\pi K(2H(t))}}{\Gamma\left(H(t)+\frac{1}{2}\right)}\left\{\int_{-\infty}^{0}\left[(t-s)^{H(t)-1/2} - (-s)^{H(t)-1/2}\right]dW(s) + \int_{0}^{t}(t-s)^{H(t)-1/2}dW(s)\right\}, \tag{1}$$

whereas the second one follows a spectral approach [6]:

$$X(t) = \frac{\sigma\sqrt{K(2H(t))}}{2}\int_{\mathbb{R}}\frac{\exp(\mathrm{i}t\lambda)-1}{|\lambda|^{H(t)+1/2}}dW(\lambda).$$

2.2 Estimating Hurst exponents

In the existing literature, the estimation of a local Hurst exponent for a mBm mainly relies on a sliding-window technique. The size of the window varies across the articles: for example, 25 days for daily financial data [10] or roughly 200 points for simulated data [13]. In each window, a static Hurst exponent is estimated. The choice of the estimator of a static Hurst exponent on an interval is thus crucial.

Several methods have been tested to estimate the Hurst exponent of a fBm, that is to say a global Hurst exponent. Among them are the rescaled range (R/S) analysis [27] and the detrended fluctuation analysis (DFA) [40]. They both consist in the regression of the amplitude of the fluctuations of the signal on given statistics computed on separate subsamples. While these methods are quite popular, they have some drawbacks in our perspective. They indeed need a big amount of data and no asymptotic theory is derived, so that one should use a bootstrap in order to obtain confidence intervals for the estimates [25, 45]. These methods thus seem more suitable for estimating global Hurst exponent than as a basis for computing local Hurst exponents.

Alternatively to the R/S analysis and to the DFA, one may use some estimators that are not based on a regression on subsamples. This group of estimators stems from the absolute moments of a Gaussian random variable [9, 10]. We assume we have N observations of a mBm X in the time interval $[0, 1]$. The observations are equally spaced in time, with a spacing equal to $1/N$. We compute the following statistic in a window of size $\varepsilon_N \leq N$, for any given $k \in \mathbb{N}$ [9]:

$$M_k(t) = \frac{1}{\varepsilon_N}\sum_{i=0}^{\varepsilon_N-1}|X(t-i/N) - X(t-(i+1)/N)|^k. \tag{2}$$

$M_k(t)$ is an estimate of $\mathbb{E}(|X(t) - X(t-1/N)|^k)$, which value is:

$$\mathbb{E}(|X(t) - X(t-1/N)|^k) = \frac{2^{k/2}\Gamma(\frac{k+1}{2})}{\Gamma(\frac{1}{2})}\sigma^k N^{-kH(t)}. \tag{3}$$

It thus leads to the following estimate of $H(t)$:

$$\check{H}(t) = -\frac{\log(\sqrt{\pi}M_k(t)/[2^{k/2}\Gamma(\frac{k+1}{2})\sigma^k])}{k\log(N)},$$

which converges towards $H(t)$ at the rate $O((\sqrt{\varepsilon_N}\log(N))^{-1})$, when N goes to infinity.[1] Before being generalised to higher moments, it was first used for $k=1$ [39]:

$$\check{H}(t) = -\frac{\log\left(\sqrt{\frac{\pi}{2}}\frac{M_1(t)}{\sigma}\right)}{\log(N)}.$$

Besides, the above moment-based estimator is for a given scale parameter, σ. However, in realistic applications, σ is unknown. To get rid of this parameter in the estimate of H, we can confront estimators at different resolutions. Let us focus on the most widespread choice, $k=2$ [8, 28, 29]. We compute the average quadratic variation $M_2(t)$ as in equation (2) as well as the same kind of statistic $M_2'(t)$ based on observations in the same window but with halved resolution:[2]

$$M_2'(t) = \frac{2}{\varepsilon_N}\sum_{i=0}^{\varepsilon_N/2-1} |X(t-2i/N) - X(t-2(i+1)/N)|^2.$$

$M_2(t)$ and $M_2'(t)$ are respectively equivalent to $2\sigma^2 N^{-2H(t)}\Gamma(\frac{3}{2})/\Gamma(\frac{1}{2})$ and $2\sigma^2(N/2)^{-2H(t)}\Gamma(\frac{3}{2})/\Gamma(\frac{1}{2})$, when N goes to infinity. Therefore, the ratio $M_2'(t)/M_2(t)$ converges towards $2^{2H(t)}$, so that we obtain a new estimate of H which does not depend on the unknown scale parameter σ:

$$\hat{H}(t) = \frac{1}{2}\log_2\left(\frac{M_2'(t)}{M_2(t)}\right). \tag{4}$$

$\hat{H}(t)$ converges almost surely toward $H(t)$ [42, 6].

Some refinements inspired by these moment-based estimators are possible:

- ▷ The estimate $\hat{H}(t)$ is based on the comparison of statistics computed at two different resolutions. Other resolutions could also be taken into account and the formula based on the ratio of statistics could then be replaced by half the slope of the linear regression of the logarithm of each statistic over the logarithm of the resolution [13]. However, such an approach would need a big amount of data so as to build significant statistics at lower resolutions.
- ▷ Similarly to $M_k(t)$, we can compute a statistic on filtered observations. More precisely, this statistic would be defined as an average of the quantities $|V_{t,i}^a(X)|^k$, for $0 \leq i \leq \varepsilon_N - 1$, in which $a \in \mathbb{R}^d$ is the filter and $V_{t,i}^a(X) = \sum_{j=1}^d a_j X(t-(i+j-1)/N)$. In particular, the filter applied in the case of $M_k(t)$ is $a=(1,-1)$. Other filters may be used [13]. For example, the filter $a=(1,-2,1)$ defines the general quadratic variation [28, 6, 13], for which the variations $|X(t-i/N) - X(t-(i+1)/N)|$ are replaced by the second-order variations $|X(t-i/N) - 2X(t-(i+1)/N) + X(t-(i+2)/N)|$. But the filter may also be more sophisticated. For example the filtered series could be its wavelet coefficient for a given wavelet function.

Beside the previously mentioned methods for estimating global Hurst exponents, other ways are possible, such as the convex combination of sample quantiles [14] or the increment ratio statistic [5, 7]. However, for simplicity, we will focus on one estimator, which is simple and widespread: the estimator based on quadratic variations at two scales, $\hat{H}(t)$.

[1] The window should be defined so that its length is asymptotically zero and the number of points it contains is asymptotically infinite: $\varepsilon_N \to 0$ and $N\varepsilon_N \to \infty$.

[2] For convenience, we impose ε_N to be even.

3 Smooth estimation procedure

The estimation techniques of the time-varying Hurst exponents, exposed in the previous section, often lead to erratic estimates. To overcome this, one may increase the size of the window of the estimator. This idea has some drawbacks, since it mitigates the local particularity with distant observations. Surprisingly, we did not find many papers in which the erratic feature of the estimates of the Hurst exponent is diminished by a smoothing. In some papers, when estimating the Hurst exponents in a window, a bigger weight is given to the closer past [37], exactly as when computing an exponentially-weighted volatility. This technique makes it possible to increase the size of the windows and thus to smooth the time-varying estimate of the Hurst exponent, while not giving too much weight to distant observations. In other papers, the raw time series of Hurst exponents built on a small window is smoothed in a second step. Indeed, the erratic feature of the series of estimated Hurst exponents is bypassed when analysing the results with the help of parametric techniques, such as linear regressions [1], or with non-parametric techniques, such as a moving average [23].

The mBm assumes a smooth evolution of the Hurst exponent. For this reason as well as for the ease of predictions, it thus seems important to smooth the series of raw and erratic estimated Hurst exponents. However, we prefer this smoothing not to be limited to an averaging but rather to respect local particularities as much as possible. For example, given a raw but already smooth series of Hurst exponents, presenting an extremum, a moving average will attenuate the extremum. We prefer a method that would leave the extremum almost unaltered while smoothing some erratic features. We thus propose a technique based on variational calculus, in which we intend to minimize the distance to the raw estimate as well as the cumulated quadratic variations of the smoothed estimate.

What we propose is thus a two-step procedure. First we estimate the Hurst exponent time series. It is a raw estimate which consists in a noisy version of the true unknown Hurst exponent time series. Then, we smooth this series of estimates and we thus obtain a new series of estimates. In this paper, we focus on the smoothing procedure, so that the raw Hurst exponent time series can be obtained by different ways. For the smoothing procedure, we propose a new method based on variational calculus and we discuss its strengths and weaknesses thereafter by comparing it with more traditional approaches.

