Chapter 1

Hurst Index Estimation for Self-similar processes with Long-Memory

Alexandra Chronopoulou* and Frederi G. Viens*

Department of Statistics, Purdue University
150 N. University St. West Lafayette
IN 47907-2067, USA.
achronop@purdue.edu, viens@purdue.edu

The statistical estimation of the Hurst index is one of the fundamental problems in the literature of long-range dependent and self-similar processes. In this article, the Hurst index estimation problem is addressed for a special class of self-similar processes that exhibit long-memory, the Hermite processes. These processes generalize the fractional Brownian motion, in the sense that they share its covariance function, but are non-Gaussian. Existing estimators such as the R/S statistic, the variogram, the maximum likelihood and the wavelet-based estimators are reviewed and compared with a class of consistent estimators which are constructed based on the discrete variations of the process. Convergence theorems (asymptotic distributions) of the latter are derived using multiple Wiener-Itô integrals and Malliavin calculus techniques. Based on these results, it is shown that the latter are asymptotically more efficient than the former.

Keywords : self-similar process, parameter estimation, long memory, Hurst parameter, multiple stochastic integral, Malliavin calculus, Hermite process, fractional Brownian motion, non-central limit theorem, quadratic variation.

2000 AMS Classification Numbers: Primary 62F12; Secondary 60G18, 60H07, 62M09.

1.1. Introduction

1.1.1. *Motivation*

A fundamental assumption in many statistical and stochastic models is that of independent observations. Moreover, many models that do not make this assumption have the convenient Markov property, according to which the future of the system is not affected by its previous states but only by the current one.

The phenomenon of *long memory* has been noted in nature long before the construction of suitable stochastic models: in fields as diverse as hydrology, economics,

*Both authors' research partially supported by NSF grant 0606615.

chemistry, mathematics, physics, geosciences, and environmental sciences, it is not uncommon for observations made far apart in time or space to be non-trivially correlated.

Since ancient times the Nile River has been known for its long periods of dryness followed by long periods of floods. The hydrologist Hurst ([13]) was the first one to describe these characteristics when he was trying to solve the problem of flow regularization of the Nile River. The mathematical study of long-memory processes was initiated by the work of Mandelbrot [16] on self-similar and other stationary stochastic processes that exhibit long-range dependence. He built the foundations for the study of these processes and he was the first one to mathematically define the fractional Brownian motion, the prototype of self-similar and long-range dependent processes. Later, several mathematical and statistical issues were addressed in the literature, such as derivation of central (and non-central) limit theorems ([5], [6], [10], [17], [27]), parameter estimation techniques ([1], [7], [8], [27]) and simulation methods ([11]).

The problem of the statistical estimation of the self-similarity and/or long-memory parameter H is of great importance. This parameter determines the mathematical properties of the model and consequently describes the behavior of the underlying physical system. Hurst ([13]) introduced the celebrated *rescaled adjusted range* or *R/S* statistic and suggested a graphical methodology in order to estimate H. What he discovered was that for data coming from the Nile River the R/S statistic behaves like a constant times k^H, where k is a time interval. This was called later by Mandelbrot the *Hurst effect* and was modeled by a *fractional Gaussian noise* (*fGn*).

One can find several techniques related to the Hurst index estimation problem in the literature. There are a lot of graphical methods including the R/S statistic, the correlogram and partial correlations plot, the variance plot and the variogram, which are widely used in geosciences and hydrology. Due to their graphical nature they are not so accurate and thus there is a need for more rigorous and sophisticated methodologies, such as the maximum likelihood. Fox and Taqqu ([12]) introduced the Whittle approximate maximum likelihood method in the Gaussian case which was later generalized for certain non-Gaussian processes. However, these approaches were lacking computational efficiency which lead to the rise of wavelet-based estimators and discrete variation techniques.

1.1.2. *Mathematical Background*

Let us first recall some basic definitions that will be useful in our analysis.

Definition 1.1. A stochastic process $\{X_n; n \in \mathbb{N}\}$ is said to be *stationary* if the vectors $(X_{n_1}, \ldots, X_{n_d})$ and $(X_{n_1+m}, \ldots, X_{n_d+m})$ have the same distribution for all integers d, $m \geq 1$ and $n_1, \ldots, n_d \geq 0$. For Gaussian processes this is equivalent to requiring that $Cov(X_m, X_{m+n}) := \gamma(n)$ does not depend on m. These two notions