3.1 Variational smoothing

We observe a stochastic process (X_t) in the time interval $[0, 1]$. We are interested in an unknown time-varying parameter[3] of this process, $H : [0, 1] \to \mathbb{R}$. In a first step, we compute an estimate $\hat{H}$ of the unknown function H. In this general outlook of the variational estimation procedure, the choice of the estimator is not overriding. We indeed focus on the smoothing layer, which constitutes the second step of the procedure. $\hat{H}$ can be considered as a noisy version of H. It can be very erratic, so that we would like to diminish these strong variations across time. In the second step, we compute a new estimate $\mathcal{H}$ of H. We aim at choosing $\mathcal{H}$ close to $\hat{H}$, so that we minimize the quadratic distance between $\mathcal{H}$ and $\hat{H}$. Moreover, we add a penalty term to this minimization problem so as to obtain a higher smoothness than for the first estimate $\hat{H}$. We thus intend to choose a function $\mathcal{H}$ which minimizes an integral:

$$\mathcal{H} = \underset{h \in \mathcal{C}^2([0,1],\mathbb{R})}{\operatorname{argmin}} \int_0^1 \left[(h(t) - \hat{H}(t))^2 + \lambda h'(t)^2 \right] dt, \tag{5}$$

[3] In this paper, we focus on the Hurst exponent, but our two-step estimation and smoothing method can be generalized to any other smooth time-varying parameter.

where h' is the derivative of a function h and $\lambda \geq 0$ is a parameter quantifying the desired smoothness. If $\lambda = 0$, then $\mathcal{H}$ is simply equal to $\hat{H}$ and no smoothing occurs. When increasing λ toward infinity, $\mathcal{H}$ should tend toward a constant function equal to the average of $\hat{H}$. The useful values of λ are in between. It must be adjusted carefully to balance accuracy and smoothness.

The problem stated in equation (5) can be solved in a non-parametric way thanks to the variational theory. Indeed, we present in the following theorem a formula for $\mathcal{H}$, in which only two unknown constants appear.

Theorem 1. *Let $\hat{H} : [0,1] \to \mathbb{R}$ be two times differentiable and $\lambda > 0$. Then, a solution $\mathcal{H} : [0,1] \to \mathbb{R}$ of equation (5) is necessarily of the form:*

$$\mathcal{H} : t \mapsto e^{t/\sqrt{\lambda}} \left(A - \frac{1}{2\sqrt{\lambda}} \int_0^t e^{-s/\sqrt{\lambda}} \hat{H}(s) ds \right) + e^{-t/\sqrt{\lambda}} \left(B + \frac{1}{2\sqrt{\lambda}} \int_0^t e^{s/\sqrt{\lambda}} \hat{H}(s) ds \right),$$

where $A, B \in \mathbb{R}$.

The proof of this theorem as well as of the following ones are reported in the appendix.

We notice the apparent contradiction between the statement that $\hat{H}$ may be very erratic and the assumption in Theorem 1 that it is also two times differentiable. In fact, we intend to use such a theorem in a very practical world, in which we observe the continuous process at discrete times. By a simple polynomial interpolation, every function observed in discrete time can be interpolated by an infinitely differentiable function, even if the discrete observations seem very erratic.

The smoothing of $\hat{H}$ thus works according to the following simple algorithm.

1. First, compute $\Phi\hat{H}$, where Φ is the operator such that, for $h \in \mathcal{C}^2([0,1], \mathbb{R})$ and $t \in [0,1]$:

$$\Phi h(t) = -\frac{e^{t/\sqrt{\lambda}}}{2\sqrt{\lambda}} \int_0^t e^{-s/\sqrt{\lambda}} h(s) ds + \frac{e^{-t/\sqrt{\lambda}}}{2\sqrt{\lambda}} \int_0^t e^{s/\sqrt{\lambda}} h(s) ds. \tag{6}$$

2. Then, project the residual in the functional space generated by $t \mapsto e^{t/\sqrt{\lambda}}$ and $t \mapsto e^{-t/\sqrt{\lambda}}$. The constants A and B in Theorem 1 can be obtained numerically, given λ. We can also use the particular structure of the explanatory variables to reduce the algorithmic complexity and get a closed-form solution. We present this method in Theorem 2.

Theorem 2. *Let $\hat{H} : [0,1] \to \mathbb{R}$ be two times differentiable and $\lambda > 0$. Then, the unique solution $\mathcal{H} : [0,1] \to \mathbb{R}$ of equation (5) such that $\int_0^1 (\mathcal{H}(t) - \hat{H}(t))^2 dt$ is minimal is:*

$$\mathcal{H} : t \mapsto \Phi\hat{H}(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}},$$

where Φ is the operator introduced in equation (6) and:

$$\begin{cases} \mathcal{G}(t) &= \hat{H}(t) - \Phi\hat{H}(t) \\ C &= \sqrt{\frac{2}{\sqrt{\lambda}(e^{2/\sqrt{\lambda}} - 1)}} \\ D &= C(e^{-2/\sqrt{\lambda}} - C^4)^{-1/2} \\ A &= C^2(1 + C^2D^2) \int_0^1 \mathcal{G}(s) e^{s/\sqrt{\lambda}} ds - C^2D^2 \int_0^1 \mathcal{G}(s) e^{-s/\sqrt{\lambda}} ds \\ B &= -C^2D^2 \int_0^1 \mathcal{G}(s) e^{s/\sqrt{\lambda}} ds + D^2 \int_0^1 \mathcal{G}(s) e^{-s/\sqrt{\lambda}} ds. \end{cases}$$

Theorem 2 provides a closed-form expression for the variational smooth estimator. However, in a practical perspective, for very low values of λ,[4] we recommend to optimize A and B numerically

[4] In the simulations and examples provided in this paper, we observe this behaviour for $\lambda < 0.02$.

instead of using the formulas provided in Theorem 2. Indeed, when λ is close to zero, the values of the exponentials in the expression of $\mathcal{H}$ diverge toward infinity, so that the computation of all the parameters as stated in Theorem 2 is inaccurate, because of the discrete approximations of the integrals. In such a case, the values obtained for A and B in Theorem 2 can be used to initiate the optimization procedure. Alternatively, in a discrete framework, we can use standard results about linear regression, as stated in Theorem 3.

Theorem 3. *Let $\hat{H}(t_1), ..., \hat{H}(t_n)$ be n statistics associated to times $t_1 = 0 < t_2 < ... < t_n = 1$, with $n > 1$. Then, the unique function $\mathcal{H} : \{t_0, ..., t_n\} \to \mathbb{R}$ such that $\sum_{i=1}^{n} (\mathcal{H}(t_i) - \hat{H}(t_i))^2 dt$ is minimal and of the form*

$$\mathcal{H} : t \mapsto \Phi^d \hat{H}(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}},$$

where Φ^d is the discretized version[5] of the operator introduced in equation (6) and $\lambda > 0$, is characterized by:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \frac{1}{\left(\sum_{i=1}^{n} e^{2t_i/\sqrt{\lambda}}\right)\left(\sum_{i=1}^{n} e^{-2t_i/\sqrt{\lambda}}\right) - n^2} \begin{pmatrix} \left(\sum_{i=1}^{n} e^{-2t_i/\sqrt{\lambda}}\right)\left(\sum_{i=1}^{n} e^{t_i/\sqrt{\lambda}}[1 - \Phi^d]\hat{H}(t_i)\right) - n\left(\sum_{i=1}^{n} e^{-t_i/\sqrt{\lambda}}[1 - \Phi^d]\hat{H}(t_i)\right) \\ \left(\sum_{i=1}^{n} e^{2t_i/\sqrt{\lambda}}\right)\left(\sum_{i=1}^{n} e^{-t_i/\sqrt{\lambda}}[1 - \Phi^d]\hat{H}(t_i)\right) - n\left(\sum_{i=1}^{n} e^{t_i/\sqrt{\lambda}}[1 - \Phi^d]\hat{H}(t_i)\right) \end{pmatrix}.$$

In Theorem 4, we show to which extent the variational smoothing method is accurate in the continuous setting. In particular, if we directly apply the smoothing procedure to the smooth function H to be estimated, we should obtain something close to H, especially when λ is close to zero. In this asymptotic case, Theorem 4 allows us understanding the link between λ and the error made when smoothing with our variational method.

Theorem 4. *Let Φ be the operator introduced in equation (6) and H an analytic function $[0, 1] \to \mathbb{R}$. Then, for $\lambda \to 0$, there exists $\xi \in [0, 1]$ such that:*

$$\min_{(A,B)\in\mathbb{R}^2} \int_0^1 \left(\Phi H(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}} - H(t)\right)^2 dt \leq 4\lambda^2 H''(\xi)^2 + o(\lambda^2).$$

In particular, if H is an affine function, the variational smoothing reproduces it without any error.[6] It is not the case for some other techniques such as the moving average. By the way, we present in a more empirical way the differences between the variational smoothing and the moving average in section 4.

3.2 Comparison with other smoothing methods

Beside the method we introduced above, several smoothing techniques exist. They are the core of non-parametric regressions [26]. Among them, we can cite averaging techniques and decompositions in orthogonal series, such as Fourier and wavelet series.