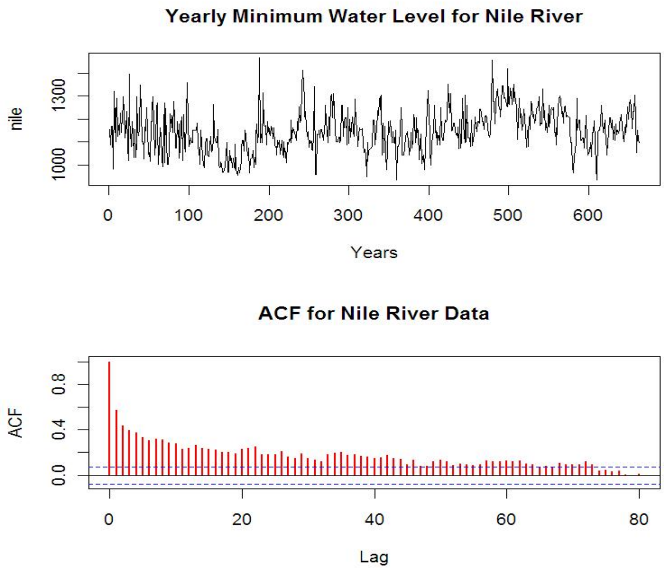


Fig. 1.1. Yearly minimum water levels of the Nile River at the Roda Gauge (622-1281 A.D.). The dotted horizontal lines represent the levels $\pm 2/\sqrt{600}$. Since our observations are above these levels, it means that they are significantly correlated with significance level 0.05.

are often called *strict stationarity* and *second-order stationarity*, respectively. The function $\gamma(n)$ is called the *autocovariance function.* The function $\rho(n) = \gamma(n)/\gamma(0)$ is the called *autocorrelation* function.

In this context, long memory can be defined in the following way:

Definition 1.2. Let $\{X_n; n \in \mathbb{N}\}$ be a stationary process. If $\sum_n \rho(n) = +\infty$ then X_n is said to exhibit *long memory* or *long-range dependence.* A sufficient condition for this is the existence of $H \in (1/2, 1)$ such that

$$\liminf_{n\to\infty} \frac{\rho(n)}{n^{2H-2}} > 0.$$

Typical long memory models satisfy the stronger condition $\lim_{n\to\infty} \rho(n)/n^{2H-2} > 0$, in which case H can be called the *long memory parameter* of X.

A process that exhibits long-memory has an autocorrelation function that decays very slowly. This is exactly the behavior that was observed by Hurst for the first time. In particular, he discovered that the yearly minimum water level of the Nile river had the long-memory property, as can been seen in Figure 1.1.

Another property that was observed in the data collected from the Nile river is the so-called *self-similarity* property. In geometry, a self-similar shape is one composed of a basic pattern which is repeated at multiple (or infinite) scale. The

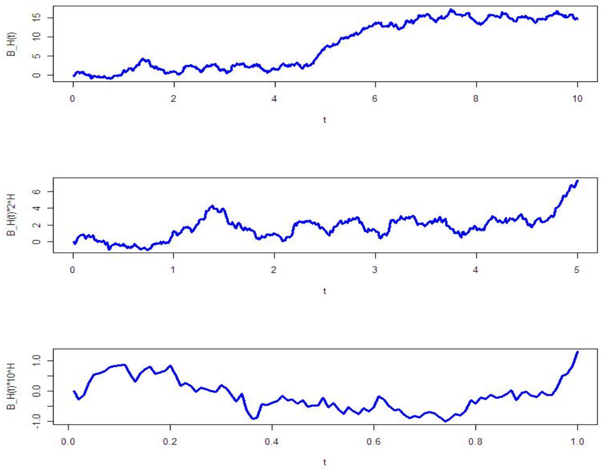


Fig. 1.2. Self-similarity property for the fractional Brownian motion with $H = 0.75$. The first graph shows the path from time 0 to 10. The second and third graph illustrate the normalized sample path for $0 < t < 5$ and $0 < t < 1$ respectively.

statistical interpretation of self-similarity is that the paths of the process will look the same, in distribution, irrespective of the distance from which we look at then. The rigorous definition of the self-similarity property is as follows:

Definition 1.3. A process $\{X_t; t \geq 0\}$ is called *self-similar* with self-similarity parameter H, if for all $c > 0$, we have the identity in distribution

$$\left\{c^{-H} X_{c\,t} : t \geq 0\right\} \overset{\mathcal{D}}{\sim} \{X_t : t \geq 0\}.$$

In Figure 1.2, we can observe the self-similar property of a simulated path of the fractional Brownian motion with parameter $H = 0.75$.

In this paper, we concentrate on a special class of long-memory processes which are also self-similar and for which the self-similarity and long-memory parameters coincide, the so-called *Hermite processes*. This is a family of processes parametrized by the order q and the self-similarity parameter H. They all share the same covariance function

$$Cov(X_t, X_s) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{1.1}$$

From the structure of the covariance function we observe that the Hermite processes have stationary increments, they are H-self-similar and they exhibit long-range dependence as defined in Definition 1.2 (in fact, $\lim_{n\to\infty} \rho(n) /\!\!/ n^{2H-2} = H(2H-1)$). The Hermite process for $q = 1$ is a standard *fractional Brownian motion* with Hurst parameter H, usually denoted by B^H, the only Gaussian process in the Hermite class. A Hermite process with $q = 2$ known as *the Rosenblatt process*. In the sequel, we will call H either long-memory parameter or self-similarity parameter or *Hurst parameter*. The mathematical definition of these processes is given in Definition 1.5.