[5] For $i \leq n$, we can define it as:

$$\Phi^d \hat{H}(t_i) = -\frac{e^{t_i/\sqrt{\lambda}}}{2\sqrt{\lambda}} \sum_{j=0}^{i} e^{-t_j/\sqrt{\lambda}} \hat{H}(t_j) V_j + \frac{e^{-t_i/\sqrt{\lambda}}}{2\sqrt{\lambda}} \sum_{j=0}^{i} e^{t_j/\sqrt{\lambda}} \hat{H}(t_j) V_j,$$

where V_j is the Voronoi cell size of t_j in the grid $\{t_0, ..., t_n\}$, as presented in [20]. In this simple case, $t_j = (t_{j+1} - t_{j-1})/2$. Theorem 3 can be applied for equispaced grids but also for random observation times. If $t_0, ..., t_n$ are independent and identically distributed, we can quantify the inaccuracy of the discretized version of Φ as the maximum spacing, whose probability distribution can be approximated [19].

[6] This statement is a direct consequence of the proof of Theorem 4, which is reported in the appendix. Indeed, if H is an affine function such that $H(t) = a_0 + a_1 t$, then $\Phi H(t) = H(t) - Ae^{t/\sqrt{\lambda}} - Be^{-t/\sqrt{\lambda}}$, with $A = (a_0 + a_1\sqrt{\lambda})/2$ and $B = (a_0 - a_1\sqrt{\lambda})/2$.

The moving average is a rule of thumb insofar as it is not based on the minimization of a quadratic error. In addition, in a predictive perspective, it can introduce some significant lag since the last estimate is smoothed with more ancient observations. Therefore, we do not expect the highest accuracy for this technique. It is however easy to compute and depends on the only choice of an exponential weighting parameter.

The decomposition of a signal (a series of observations or statistics) in a series of orthogonal functions makes it then possible to smooth the signal. This can be achieved either by truncating the orthogonal series or by modifying those coefficients which are supposed to be largely disrupted by some noise. Wavelets series are often preferred to Fourier series since they describe more precisely local peculiarities. They have been widely used in image processing [32], but also for dynamical systems [17, 21], for estimating mBm's [8] or in finance [18, 41]. The use of this technique may seem painstaking since many technical choices must be made, which could generate as many outputs: mother wavelet, resolution level, denoising threshold, thresholding technique, etc. The computational cost is also high due to nested integrals. Moreover, wavelets cannot be used directly to make predictions. Border effects indeed have to be managed in a proper way [47]. For all these reasons, wavelet techniques can be considered as accurate but complicated.

In comparison, the variational smoothing appears to be both accurate and easy to use. It is based on the choice of only one parameter. We also highlight that the wavelet technique uses a projection of the signal into a functional space which is generated by numerous wavelet functions. On the contrary, the variational smoothing, as in Theorem 2, implies the projection in a space generated by only two functions. How is it possible to reduce as such the complexity of the decomposition of the signal? In fact, in the variational smoothing technique, the family of generating functions is adaptive. Contrary to the wavelet framework, we do not impose *ex ante* the family but we define a family of functions which depend on the signal. The easy use of the variational smoothing technique is also due to the very simple mathematical objects involved: sums and exponential functions.

We discuss empirically the accuracy of the variational smoothing in the next section as well as in section 5, dedicated to forecasting.

4 A simulation study

In this section, we compare the accuracy of the weighted moving average and of the variational smoothing for a large range of parameters. We exclude the wavelets from this analysis since they do not imply the choice of a sole parameter, contrary to both the moving average and the variational smoothing. The accuracy is determined by three metrics:

- ▷ The mean integrated squared error (MISE) depicts the average accuracy of the method:

$$R_0 = \frac{1}{N}\sum_{i=1}^{N}\left[\mathcal{H}\left(\frac{i}{N}\right) - H\left(\frac{i}{N}\right)\right]^2,$$

  for N observations in $[0,1]$, $\mathcal{H}$ being the smooth estimate of H.

- ▷ The average bias in $t=1$, that is on the right of the sample:

$$R_1 = \mathcal{H}(1) - H(1).$$

  It depicts the ability of the technique not to create artificial gaps at the border of the sample. This is an important feature in the predictive framework we develop further in this paper.

▷ The average quadratic slope, which determines to which extent the erratic estimate has been smoothed:

$$R_2 = \frac{1}{N-1}\sum_{i=2}^{N}\left[\frac{\mathcal{H}\left(\frac{i}{N}\right) - \mathcal{H}\left(\frac{i-1}{N}\right)}{1/N}\right]^2.$$

We now introduce the simulation examples. They consist in two samples. We first simulate a mBm on the time interval $[0,1]$ with an affine Hurst exponent, ranging from 0.1 to 0.9. The time interval is discretized in 200 points. The raw estimate of the Hurst exponent for each of these 200 dates is computed on a window of 25 observations. We then obtain the smoothed Hurst exponents by two methods: a variational smoothing of parameter 0.02 and an exponentially weighted moving average of parameter 0.04. Both parameters have been chosen so as to produce an overall good accuracy for each method. We thus obtain the estimates represented in Figure 1. We also represent the three metrics R_0, R_1 and R_2 for 100 simulated trajectories in Figure 2 in a more general framework, that is for different choices of parameter for both techniques.

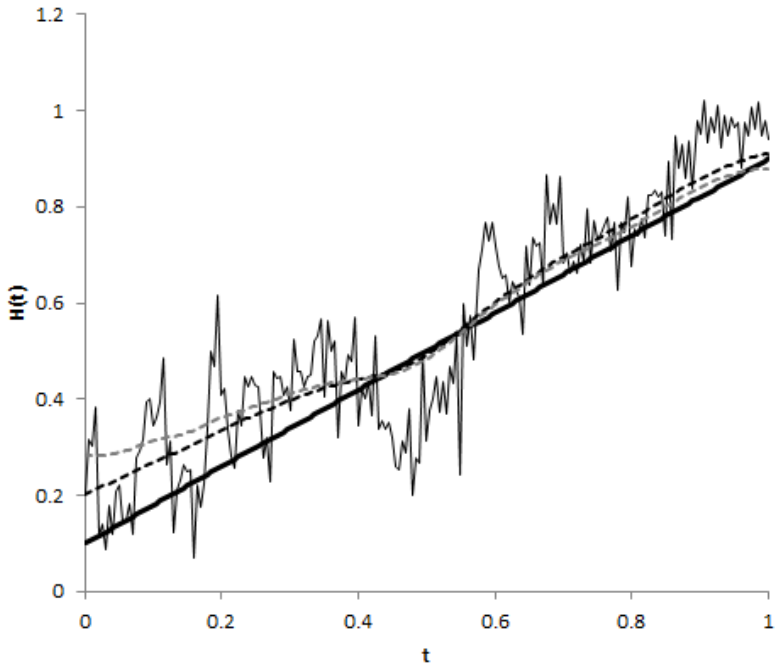

Figure 1: Theoretical affine Hurst exponent (thick line), raw estimate based on a simulated mBm (thin line), its variational smoothing (black dotted line) and its moving average (grey dotted line).

The second example follows the same method. The only difference in the input is the theoretical Hurst exponent, which is an exponential function in $[0,1]$ ranging from 0.2 to 0.8:

$$t \mapsto 0.2 + 0.6\frac{\exp(t^2)-1}{\exp(1)-1}.$$

As for the affine Hurst exponent, we represent the estimates for one simulated trajectory (Figure 3) and the average accuracy metrics for 100 trajectories for various choices of parameter for both methods (Figure 4).

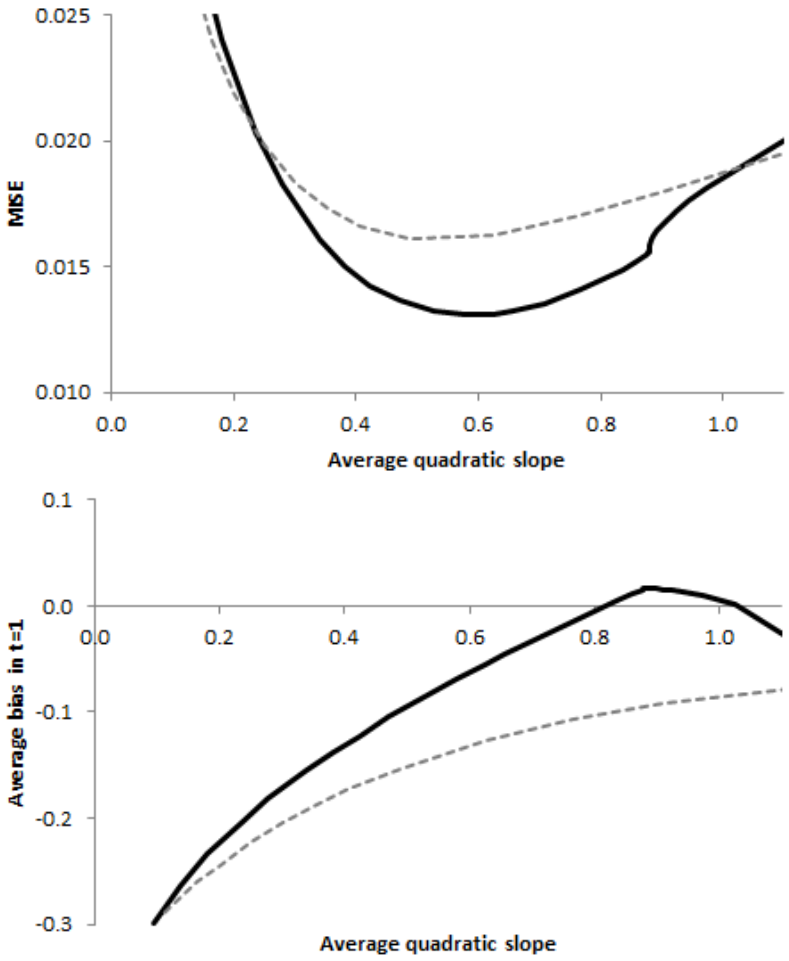

Figure 2: R_0 *vs* R_2 and R_1 *vs* R_2 for the affine Hurst exponent. The black line is for the variational smoothing, for a parameter between 0.02 (corresponding to the highest quadratic slope) and 280 (corresponding to the lowest quadratic slope). The grey dotted line is for the moving average, for a parameter between 0.01 (corresponding to the lowest quadratic slope) and 0.05 (corresponding to the highest quadratic slope). The average quadratic slope of the theoretical Hurst exponent is 0.64.

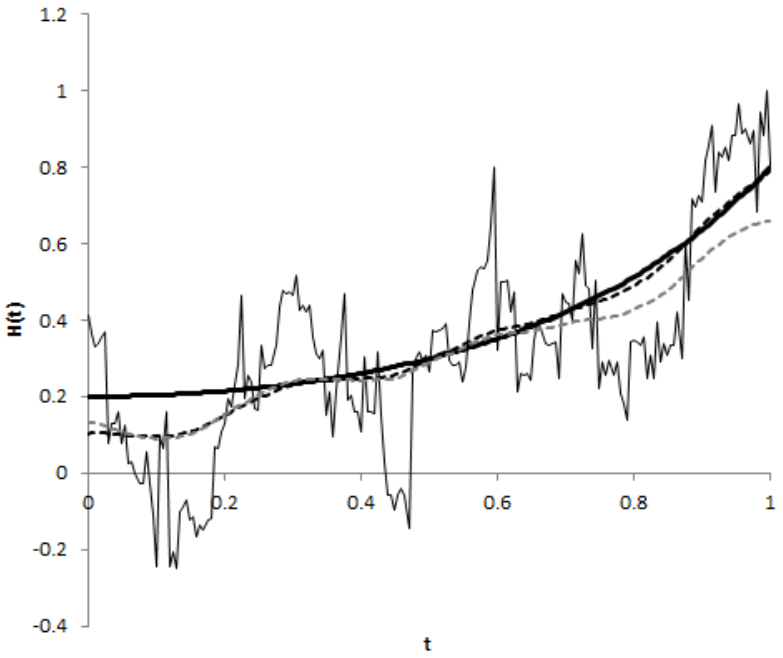

Figure 3: Theoretical exponential Hurst exponent (thick line), raw estimate based on a simulated mBm (thin line), its variational smoothing (black dotted line) and its moving average (grey dotted line).

In the two trajectories in Figures 1 and 3, we observe that the variational technique smooths more accurately the local raw Hurst exponents. For instance, in Figure 1, the moving average is biased by the increasing Hurst exponents, so that the raw curve of Hurst exponents is overestimated for the extremum at low t, whereas high Hurst exponents for the extremum at high t are underestimated. It is due to the fact that the moving average mitigates the local raw estimates with neighbours, which are higher in the first case and lower in the second case. Thus, despite the quite adequate averaging of the raw estimate by the moving average in the middle of the sample, border effects are significant. We also observe this phenomenon in Figure 3 for high t, but not that much for low t. It is not fortuitous: the slow variation of the exponential Hurst function for low t limits the distortion.

The superior accuracy of the variational smoothing over the moving average can also be noted while comparing average errors. In Figures 2 and 4, the choice of the parameter for each method makes it possible to tune monotonically the average smoothness R_2 of the estimate: the higher (respectively lower) the parameter, the smoother the variational (resp. moving-average) estimate. In general, in the affine and exponential examples, for a given value of R_2, the variational smoothing makes it possible to reach a lower MISE (R_0) and a bias in $t = 1$ (R_1) much closer to zero than does the moving average. The difference of the lowest MISE reached by each method is also significant and it is always in favour of the variational smoothing: 0.013 *versus* 0.016 in the affine case and 0.017 *versus* 0.019 in the exponential. By the way, in both cases, if we want to minimize the bias in $t = 1$, we are inclined to give up a low average MISE for the moving average, whereas an acceptable balance between R_0 and R_1 seems more achievable for the variational smoothing.

One could ask why the average bias is almost always negative in the examples provided in Figures 2 and 4. This underestimate of $H(1)$, whatever the parameter, is inherent in the growing feature of the theoretical Hurst exponent in the examples. For the parameters illustrated in both figures,

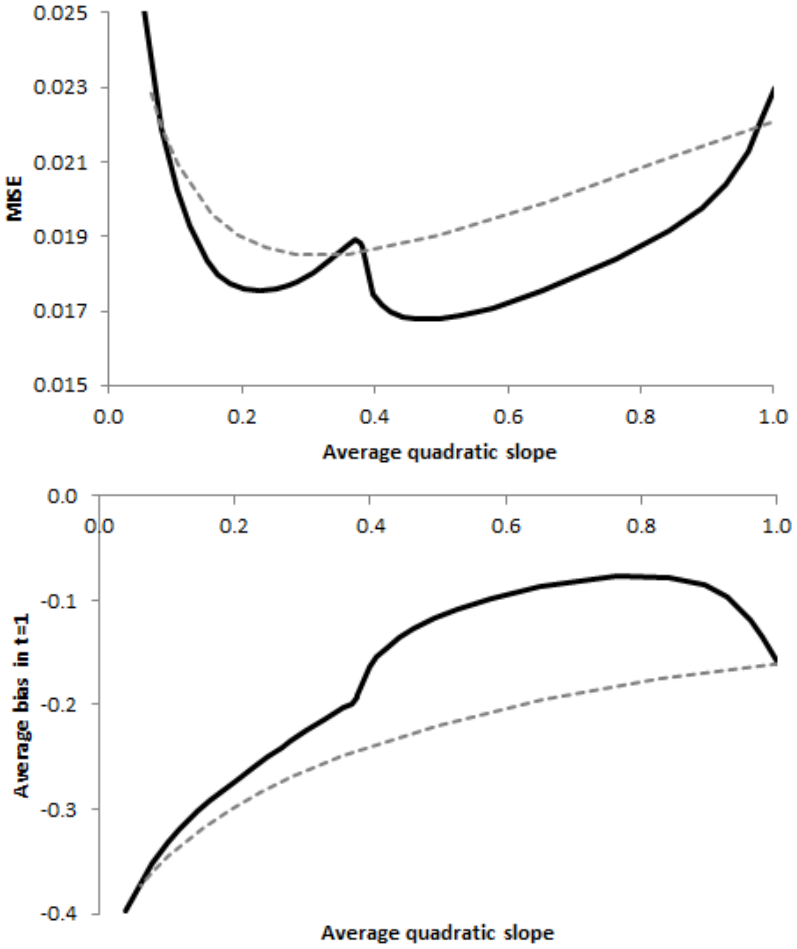

Figure 4: R_0 *vs* R_2 and R_1 *vs* R_2 for the exponential Hurst exponent. The black line is for the variational smoothing, for a parameter between 0.02 (corresponding to the highest quadratic slope) and 280 (corresponding to the lowest quadratic slope). The grey dotted line is for the moving average, for a parameter between 0.01 (corresponding to the lowest quadratic slope) and 0.05 (corresponding to the highest quadratic slope). The average quadratic slope of the theoretical Hurst exponent is 0.61.

the bias is anyway always closer to zero for the variational smoothing and even reaches zero and positive values for some parameters in the affine case. However, when increasing the parameter of the moving average, the bias monotonically reduces and it asymptotically reaches zero. By the way, it also reaches a very high MISE, since it asymptotically converges toward the raw series.

If we focus on the accuracy of the variational smoothing, the shape of the curves in Figures 2 and 4 shows a striking critical point. In fact, this point corresponds to a smoothing parameter which leads to an affine approximation of the raw series. For lower parameters and thus for a higher R_2, the approximation is closer to the raw series. Higher parameters and thus lower R_2 correspond to a more degenerate situation in which increasing the smoothness of the estimate go farther than a simple affine approximation. The slope of the estimate indeed slowly decreases up to reach a constant approximation of the raw estimate.

5 Forecasting prices of financial assets using Hurst exponents

We now use our variational smoothing method for estimating the Hurst exponents of a mBm in a predictive perspective. We first present the forecasting procedure and then an application to financial asset prices and more precisely to foreign exchange rates.