Another class of processes used to model long-memory phenomena are the fractional ARIMA (Auto Regressive, Integrated, Moving Average) or FARIMA processes. The main technical difference between a FARIMA and a Hermite process is that the first one is a discrete-time process and the second one a continuous-time process. Of course, in practice, we can only have discrete observations. However, most phenomena in nature evolve continuously in time and the corresponding observations arise as samplings of continuous time processes. A discrete-time model depends heavily on the sampling frequency: daily observations will be described by a different FARIMA model than weekly observations. In a continuous time model, the observation sampling frequency does not modify the model. These are compelling reasons why one may choose to work with the latter.

In this article we study the Hurst parameter estimation problem for the Hermite processes. The structure of the paper is as follows: in Section 2, we provide a survey of the most widely used estimators in the literature. In Section 3 we describe the main ingredients and the main definitions that we need for our analysis. In Section 4, we construct a class of estimators based on the discrete variations of the process and describe their asymptotic properties, including a sketch of the proof of the main theoretical result, Theorem 1.4, which summarizes the series of papers [6], [7], [27] and [28]. In the last section, we compare the variations-based estimators with the existing ones in the literature, and provide an original set of practical recommendations based on theoretical results and on simulations.

1.2. Most Popular Hurst parameter Estimators

In this section we discuss the main estimators for the Hurst parameter in the literature. We start with the description of three heuristic estimators: the R/S estimator, the correlogram and the variogram. Then, we concentrate on a more traditional approach: the maximum likelihood estimation. Finally, we briefly describe the wavelet-based estimator.

The description will be done in the case of the *fractional Brownian motion* (fBm) $\{B_t^H;\, t \in [0,1]\}$. We assume that it is observed in discrete times $\{0, 1, \ldots, N-1, N\}$. We denote by $\{X_t^H; t \in [0,1]\}$ the corresponding increment process of the fBm (i.e. $X^H_{\frac{i}{N}} = B^H_{\frac{i}{N}} - B^H_{\frac{i}{N}}$) , also known as *fractional Gaussian noise*.

1.2.1. *Heuristic Estimators*

R/S Estimator :

The most famous among these estimators is the so-called R/S estimator that was first proposed by Hurst in 1951, [13], in the hydrological problem regarding the storage of water coming from the Nile river. We start by dividing our data in K non-intersecting blocks, each one of which contains $M = \left[\frac{N}{K}\right]$ elements. The *rescaled adjusted range* is computed for various values of N by

$$Q := Q(t_i, N) = \frac{R(t_i, N)}{S(t_i, n)}$$

at times $t_i = M(i-1),\ i = 1, \ldots, K$. For

$$Y(t_i, k) := \sum_{j=0}^{k-1} X^H_{t_i+j} - k\left(\frac{1}{n}\sum_{j=0}^{n-1} X^H_{t_i+j}\right), \quad k = 1, \ldots, n$$

we define $R(t_i, n)$ and $S(t_i, n)$ to be

$$R(t_i, n) := \max\{Y(t_i, 1), \ldots, Y(t_i, n)\} - \min\{Y(t_i, 1), \ldots, Y(t_i, n)\} \quad \text{and}$$

$$S(t_i, n) := \sqrt{\frac{1}{n}\sum_{j=0}^{n-1} X^{H\,2}_{t_i+j} - \left(\frac{1}{n}\sum_{j=0}^{n-1} X^H_{t_i+j}\right)^2}.$$

Remark 1.1. It is interesting to note that the numerator $R(t_i, n)$ can be computed only when $t_i + n \leq N$.

In order to compute a value for H we plot the logarithm of R/S (i.e $\log Q$) with respect to $\log n$ for several values of n. Then, we fit a least-squares line $y = a + b\log n$ to a central part of the data, that seem to be nicely scattered along a straight line. The slope of this line is the estimator of H.

This is a graphical approach and it is really in the hands of the statistician to determine the part of the data that is "nicely scattered along the straight line". The problem is more severe in small samples, where the distribution of the R/S statistic is far from normal. Furthermore, the estimator is biased and has a large standard error. More details on the limitations of this approach in the case of fBm can be found in [2].

Correlogram :

Recall $\rho(N)$ the autocorrelation function of the process as in Definition 1.1. In the Correlogram approach, it is sufficient to plot the *sample autocorrelation function*

$$\hat{\rho}(N) = \frac{\hat{\gamma}(N)}{\hat{\gamma}(0)}$$

against N. As a rule of thumb we draw two horizontal lines at $\pm 2/\sqrt{N}$. All observations outside the lines are considered to be significantly correlated with significance level 0.05. If the process exhibits long-memory, then the plot should have a very slow decay.

The main disadvantage of this technique is its graphical nature which cannot guarantee accurate results. Since long-memory is an asymptotic notion, we should analyze the correlogram at high lags. However, when for example $H = 0.6$ it is quite hard to distinguish it from short-memory. To avoid this issue, a more suitable plot will be this of $\log \hat{\rho}(N)$ against $\log N$. If the asymptotic decay is precisely hyperbolic, then for large lags the points should be scattered around a straight line with negative slope equal to $2H - 2$ and the data will have long-memory. On the other hand when the plot diverges to $-\infty$ with at least exponential rate, then the memory is short.

Variogram :
The variogram for the lag N is defined as

$$V(N) := \frac{1}{2}\mathbf{E}\left[\left(B_t^H - B_{t-N}^H\right)^2\right].$$

Therefore, it suffices to plot $V(N)$ against N. However, we can see that the interpretation of the variogram is similar to that of the correlogram, since if the process is stationary (which is true for the increments of fractional Brownian motion and all other Hermite processes), then the variogram is asymptotically finite and

$$V(N) = V(\infty)(1 - \rho(N)).$$

In order to determine whether the data exhibit short or long memory this method has the same problems as the correlogram.

The main advantage of these approaches is their simplicity. In addition, due to their non-parametric nature, they can be applied to any long-memory process. However, none of these graphical methods are accurate. Moreover, they can frequently be misleading, indicating existence of long-memory in cases where none exists. For example, when a process has short-memory together with a trend that decays to zero very fast, a correlogram or a variogram could show evidence of long-memory.

In conclusion, a good approach would be to use these methods as a first heuristic analysis to detect the possible presence of long-memory and then use a more rigorous technique, such as those described in the remainder of this section, in order to estimate the long-memory parameter.

1.2.2. *Maximum Likelihood Estimation*

The Maximum Likelihood Estimation (*mle*) is the most common technique of parameter estimation in Statistics. In the class of Hermite processes, its use is limited to fBm, since for the other processes we do not have an expression for their distribution function. The *mle* estimation is done in the spectral domain using the spectral density of fBm as follows.

Denote by $X^H = \left(X_0^H, X_1^H, \ldots, X_N^H\right)$ the vector of the fractional Gaussian noise (increments of fBm) and by $X^{H\,'}$ the transposed (column) vector; this is a Gaussian vector with covariance matrix $\Sigma_N(H) = [\sigma_{ij}(H)]_{i,j=1,\ldots,N}$; we have

$$\sigma_{ij} := Cov\left(X_i^H \ ; \ X_j^H\right) = \frac{}{2}\left(i^{2H} + j^{2H} - |i-j|^{2H}\right).$$

Then, the *log-likelihood function* has the following expression:

$$\log f(x;H) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log\left[\det\left(\Sigma_N(H)\right)\right] - \frac{1}{2}X^H\left(\Sigma_N(H)\right)^{-1}X^{H\,'}.$$

In order to compute $\hat{H}_{mle}$, the *mle* for H, we need to maximize the log-likelihood equation with respect to H. A detailed derivation can be found in [3] and [9]. The asymptotic behavior of $\hat{H}_{mle}$ is described in the following theorem.

Theorem 1.1. *Define the quantity* $D(H) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial H}\log f(x;H)\right)^2 dx$. *Then under certain regularity conditions (that can be found in [9]) the maximum likelihood estimator is weakly consistent and asymptotically normal:*

(i) $\hat{H}_{mle} \to H$ *, as* $N \to \infty$ *in probability;*
(ii) $\sqrt{N}\sqrt{2\,D(H)}\left(\hat{H}_{mle} - H\right) \to \mathcal{N}(0,1)$ *in distribution, as* $N \to \infty$.

In order to obtain the *mle* in practice, in almost every step we have to maximize a quantity that involves the computation of the inverse of $\Sigma(H)$, which is not an easy task.

In order to avoid this computational burden, we approximate the likelihood function with the so-called *Whittle* approximate likelihood which can be proved to converge to the true likelihood, [29]. In order to introduce Whittle's approximation we first define the density on the spectral domain.

Definition 1.4. Let X_t be a process with autocovariance function $\gamma(h)$, as in Definition 1.1. The *spectral density function* is defined as the inverse Fourier transform of $\gamma(h)$

$$f(\lambda) := \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-i\lambda h}\gamma(h).$$

In the fBm case the spectral density can be written as

$$f(\lambda;H) = \frac{1}{2\pi}exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_1 \, d\lambda\right\}, \quad \text{where}$$

$$f_1(\lambda;H) = \frac{1}{\pi}\Gamma(2H+1)\sin(\pi H)(1-\cos\lambda)\sum_{j=-\infty}^{\infty}|2\pi j + \lambda|^{-2H-1}$$

The Whittle method approximates each of the terms in the log-likelihood function as follows:

(i) $\lim_{N\to\infty} \log\det(\Sigma_N(H)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda; H)d\lambda$.
(ii) The matrix $\Sigma_N^{-1}(H)$ itself is asymptotically equivalent to the matrix $A(H) = [\alpha(j-\ell)]_{j\ell}$, where

$$\alpha(j-\ell) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} \frac{e^{-i(j-\ell)\lambda}}{f(\lambda; H)} d\lambda$$

Combining the approximations above, we now need to minimize the quantity

$$(\log f(\lambda; H))^* = -\frac{N}{2}\log 2\pi - \frac{n}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda; H)d\lambda - \frac{1}{2}X\ A(H)X'.$$

The details in the Whittle *mle* estimation procedure can be found in [3]. For the Whittle *mle* we have the following convergence in distribution result as $N \to \infty$

$$\sqrt{\frac{N}{[2\ D(H)]^{-1}}}\ \left(\hat{H}_{Wmle} - H\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \tag{1.2}$$

It can also be shown that the Whittle approximate *mle* remains weakly consistent.