5.1 Forecasting procedure

Based on equation (1), which defines the mBm, we introduce a forecasting procedure. Let X be the log-price of a financial asset. We make the assumption that it is driven by a mBm. We will use the dependence feature of the log-returns to make some predictions. More precisely, from the time series X, we first estimate raw Hurst exponents, whose calculation, for each time t, takes into account a given number of preceding observations of X. We thus have a series $\hat{H}$, defined for the equispaced times 0, δ, ..., $T\delta$. We choose $\tau < T$, which defines the length of a sample on which we smooth $\hat{H}$. For $t \in \{0, ..., T-\tau\}$, we make successive out-of-sample predictions for $X((t+\tau)\delta)$, based on the sole knowledge of $(X(s\delta))_{s \leq t+\tau-1}$. It is done as in the following algorithm, in which a one-step prediction is made in $t+\tau-1$ for the time $t+\tau$, with a given t:

1. Compute the smooth version $(\mathcal{H}(s\delta))_{t \leq s \leq t+\tau-1}$ of the raw estimate $(\hat{H}(s\delta))_{t \leq s \leq t+\tau-1}$ of the Hurst exponents, as exposed in section 3.1.

2. Forecast the Hurst exponent at time $(t+\tau)\delta$. The forecasted value, $\mathcal{H}((t+\tau)\delta)$, can be obtained by a $\mathcal{C}^2$-continuation of $(\mathcal{H}(s\delta))_{t \leq s \leq t+\tau-1}$:

$$\mathcal{H}((t+\tau)\delta) = \frac{5}{2}\mathcal{H}((t+\tau-1)\delta) - 2\mathcal{H}((t+\tau-2)\delta) + \frac{1}{2}\mathcal{H}((t+\tau-3)\delta).$$

3. Estimate the scale parameter σ using the definition of the total variation of a mBm, stemming from equation (3) with $k = 1$. The estimate is noted $\hat{\sigma}$ and uses the estimate of the Hurst exponents in the time interval $[t\delta, (t+\tau-1)\delta]$:

$$\hat{\sigma} = \tau^{-1}\sqrt{\frac{\pi}{2}}\sum_{s=t}^{t+\tau-1}\frac{|X(s\delta) - X((s-1)\delta)|}{\delta^{\mathcal{H}(s\delta)}}.$$

4. Estimate the underlying standard Brownian motion W as exposed in the definition of the mBm in equation (1). This estimate, $\hat{W}$, uses $\hat{\sigma}$ and the estimate of the Hurst exponents in the time interval $[t\delta, (t+\tau-1)\delta]$. As at the origin t of the time interval the mBm should be 0, we translate X by $X(t\delta)$. We thus compute it iteratively for $s = t$ to $s = t+\tau-1$:

$$\hat{W}(s\delta) = \hat{W}((s-1)\delta) + (X(s\delta) - X(t\delta))\Pi(s\delta) - \mathbf{1}_{s>t} \sum_{u=t}^{s-1} [\hat{W}(u\delta) - \hat{W}((u-1)\delta)](s-u+1)^{\mathcal{H}(s\delta)-1/2},$$

where $\mathbf{1}_{s>t}$ is the indicator function, whose value is 1 if $s > t$ and 0 else, and:

$$\Pi(s\delta) = \frac{\Gamma(\mathcal{H}(s\delta) + \frac{1}{2})}{\hat{\sigma}\delta^{\mathcal{H}(s\delta)-1/2}\sqrt{\Gamma(2\mathcal{H}(s\delta)+1)\sin(\pi\mathcal{H}(s\delta))}}.$$

Since only the increments of W are of interest in our method, we choose $\hat{W}((t-1)\delta) = 0$.

5. Compute the forecasted log-return[7] $r((t+\tau)\delta)$:

$$r((t+\tau)\delta) = \sum_{s=t}^{t+\tau-1} \left[\hat{W}(s\delta) - \hat{W}((s-1)\delta)\right] \left[\frac{(t+\tau-s+1)^{\mathcal{H}((t+\tau)\delta)-1/2}}{\Pi((t+\tau)\delta)} - \frac{(t+\tau-s)^{\mathcal{H}((t+\tau-1)\delta)-1/2}}{\Pi((t+\tau-1)\delta)}\right].$$

This value is the expected log-return in $t+\tau$, given the estimates of all the useful terms indexed by times $s < t+\tau$, as well as the forecast $\mathcal{H}((t+\tau)\delta)$. It is equivalent to replacing $\hat{W}((t+\tau)\delta)$ by its expectation at time $(t+\tau-1)\delta$, that is to say $\hat{W}((t+\tau-1)\delta)$.

All the estimates $\mathcal{H}$, $\hat{\sigma}$ and $\hat{W}$ are specific to the time interval used for the forecast in $t+\tau$. Therefore, it must be estimated again for the next forecast in $t+\tau+1$ and so on until reaching T. We could thus have indexed these estimates to stress the fact that they are specific to the sample $[t, t+\tau-1]$ and to avoid any ambiguity. However, we have decided not to overload the equations.

To be consistent with the framework of the previous sections, in which mBm's were observed in the time interval $[0, 1]$, we choose $\delta = (\tau-1)^{-1}$.

5.2 Results

We apply the forecasting procedure we have set out on nine foreign-exchange rates *versus* Euro: the British pound (GBP), the Swiss franc (CHF), the Swedish krona (SEK), the United States dollar (USD), the Canadian dollar (CAD), the Australian dollar (AUD), the Japanese yen (JPY), the Chinese yuan (CNY) and the Singapore dollar (SGD). We focus on a high-frequency framework, in which prices are observed every 15 minutes.

This paper is not the first to deal with the analysis of Hurst exponents of foreign exchange rates. But existing papers on the subject focus on the estimate of a piecewise fBm for different market phases without smoothing [2, 31], on describing the long-memory feature of FX time series [15, 46], on estimating a mBm with random Hurst exponents [11] or on the multifractal spectrum [16, 38]. To our knowledge, our paper is the first to deal with predictions for FX markets deduced from the estimation of a mBm. In the litterature, mBm have been estimated on other types of financial time series, mainly on equity indexes [8, 10]. The value of Hurst exponents has also already been confronted to the performance of a simple strategy [36]: this strategy has not been based on forecasting Hurst exponents but it has revealed gains which increase with the value of the Hurst exponent. It is consistent with the findings we develop in this section.

[7] Forecasting the log-price follows the same method.

We now detail the data we use for this example. We observe each of the 9 foreign-exchange rates between the 7th March 2016 and the 7th September 2016 every 15 minutes of open market. The exchange rate considered is the average between the highest and the lowest rate in the 15-minute time interval. This makes roughly 12,930 observations for each exchange rate, with the notable exception of the CNY/EUR pair, for which we only have 7,422 observations. Then, for each series, we compute a raw estimate of the Hurst exponent using 25-observation windows. The estimate follows equation (4) and is only based on preceding observations so as not to introduce unfairness in the forecasting procedure. We then smooth this raw series of Hurst exponents on a sample of 50 consecutive raw estimates, which represent 12.5 hours. Variations of the Hurst exponent can be significant in such a small interval as one can see, for example, in Figure 5. This remark is a strong justification for the analysis of financial time series as mBm's rather than the more popular piecewise fBm's. With the help of the prediction method presented in section 5.1, we make a forecast for the exchange rate 15 minutes after the end of the smoothing sample. We repeat this procedure until the end of the sample, using overlapping windows of 50 raw estimates.

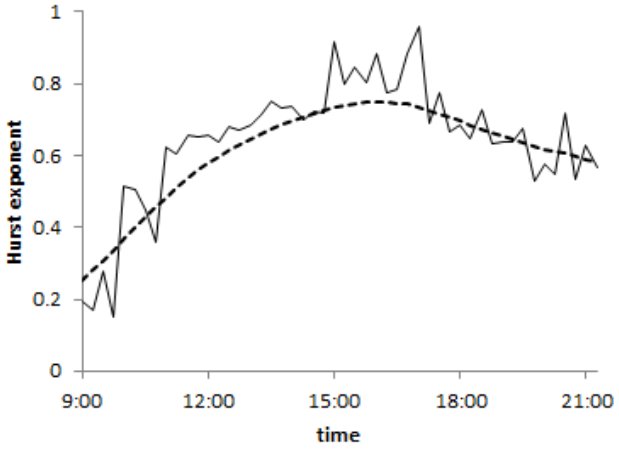

Figure 5: Raw Hurst exponent estimate (solid line) and smoothed estimate (dotted line) for the SEK/EUR the 21st March 2016 between 9:00 and 21:15.

To assess whether the forecasting procedure is efficient or not, we calculate a hit ratio: we count the proportion of times that the forecasted exchange rate evolution has the same sign than the real one. We exclude the few observations for which the real return is equal to 0. We report the results in Table 1. Be observe a hit ratio above 50% for all the currencies. We compare it with the hit ratio obtained if we replace the smoothed estimate by the raw series in the forecasting procedure. The hit ratio for the raw series is always smaller and is sometimes lower than 50%.

However, on could argue that the forecasting procedure using the smoothed estimate of the Hurst exponent only reaches 52.7%, which is not far from 50%. We have three answers to this remark. The first one focus on the significance of the hit ratio. As mentioned in the presentation of the data, we work on huge samples, so that the hit ratios obtained for the forecasting procedure are very significantly better than the 50% of a simple coin toss. We legitimate this assertion with a two-tailed binomial test, whose null assumption is a hit ratio of 50%. The results are also reported in Table 1. The only exchange rate whose p-value is greater than 1% is the CNY/EUR for the variational smoothing,[8] whereas the procedure based on the raw Hurst exponent never obtain p-values lower than 1%.