1.2.3. *Wavelet Estimator*

Much attention has been devoted to the wavelet decomposition of both fBm and the Rosenblatt process. Following this trend, an estimator for the Hurst parameter based on wavelets has been suggested. The details of the procedure for the constructing this estimator, and the underlying wavelets theory, are beyond the scope of this article. For the proofs and the detailed exposition of the method the reader can refer to [1], [11] and [14]. This section provides a brief exposition.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a continuous function with support in $[0,1]$. This is also called the *mother wavelet*. $Q \geq 1$ is the number of vanishing moments where

$$\int_{\mathbb{R}} t^p \psi(t)dt = 0, \ \text{ for } \ p = 0, 1, \ldots, Q-1,$$
$$\int_{\mathbb{R}} t^Q \psi(t)dt \neq 0.$$

For a "scale" $\alpha \in \mathbb{N}^*$ the corresponding wavelet coefficient is given by

$$d(\alpha, i) = \frac{1}{\sqrt{\alpha}}\int_{-\infty}^{\infty} \psi\left(\frac{t}{\alpha} - i\right) Z_t^H dt,$$

for $i = 1, 2, \ldots, N_\alpha$ with $N_\alpha = \left[\frac{N}{\alpha}\right] - 1$, where N is the sample size. Now, for (α, b) we define the *approximate wavelet coefficient* of $d(\alpha, b)$ as the following Riemann approximation

$$e(\alpha, b) = \frac{1}{\sqrt{\alpha}}\sum_{k=1}^{N} Z_k^H \psi\left(\frac{k}{\alpha} - b\right),$$

where Z^H can be either fBm or Rosenblatt process. Following the analysis by J.-M. Bardet and C.A. Tudor in [1], the suggested estimator can be computed by performing a *log-log regression* of

$$\left(\frac{1}{N_{\alpha_i(N)}} \sum_{j=1}^{N_{\alpha_i(N)}} e^2\left(\alpha_i(N), j\right) \right)_{1 \le i \le \ell}$$

against $(i\ \alpha_i(N))_{1\le i\le \ell}$, where $\alpha(N)$ is a sequence of integer numbers such that $N\alpha(N)^{-1} \to \infty$ and $\alpha(N) \to \infty$ as $N \to \infty$ and $\alpha_i(N) = i\alpha(N)$. Thus, the obtained estimator, in vectors notation, is the following

$$\hat{H}_{wave} := \left(\frac{1}{2}, 0\right)' \left(Z_\ell', Z_\ell\right)^{-1} Z_\ell^{-1} \left(\frac{1}{2} \sum_{j=1}^{N_{\alpha_i(N)}} e^2\left(\alpha_i(N), j\right) \right)_{1\le i \le \ell} - \frac{1}{2}, \tag{1.3}$$

where $Z_\ell(i,1) = 1$, $Z_\ell(i,2) = \log i$ for all $i = 1, \ldots, \ell$, for $\ell \in \mathbb{N} \smallsetminus \{1\}$.

Theorem 1.2. *Let $\alpha(N)$ as above. Assume also that $\psi \in \mathcal{C}^m$ with $m \ge 1$ and ψ is supported on $[0,1]$. We have the following convergences in distribution.*

(1) Let Z^H be a fBm; assume $N\alpha(N)^{-2} \to 0$ as $N \to \infty$ and $m \ge 2$; if $Q \ge 2$, or if $Q = 1$ and $0 < H < 3/4$, then there exists $\gamma^2(H, \ell, \psi) > 0$ such that

$$\sqrt{\frac{N}{\alpha(N)}} \left(\hat{H}_{wave} - H\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2(H, \ell, \psi)), \quad as \ N \to \infty. \tag{1.4}$$

(2) Let Z^H be a fBm; assume $N\alpha(N)^{-\frac{5-4H}{4-4H}} \to 0$ as $N\ \alpha(N)^{-\frac{3-2H+m}{3-2H}} \to 0$; if $Q = 1$ and $3/4 < H < 1$, then

$$\left(\frac{N}{\alpha(N)}\right)^{2-2H} \left(\hat{H}_{wave} - H\right) \xrightarrow{\mathcal{D}} L, \quad as \ N \to \infty \tag{1.5}$$

where the distribution law L depends on H, ℓ and ψ.

(3) Let Z^H is be Rosenblatt process; assume $N\alpha(N)^{-\frac{2-2H}{3-2H}} \to 0$ as $N\ \alpha(N)^{-(1+m)} \to 0$; then

$$\left(\frac{N}{\alpha(N)}\right)^{1-H} \left(\hat{H}_{wave} - H\right) \xrightarrow{\mathcal{D}} L, \quad as \ N \to \infty \tag{1.6}$$

where the distribution law L depends on H, ℓ and ψ.

The limiting distributions L in the theorem above are not explicitly known: they come from a non-trivial non-linear transformation of quantities which are asymptotically normal or Rosenblatt-distributed. A very important advantage of $\hat{H}_{wave}$ over the *mle* for example is that it can be computed in an efficient and fast way. On the other hand, the convergence rate of the estimator depends on the choice of $\alpha(N)$.

1.3. Multiplication in the Wiener Chaos & Hermite Processes

1.3.1. *Basic tools on multiple Wiener-Itô integrals*

In this section we describe the basic framework that we need in order to describe and prove the asymptotic properties of the estimator based on the discrete variations of the process. We denote by $\{W_t : t \in [0,1]\}$ a classical Wiener process on a standard Wiener space $(\Omega, \mathcal{F}, P)$. Let $\left\{B_t^H; t \in [0,1]\right\}$ be a fractional Brownian motion with Hurst parameter $H \in (0,1)$ and covariance function

$$\left\langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \right\rangle = R_H(t,s) := \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{1.7}$$

We denote by $\mathcal{H}$ its canonical Hilbert space. When $H = \frac{1}{2}$, then $B^{\frac{1}{2}}$ is the standard Brownian motion on $L^2([0,1])$. Otherwise, $\mathcal{H}$ is a Hilbert space which contains functions on $[0,1]$ under the inner product that extends the rule $\left\langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \right\rangle$. Nualart's textbook (Chapter 5, [19]) can be consulted for full details.

We will use the representation of the fractional Brownian motion B^H with respect to the standard Brownian motion W: there exists a Wiener process W and a deterministic kernel $K^H(t,s)$ for $0 \le s \le t$ such that

$$B^H(t) = \int_0^1 K^H(t,s)dW_s = I_1\left(K^H(\cdot,t)\right), \tag{1.8}$$

where I_1 is the Wiener-Itô integral with respect to W. Now, let $I_n(f)$ be the multiple Wiener-Itô integral, where $f \in L^2([0,1]^n)$ is a symmetric function. One can construct the multiple integral starting from simple functions of the form $f := \sum_{i_1,\ldots,i_n} c_{i_1,\ldots i_n} 1_{A_{i_1} \times \ldots \times A_{i_n}}$ where the coefficient $c_{i_1,..,i_n}$ is zero if two indices are equal and the sets A_{i_j} are disjoint intervals by

$$I_n(f) := \sum_{i_1,\ldots,i_n} c_{i_1,\ldots i_n} W(A_{i_1}) \ldots W(A_{i_n}),$$

where $W\left(1_{[a,b]}\right) = W([a,b]) = W_b - W_a$. Using a density argument the integral can be extended to all symmetric functions in $L^2([0,1]^n)$. The reader can refer to Chapter 1 [19] for its detailed construction. Here, it is interesting to observe that this construction coincides with the iterated Itô stochastic integral

$$I_n(f) = n! \int_0^1 \int_0^{t_n} \ldots \int_0^{t_2} f(t_1,\ldots,t_n) dW_{t_1} \ldots dW_{t_n}. \tag{1.9}$$

The application I_n is extended to non-symmetric functions f via

$$I_n(f) = I_n(\tilde{f})$$

where $\tilde{f}$ denotes the symmetrization of f defined by $\tilde{f}(x_1,\ldots,x_N) = \frac{1}{n!}\sum_{\sigma \in \mathcal{S}_n} f(x_{\sigma(1)},\ldots,x_{\sigma(n)})$.

I_n is an isometry between the Hilbert space $\mathcal{H}^{\odot n}$ equipped with the scaled norm $\frac{1}{\sqrt{n!}}||\cdot||_{\mathcal{H}^{\otimes n}}$. The space of all integrals of order n, $\left\{I_n(f) : f \in L^2([0,1]^n)\right\}$, is called

n^{th} Wiener chaos. The Wiener chaoses form orthogonal sets in $L^2(\Omega)$:

$$\mathbf{E}\left(I_n(f)I_m(g)\right) = n!\langle f,g\rangle_{L^2([0,1]^n)} \quad \text{if } m=n, \qquad (1.10)$$
$$= 0 \quad \text{if } m \neq n.$$

The next multiplication formula will plays a crucial technical role: if $f \in L^2([0,1]^n)$ and $g \in L^2([0,1]^m)$ are symmetric functions, then it holds that

$$I_n(f)I_m(g) = \sum_{\ell=0}^{m\wedge n} \ell! C_m^\ell C_n^\ell I_{m+n-2\ell}(f \otimes_\ell g), \qquad (1.11)$$

where the contraction $f \otimes_\ell g$ belongs to $L^2([0,1]^{m+n-2\ell})$ for $\ell = 0,1,\ldots,m\wedge n$ and is given by

$$(f \otimes_\ell g)(s_1,\ldots,s_{n-\ell},t_1,\ldots,t_{m-\ell})$$
$$= \int_{[0,1]^\ell} f(s_1,\ldots,s_{n-\ell},u_1,\ldots,u_\ell)g(t_1,\ldots,t_{m-\ell},u_1,\ldots,u_\ell)du_1\ldots du_\ell.$$

Note that the contraction $(f \otimes_\ell g)$ is not necessarily symmetric. We will denote its symmetrization by $(f\tilde{\otimes}_\ell g)$.