[8] For the CNY/EUR exchange rate, only 7,422 observations are provided in the time interval.

	Raw Hurst exponents	Variational smoothing
GBP/EUR	50.5% (2.5×10^{-1})	52.4% (5.8×10^{-8})
CHF/EUR	50.5% (2.7×10^{-1})	51.6% (3.2×10^{-4})
SEK/EUR	51.1% (1.6×10^{-2})	52.7% (2.6×10^{-9})
USD/EUR	50.0% (1.0)	52.5% (3.5×10^{-8})
CAD/EUR	50.7% (1.4×10^{-1})	52.0% (5.4×10^{-6})
AUD/EUR	50.5% (3.1×10^{-1})	52.4% (1.4×10^{-7})
JPY/EUR	50.7% (1.2×10^{-1})	52.2% (6.5×10^{-7})
CNY/EUR	49.7% (5.9×10^{-1})	51.1% (7.3×10^{-2})
SGD/EUR	51.0% (2.6×10^{-2})	51.3% (2.9×10^{-3})

Table 1: Overall proportion of good forecasts. In parenthesis, the p-value of the test of equality with Bernoulli variables of parameter 50%.

The second explanation for the limited performance of the forecasting procedure is linked to the fact that some predictions are more reliable than others. We represent in Figure 6 the hit ratio conditionally to the local smooth Hurst exponent. For Hurst exponents higher than 0.5, the hit ratio is high, always larger than 50%, whatever the currency. It goes up to 56%. On the contrary, the hit ratio is varying a lot for Hurst exponents lower than 0.5 and we cannot assert with confidence that forecasts are good or are wrong. This point is interesting for practitioners: if one estimates a Hurst exponent greater than 0.5, one can make some prediction. Otherwise, a low Hurst exponent would lead to unreliable forecasts. This finding is consistent with the results reported in the literature [36]. Since Hurst exponents greater than 0.5 correspond to a long-memory feature, contrary to lower Hurst exponents, the better performance of the forecasting procedure in that case is quite intuitive and we have here an empirical evidence of this idea. Nevertheless, we observe for half the exchange rates that the hit ratio slightly decreases for very high Hurst exponents (approximatively between 0.9 and 1). We can provide an explanation for this phenomenon. Such very high Hurst exponents indeed correspond to a very long memory of the process, so that any reliable forecast must be based on many observations in the past. We suppose that the windows used in the forecasting procedure are then too narrow to provide reliable predictions.

The way the last increment $(\hat{W}((t+\tau)\delta) - \hat{W}((t+\tau-1)\delta))$ of the standard Brownian motion is treated is the third and last source of inaccuracy in the forecasting procedure. Indeed, we decided to replace this increment by its expected value, which is zero. However, this increment is a random variable which can be either positive or negative. Therefore, assuming a perfect estimation of the dynamic, this last increment is uncertain and its realized value could lead the realized return away from our forecast.

6 Conclusion

We have presented a method to smooth a raw estimate of any time-dependent variable. This method offers an improvement to existing techniques. The moving average is indeed not as accurate on borders and the wavelet-based techniques are more intricate due to an important dependence on parameter choices. On the contrary, the variational smoothing is very easy to implement and one only has to choose a penalizing parameter. The application to estimating time-varying Hurst exponents and making some predictions on the FX market is quite promising, in particular for Hurst exponents higher than 0.5. For this purpose, we have proposed a forecasting method which is based on the estimated dynamics of a mBm. Further works would consist in applying the

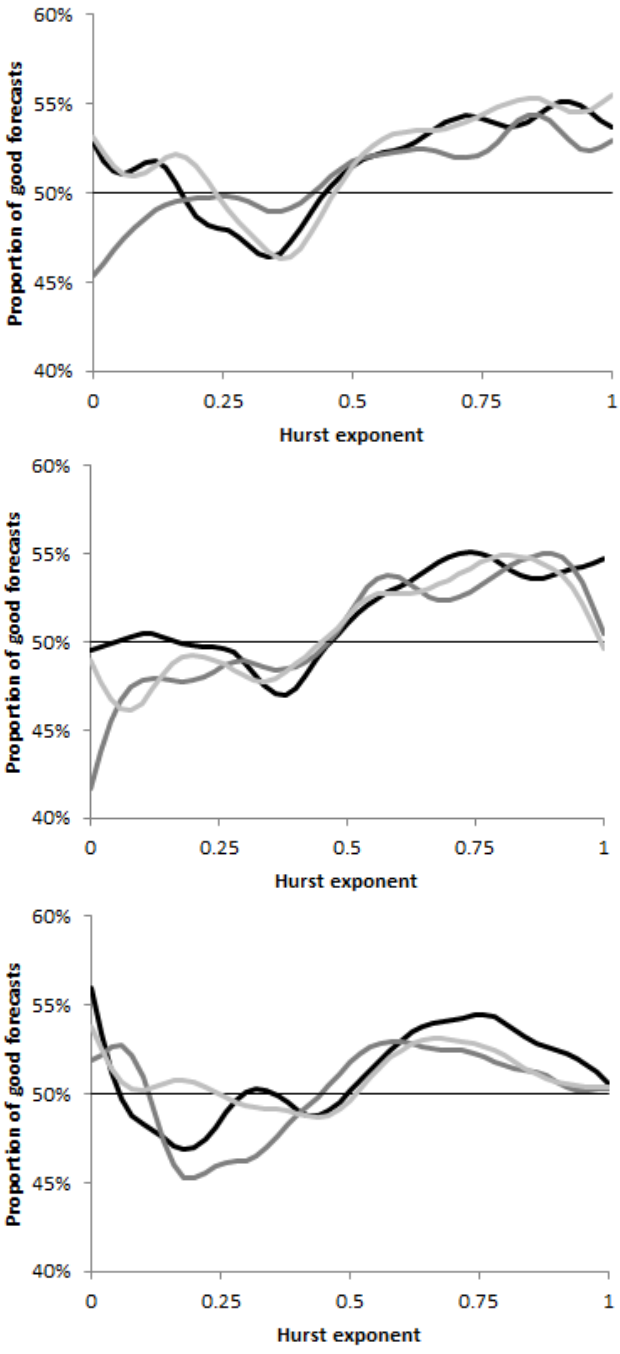

Figure 6: Frequency (smoothed with a Gaussian kernel) with which the model correctly forecasts the sign of the return, as a function of the local Hurst exponent. Top: GBP/EUR (black), CHF/EUR (dark grey), SEK/EUR (light grey). Middle: USD/EUR (black), CAD/EUR (dark grey), AUD/EUR (light grey). Bottom: JPY/EUR (black), CNY/EUR (dark grey), SGD/EUR (light grey).

variational smoothing technique to other statistics than the Hurst exponents, or to use it on time series analysis to have a more accurate description of mBm's and long-memory features.

A Proof of Theorem 1

Proof. The equation (5) can be written:

$$\mathcal{H} = \underset{h \in \mathcal{C}^2([0,1],\mathbb{R})}{\operatorname{argmin}} \int_0^1 \mathcal{L}\left(t, h(t), h'(t)\right) dt,$$

where

$$\mathcal{L}\left(t, h(t), h'(t)\right) = \left[(h(t) - \hat{H}(t))^2 + \lambda h'(t)^2\right]. \tag{7}$$

Since $\hat{H}$ is assumed to be two times differentiable, then $\mathcal{L}$ is also two times differentiable, as well as the unknown function h, by assumption. We can therefore apply the Euler-Lagrange equation, which states that the function $\mathcal{H}$ which minimizes the equation (5) is such that:

$$\frac{\partial}{\partial h}\mathcal{L}\left(t, \mathcal{H}(t), \mathcal{H}'(t)\right) = \frac{d}{dt}\frac{\partial}{\partial h'}\mathcal{L}\left(t, \mathcal{H}(t), \mathcal{H}'(t)\right). \tag{8}$$

Using together equations (7) and (8), we get, for $\lambda > 0$:

$$\mathcal{H}''(t) - \frac{\mathcal{H}(t)}{\lambda} = -\frac{\hat{H}(t)}{\lambda}. \tag{9}$$