We now introduce the Malliavin derivative for random variables in a finite chaos. The derivative operator D is defined on a subset of $L^2(\Omega)$, and takes values in $L^2(\Omega \times [0,1])$. Since it will be used for random variables in a finite chaos, it is sufficient to know that if $f \in L^2([0,1]^n)$ is a symmetric function, $DI_n(f)$ exists and it is given by

$$D_t I_n(f) = n\, I_{n-1}(f(\cdot,t)), \quad \in [0,1].$$

D. Nualart and S. Ortiz-Latorre in [21] proved the following characterization of convergence in distribution for any sequence of multiple integrals to the standard normal law.

Proposition 1.1. *Let n be a fixed integer. Let $F_N = I_n(f_N)$ be a sequence of square integrable random variables in the n^{th} Wiener chaos such that $\lim_{N\to\infty} \mathbf{E}\left[F_N^2\right] = 1$. Then the following are equivalent:*

(i) *The sequence $(F_N)_{N\geq 0}$ converges to the normal law $\mathcal{N}(0,1)$.*
(ii) $\|DF_N\|^2_{L^2[0,1]} = \int_0^1 |D_t I_n(f)|^2\, dt$ *converges to the constant n in $L^2(\Omega)$ as $N \to \infty$.*

There also exists a multidimensional version of this theorem due to G. Peccati and C. Tudor in [22].

1.3.2. *Main Definitions*

The Hermite processes are a family of processes parametrized by the order and the self-similarity parameter with covariance function given by (1.7). They are well-suited to modeling various phenomena that exhibit long-memory and have the self-similarity property, but which are not Gaussian. We denote by $(Z_t^{(q,H)})_{t\in[0,1]}$ the

Hermite process of order q with self-similarity parameter $H \in (1/2, 1)$ (here $q \geq 1$ is an integer). The Hermite process can be defined in two ways: as a multiple integral with respect to the standard Wiener process $(W_t)_{t\in[0,1]}$; or as a multiple integral with respect to a fractional Brownian motion with suitable Hurst parameter. We adopt the first approach throughout the paper, which is the one described in (1.8).

Definition 1.5. The Hermite process $(Z_t^{(q,H)})_{t\in[0,1]}$ of order $q \geq 1$ and with self-similarity parameter $H \in (\frac{1}{2}, 1)$ for $t \in [0, 1]$ is given by

$$Z_t^{(q,H)} = d(H) \int_0^t \ldots \int_0^t dW_{y_1} \ldots dW_{y_q} \left(\int_{y_1 \vee \ldots \vee y_q}^t \partial_1 K^{H'}(u, y_1) \ldots \partial_1 K^{H'}(u, y_q) du \right), \tag{1.12}$$

where $K^{H'}$ is the usual kernel of the fractional Brownian motion, $d(H)$ a constant depending on H and

$$H' = 1 + \frac{H-1}{q} \Longleftrightarrow (2H' - 2)q = 2H - 2. \tag{1.13}$$

Therefore, the Hermite process of order q is defined as a q^{th} order Wiener-Itô integral of a non-random kernel, i.e.

$$Z_t^{(q,H)} = I_q\left(L(t, \cdot)\right),$$

where $L(t, y_1, \ldots, y_q) = \partial_1 K^{H'}(u, y_1) \ldots \partial_1 K^{H'}(u, y_q) du$.

The basic properties of the Hermite process are listed below:

- the Hermite process $Z^{(q,H)}$ is H-selfsimilar and it has stationary increments;
- the mean square of its increment is given by

$$\mathbf{E}\left[\left|Z_t^{(q,H)} - Z_s^{(q,H)}\right|^2\right] = |t - s|^{2H};$$

  as a consequence, it follows from the Kolmogorov continuity criterion that, almost surely, $Z^{(q,H)}$ has Hölder-continuous paths of any order $\delta < H$;
- $Z^{(q,H)}$ exhibits long-range dependence in the sense of Definition 1.2. In fact, the autocorrelation function $\rho(n)$ *of its increments* of length 1 is asymptotically equal to $H(2H-1)n^{2H-2}$. This property is identical to that of fBm since the processes share the same covariance structure, and the property is well-known for fBm with $H > 1/2$. In particular for Hermite processes, the self-similarity and long-memory parameter coincide.

In the sequel, we will also use the *filtered* process to construct an estimator for H.

Definition 1.6. A filter α of length $\ell \in \mathbb{N}$ and order $p \in \mathbb{N} \setminus 0$ is an $(\ell+1)$-dimensional vector $\alpha = \{\alpha_0,\ \alpha_1, \ldots, \alpha_\ell\}$ such that

$$\sum_{q=0}^{\ell} \alpha_q q^r = 0, \quad \text{for} \quad 0 \le r \le p-1,\ r \in \mathbf{Z}$$

$$\sum_{q=0}^{\ell} \alpha_q q^p \neq 0$$

with the convention $0^0 = 1$.