It is easy to solve the linear differential equation (9). It can be done for example by variation of parameters. The functions $v_1 : t \mapsto \exp(t/\sqrt{\lambda})$ and $v_2 : t \mapsto \exp(-t/\sqrt{\lambda})$ are two independent solutions of the homogeneous equation corresponding to equation (9). Their Wronskian, $W(v_1, v_2) = v_1 v_2' - v_1' v_2$, is simply a constant function equal to $-2/\sqrt{\lambda}$. Therefore, any solution to the non-homogeneous equation (9) can be written:

$$\mathcal{H}(t) = v_1(t)\left(A - \int_0^t \frac{v_2(s)}{W(v_1,v_2)(s)} \times \left(-\frac{\hat{H}(s)}{\lambda}\right) ds\right) + v_2(t)\left(B + \int_0^t \frac{v_1(s)}{W(v_1,v_2)(s)} \times \left(-\frac{\hat{H}(s)}{\lambda}\right) ds\right),$$

with $A, B \in \mathbb{R}$. This last equation directly leads to Theorem 1. □

B Proof of Theorem 2

Proof. Any solution of equation (5) is provided by Theorem 1:

$$\mathcal{H} : t \mapsto \Phi\hat{H}(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}},$$

where the parameters A and B can be adjusted to fit the solution with the *initial* condition. In our case, the condition to fit with is the minimal quadratic distance between $\mathcal{H}$ and $\hat{H}$. In other words, parameters A and B are given by the orthogonal projection of $\mathcal{G} = \hat{H} - \Phi\hat{H}$ in the two-dimensional functional space Π, with a $L^2([0,1])$ norm, generated by:

$$\begin{cases} v_1 : t \mapsto e^{t/\sqrt{\lambda}} \\ v_2 : t \mapsto e^{-t/\sqrt{\lambda}}. \end{cases}$$

Therefore,

$$P_\Pi(\mathcal{G}) = Av_1 + Bv_2. \tag{10}$$

From (v_1, v_2), we create an orthonormal set of vectors (w_1, w_2), generating the same space than (v_1, v_2), thanks to the Gram-Schmidt method. We thus set, for $\langle ., . \rangle$ and $\|.\|$ the scalar product and the norm in $L^2([0,1])$:

$$w_1(t) = v_1(t)/\|v_1\| = Cv_1(t),$$

where $C = \sqrt{\frac{2}{\sqrt{\lambda}(e^{2/\sqrt{\lambda}}-1)}}$, and

$$\begin{aligned} w_2(t) &= \frac{v_2(t)-\langle v_2,w_1\rangle w_1(t)}{\|v_2-\langle v_2,w_1\rangle w_1\|} \\ &= \frac{v_2(t)-C^2 v_1(t)}{\sqrt{\|v_2\|^2+C^4\|v_1\|^2-2C^2\langle v_2,v_1\rangle}} \\ &= \frac{v_2(t)-C^2 v_1(t)}{\sqrt{e^{-2/\sqrt{\lambda}}C^{-2}+C^4C^{-2}-2C^2}} \\ &= \frac{Cv_2(t)-C^3 v_1(t)}{\sqrt{e^{-2/\sqrt{\lambda}}-C^4}} \\ &= Dv_2(t) - C^2 Dv_1(t), \end{aligned}$$

where $D = C(e^{-2/\sqrt{\lambda}} - C^4)^{-1/2}$. We then project $\mathcal{G}$ in the space Π generated by the pair (w_1, w_2). We write the projection $P_\Pi(\mathcal{G}) = c_1 w_1 + c_2 w_2$. By definition of the projection, c_1 and c_2 are given by:

$$c_1 = \langle \mathcal{G}, w_1 \rangle = C\langle \mathcal{G}, v_1 \rangle$$

and

$$c_2 = \langle \mathcal{G}, w_2 \rangle = D\langle \mathcal{G}, v_2 \rangle - C^2 D\langle \mathcal{G}, v_1 \rangle.$$

Thanks to equation (10), we get

$$Av_1 + Bv_2 = c_1 w_1 + c_2 w_2 = (Cc_1 - C^2 Dc_2)v_1 + Dc_2 v_2.$$

Since v_1 and v_2 are independent, we obtain the following system:

$$\begin{cases} A = Cc_1 - C^2 Dc_2 \\ B = Dc_2, \end{cases}$$

which directly leads to the solution given in Theorem 2 □

C Proof of Theorem 3

In the following proof, the notation M' is the transpose of a matrix M. The framework is here linear algebra, so that no confusion should arise with the similar notation f' used in other parts of the paper for the derivative of a function f.

Proof. The problem stated in Theorem 3 is an ordinary least-square problem of the form $Y = Xa + u$, with $Y = ([1 - \Phi^d]\hat{H}(t_1), ..., [1 - \Phi^d]\hat{H}(t_n))'$, $a = (A, B)'$, u a residual and

$$X = \begin{pmatrix} e^{t_1/\sqrt{\lambda}} & e^{-t_1/\sqrt{\lambda}} \\ \vdots & \vdots \\ e^{t_n/\sqrt{\lambda}} & e^{-t_n/\sqrt{\lambda}} \end{pmatrix},$$

whose solution is provided by the estimate $\hat{a} = (X'X)^{-1}X'Y$. In particular,

$$X'X = \begin{pmatrix} \sum_{i=1}^n e^{2t_i/\sqrt{\lambda}} & n \\ n & \sum_{i=1}^n e^{-2t_i/\sqrt{\lambda}} \end{pmatrix},$$

$$det(X'X) = \left(\sum_{i=1}^n e^{2t_i/\sqrt{\lambda}}\right)\left(\sum_{i=1}^n e^{-2t_i/\sqrt{\lambda}}\right) - n^2 \neq 0,$$

for any finite λ and $n > 1$, and

$$(X'X)^{-1}X'Y = \frac{1}{det(X'X)} \begin{pmatrix} \sum_{i=1}^n e^{-2t_i/\sqrt{\lambda}} & -n \\ -n & \sum_{i=1}^n e^{2t_i/\sqrt{\lambda}} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n e^{t_i/\sqrt{\lambda}}[1-\Phi^d]\hat{H}(t_i) \\ \sum_{i=1}^n e^{-t_i/\sqrt{\lambda}}[1-\Phi^d]\hat{H}(t_i) \end{pmatrix},$$

which leads to the solution given in Theorem 3 after some standard matrix calculus. □

D Proof of Theorem 4

Proof. Let $f : t \in [0,1] \mapsto t^p$, where $p \in \mathbb{N}$. If $p = 0$, then $\Phi f(t) = 1 - (e^{t/\sqrt{\lambda}} + e^{-t/\sqrt{\lambda}})/2$ for all $t \in [0,1]$. We now assume that $p \geq 1$. Then, thanks to p successive integrations by parts, we get:

$$\begin{cases} \int_0^t e^{-s/\sqrt{\lambda}} s^p ds = -\lambda^{(p+1)/2} p! \left(e^{-t/\sqrt{\lambda}} - 1\right) + \sum_{k=1}^p \left(-\lambda^{k/2} \frac{p!}{(p-k+1)!} t^{p+1-k}\right) e^{-t/\sqrt{\lambda}} \\ \int_0^t e^{s/\sqrt{\lambda}} s^p ds = (-1)^p \lambda^{(p+1)/2} p! \left(e^{t/\sqrt{\lambda}} - 1\right) + \sum_{k=1}^p \left((-1)^{k+1} \lambda^{k/2} \frac{p!}{(p-k+1)!} t^{p+1-k}\right) e^{t/\sqrt{\lambda}}, \end{cases}$$

so that:

$$\Phi f(t) = \begin{cases} -\frac{\lambda^{p/2}}{2} p! \left(e^{t/\sqrt{\lambda}} + e^{-t/\sqrt{\lambda}}\right) + \lambda^{p/2} p! + \sum_{k=0}^{\frac{p}{2}-1} \lambda^k \frac{p!}{(p-2k)!} t^{p-2k} & \text{if } p \text{ is even} \\ -\frac{\lambda^{p/2}}{2} p! \left(e^{t/\sqrt{\lambda}} - e^{-t/\sqrt{\lambda}}\right) + \sum_{k=0}^{\frac{p-1}{2}} \lambda^k \frac{p!}{(p-2k)!} t^{p-2k} & \text{if } p \text{ is odd.} \end{cases}$$

We now use the assumption of H being an analytic function. Then, H can be written as a convergent power series: $H(t) = \sum_{p=0}^{+\infty} a_p t^p$. Therefore:

$$\begin{aligned} \Phi H(t) &= a_0 \left[1 - (e^{t/\sqrt{\lambda}} + e^{-t/\sqrt{\lambda}})/2\right] + \sum_{p=1}^{\infty} a_p \sum_{k=0}^{\lfloor \frac{p-1}{2} \rfloor} \lambda^k \frac{p!}{(p-2k)!} t^{p-2k} \\ &\quad - \sum_{p=1}^{+\infty} a_p \frac{\lambda^{p/2}}{2} p! \left(e^{t/\sqrt{\lambda}} + (-1)^p e^{-t/\sqrt{\lambda}}\right) + \sum_{p=1}^{\infty} a_{2p} \lambda^p (2p)!. \end{aligned}$$

If we choose $A = \sum_{p=0}^{+\infty} a_p \lambda^{p/2} p!/2$ and $B = \sum_{p=0}^{+\infty} (-1)^p a_p \lambda^{p/2} p!/2$, we get:

$$\Phi H(t) + A e^{t/\sqrt{\lambda}} + B e^{-t/\sqrt{\lambda}} = H(t) + \sum_{k=1}^{+\infty} \lambda^k \left(H^{(2k)}(t) + H^{(2k)}(0)\right),$$

where $H^{(n)}$ is the n-th derivative of H. For $\lambda \to 0$, we get:

$$\Phi H(t) + A e^{t/\sqrt{\lambda}} + B e^{-t/\sqrt{\lambda}} - H(t) = \lambda \left(H^{(2)}(t) + H^{(2)}(0)\right) + o(\lambda).$$

The arbitrary choice we made for A and B may not be optimal in the sense of a minimization of $\int_0^1 (\Phi H(t) + A e^{t/\sqrt{\lambda}} + B e^{-t/\sqrt{\lambda}} - H(t))^2 dt$. This remark leads to the inequality stated in Theorem 4. □

References

[1] Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E. and Fernandez-Anaya, G. (2008), *Time-varying Hurst exponent for US stock markets*, Physica A: statistical mechanics and its applications, 387, 24: 6159-6169

[2] Ausloos, M. (2000), *Statistical physics in foreign exchange currency and stock market*, Physica A: statistical mechanics and its applications, 285, 1: 48-65

[3] Ayache, A. and Lévy Véhel, J. (2000), *The generalized multifractional Brownian motion*, Statistical inference for stochastic processes, 3, 1-2: 7-18

[4] Ayache, A. and Taqqu, M.S. (2005), *Multifractional processes with random exponent*, Publicacions matemàtiques, 49, 2: 459-486

[5] Bardet, J.-M. and Surgailis, D. (2013), *Nonparametric estimation of the local Hurst function of multifractional Gaussian processes*, Stochastic processes and their applications, 123, 3: 1004-1045

[6] Benassi, A., Cohen, S. and Istas, J. (1998), *Identifying the multifractional function of a Gaussian process*, Statistics and probability letters, 39, 4: 337-345

[7] Bertrand, P.R., Fhima, M. and Guillin, A. (2013), *Local estimation of the Hurst index of multifractional Brownian motion by Increment Ratio Statistic method*, ESAIM: probability and statistics, 17: 307-327

[8] Bertrand, P.R., Hamdouni, A. and Khadhraoui, S. (2012), *Modelling NASDAQ series by sparse multifractional Brownian motion*, Methodology and computing in applied probability, 14, 1: 107-124

[9] Bianchi, S. (2005), *Pathwise identification of the memory function of multifractional Brownian motion with application to finance*, International journal of theoretical and applied finance, 8, 2: 255-281

[10] Bianchi, S. and Pantanella, A. (2010), *Pointwise regularity exponents and market cross-correlations*, International review of business research papers, 6, 2: 39-51

[11] Bianchi, S., Pantanella, A. and Pianese, A. (2012), *Modeling and simulation of currency exchange rates using multifractional process with random exponent*, International journal of modeling and optimization, 2, 3: 309-314

[12] Chen, K. (2013), *Introduction to variational image-processing models and applications*, International journal of computer mathematics, 90, 1: 1-8

[13] Coeurjolly, J.-F. (2005), *Identification of multifractional Brownian motion*, Bernoulli, 11, 6: 987-1008

[14] Coeurjolly, J.-F. (2008), *Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles*, The annals of statistics, 36, 3: 1404-1434

[15] Da Silva, S., Matsushita, R., Gleria, I. and Figueiredo, A. (2007), *Hurst exponents, power laws, and efficiency in the Brazilian foreign exchange market*, Economics bulletin, 7, 1: 1-11

[16] Dutta, S., Ghosh, D. and Chatterjee, S. (2016), *Multifractal detrended cross correlation analysis of foreign exchange and SENSEX fluctuation in Indian perspective*, Physica A: statistical mechanics and its applications, 463: 188-201

[17] Garcin, M. (2016), *Empirical wavelet coefficients and denoising of chaotic data in the phase space*, in Skiadas, C.H. and Skiadas, C. (editors), *Handbook of applications of chaos theory*, CRC/Taylor & Francis

[18] Garcin, M. and Goulet, C. (2016), *Non-parametric news impact curve: a variational approach*, working paper

[19] Garcin, M. and Guégan, D. (2012), *Extreme values of random or chaotic discretization steps and connected networks*, Applied mathematical sciences, 119, 6: 5901-5926

[20] Garcin, M. and Guégan, D. (2014), *Probability density of the empirical wavelet coefficients of a noisy chaos*, Physica D: nonlinear phenomena, 276: 28-47

[21] Garcin, M. and Guégan, D. (2016), *Wavelet shrinkage of a noisy dynamical system with non-linear noise impact*, Physica D: nonlinear phenomena, 325: 126-145

[22] 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

[23] Grech, D. and Pamuła, G. (2008), *The local Hurst exponent of the financial time series in the vicinity of crashes on the Polish stock exchange market*, Physica A: statistical mechanics and its applications, 387, 16: 4299-4308

[24] Greville, T.N. (1981), *Moving-weighted-average smoothing extended to the extremities of the data. I. Theory*, Scandinavian actuarial journal, 1981, 1: 39-55

[25] Hall, P., Härdle, W., Kleinow, T. and Schmidt, P. (2000), *Semiparametric bootstrap approach to hypothesis tests and confidence intervals for the Hurst coefficient*, Statistical inference for stochastic processes, 3, 3: 263-276

[26] Härdle, W.K., Müller, M., Sperlich, S. and Werwatz, A. (2012), *Nonparametric and semiparametric models*, Springer science and business media

[27] Hurst, H. (1951), *Long-term storage capacity of reservoirs*, Transactions of the American society of civil engineering, 116, 1: 770-808

[28] Istas, J. and Lang, G. (1997), *Quadratic variations and estimation of the local Hölder index of a Gaussian process*, Annales de l'IHP, probabilités et statistiques, 33, 4: 407-436

[29] Kent, J.T. and Wood, A.T. (1997), *Estimating the fractal dimension of a locally self-similar Gaussian process by using increments*, Journal of the royal statistical society, series B (methodological), 59, 3: 679-699

[30] Kolmogorov, A. (1940), *The Wiener spiral and some other interesting curves in Hilbert space*, Doklady akad. nauk SSSR, 26, 2: 115-118

[31] Li, W.S., Tsai, Y.J., Shen, Y.H. and Liaw, S.S. (2016), *General and specific statistical properties of foreign exchange markets during a financial crash*, Physica A: statistical mechanics and its applications, 451: 601-622

[32] Mallat, S. (2000), *Une exploration des signaux en ondelettes*, Ellipses, Éditions de l'École Polytechnique, Paris, France

[33] Mandelbrot, B. (1967), *The variation of the prices of cotton, wheat, and railroad stocks, and of some financial rates*, The journal of business, 40, 1: 393-413

[34] Mandelbrot, B. and van Ness, J. (1968), *Fractional Brownian motions, fractional noises and applications*, SIAM review, 10, 4: 422-437

[35] Mantegna, R. and Stanley, H. (1995), *Scaling behaviour in the dynamics of an economic index*, Nature, 376, 6535: 46-49

[36] Mitra, S.K. (2012), *Is Hurst exponent value useful in forecasting financial time series?*, Asian social science, 8, 8: 111-120

[37] Morales, R., Di Matteo, T., Gramatica, R. and Aste, T. (2012), *Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series*, Physica A: statistical mechanics and its applications, 391,11: 3180-3189

[38] Pal, M., Rao, P.M. and Manimaran, P. (2014), *Multifractal detrended cross-correlation analysis on gold, crude oil and foreign exchange rate time series*, Physica A: statistical mechanics and its applications, 416: 452-460

[39] Peltier, R.F. and Lévy Véhel, J. (1995), *Multifractional Brownian motion: definition and preliminary results*, Rapport de recherche de l'INRIA, 2645

[40] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E. and Goldberger, A.L. (1994), *Mosaic organization of DNA nucleotides*, Physical review E, 49, 2: 1685-1689

[41] Ranta, M. (2010), *Wavelet multiresolution analysis of financial time series*, PhD thesis, Universitas Wasaensis, 223

[42] Roux, D., Jaffard, S. and Benassi, A. (1997), *Elliptic Gaussian random processes*, Revista matemática iberoamericana, 13, 1: 19-90

[43] Rudin, L.I., Osher, S. and Fatemi, E. (1992), *Nonlinear total variation based noise removal algorithms*, Physica D: nonlinear phenomena, 60, 1: 259-268

[44] Weinert, H.L. (2007), *Efficient computation for Whittaker-Henderson smoothing*, Computational statistics and data analysis, 52, 2: 959-974

[45] Weron, R. (2002), *Estimating long-range dependence: finite sample properties and confidence intervals*, Physica A: statistical mechanics and its applications, 312, 1: 285-299

[46] Yao, J. and Tan, C.L. (2000), *A case study on using neural networks to perform technical forecasting of forex*, Neurocomputing, 34, 1: 79-98

[47] Yu, P., Goldenberg, A. and Bi, Z. (2001), *Time series forecasting using wavelets with predictor-corrector boundary treatment*, in 7th ACM SIGKDD international conference on knowledge discovery and data mining, San Francisco, USA

[48] Zhi, Z., Shi, B. and Sun, Y. (2016), *Primal-dual method to smoothing TV-based model for image denoising*, Journal of algorithms and computational technology, 1748301816656298