We assume that we observe the process in discrete times $\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\}$. The filtered process $Z^{(q,H)}(\alpha)$ is the convolution of the process with the filter, according to the following scheme:

$$Z(\alpha) := \sum_{q=0}^{\ell} \alpha_q Z^{(q,H)}\left(\frac{i-q}{N}\right), \quad \text{for } i = \ell, \ldots, N-1 \tag{1.14}$$

Some examples are the following:

(1) For $\alpha = \{1, -1\}$

$$Z^{(q,H)}(\alpha) = Z^{(q,H)}\left(\frac{i}{N}\right) - Z^{(q,H)}\left(\frac{i-1}{N}\right).$$

This is a filter of length 1 and order 1.

(2) For $\alpha = \{1, -2, 1\}$

$$Z^{(q,H)}(\alpha) = Z^{(q,H)}\left(\frac{i}{N}\right) - 2Z^{(q,H)}\left(\frac{i-1}{N}\right) + Z^{(q,H)}\left(\frac{i-2}{N}\right).$$

This is a filter of length 2 and order 2.

(3) More generally, longer filters produced by finite-differencing are such that the coefficients of the filter α are the binomial coefficients with alternating signs. Borrowing the notation ∇ from time series analysis, $\nabla Z^{(q,H)}(i/N) = Z^{(q,H)}(i/N) - Z^{(q,H)}((i-1)/N)$, we define $\nabla^j = \nabla\nabla^{j-1}$ and we may write the jth-order finite-difference-filtered process as follows

$$Z^{(q,H)}(\alpha) := \left(\nabla^j Z^{(q,H)}\right)\left(\frac{i}{N}\right).$$

1.4. Hurst parameter Estimator based on Discrete Variations

The estimator based on the discrete variations of the process is described by Coeurjolly in [8] for fractional Brownian motion. Using previous results by Breuer and Major, [5], he was able to prove consistency and derive the asymptotic distribution for the suggested estimator in the case of filter of order 1 for $H < 3/4$ and for all H in the case of a longer filter.

Herein we see how the results by Coeurjolly are generalized: we construct consistent estimators for the self-similarity parameter of a Hermite process of order q based on the discrete observations of the underlying process. In order to determine the corresponding asymptotic behavior we use properties of the Wiener-Itô integrals as well as Malliavin calculus techniques. The estimation procedure is the same irrespective of the specific order of the Hermite process, thus in the sequel we denote the process by $Z := Z^{(q,H)}$.

1.4.1. *Estimator Construction*

Filter of order 1 : $\alpha = \{-1, +1\}$.

We present first the estimation procedure for a filter of order 1, i.e. using the increments of the process. The quadratic variation of Z is

$$S_N(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \left(Z\left(\frac{i}{N}\right) - Z\left(\frac{i-1}{N}\right) \right)^2. \tag{1.15}$$

We know that the expectation of $S_N(\alpha)$ is $\mathbf{E}[S_N(\alpha)] = N^{-2H}$; thus, given good concentration properties for $S_N(\alpha)$, we may attempt to estimate $S_N(\alpha)$'s expectation by its actual value, i.e. $\mathbf{E}[S_N(\alpha)]$ by $S_N(\alpha)$; suggesting the following estimator for H:

$$\hat{H}_N = -\frac{\log S_N(\alpha)}{2 \log N}. \tag{1.16}$$

Filter of order p :

In this case we use the filtered process in order to construct the estimator for H. Let α be a filter (as defined in (1.6)) and the corresponding filtered process $Z(\alpha)$ as in (1.14): First we start by computing the quadratic variation of the filtered process

$$S_N(\alpha) = \frac{1}{N} \sum_{i=\ell}^{N} \left(\sum_{q=0}^{\ell} \alpha_q Z\left(\frac{i-q}{N}\right) \right)^2. \tag{1.17}$$

Similarly as before, in order to construct the estimator, we estimate S_N by its expectation, which computes as $\mathbf{E}[S_N] = -\frac{N^{-2H}}{2} \sum_{q,r=0}^{\ell} \alpha_q \alpha_r |q-r|^{2H}$. Thus, we can obtain $\hat{H}_N$ by solving the following non-linear equation with respect to H

$$S_N = -\frac{N^{-2H}}{2} \sum_{q,r=0}^{\ell} \alpha_q \alpha_r |q-r|^{2H}. \tag{1.18}$$

We write that $\hat{H}_N = g^{-1}(S_N)$, where $g(x) = -\frac{N^{-2x}}{2} \sum_{q,r=0}^{\ell} \alpha_q \alpha_r |q-r|^{2x}$. In this case, it is not possible to compute an analytical expression for the estimator. However, we can show that there exists a unique solution for $H \in [\frac{1}{2}, 1]$ as long as

$$N > \max_{H\in[\frac{1}{2},1]} \exp\left\{ \frac{\sum_{q,r=0}^{\ell} \alpha_q\alpha_r \log|q-r|\;|q-r|^{2H}}{\sum_{q,r=0}^{\ell} \alpha_q\alpha_r |q-r|^{2H}} \right\}.$$

This restriction is typically satisfied, since we work with relatively large sample sizes.

1.4.2. *Asymptotic Properties of $\hat{H}_N$*

The first question is whether the suggested estimator is consistent. This is indeed true: if sampled sufficiently often (i.e. as $N \to \infty$), the estimator converges to the true value of H almost surely, for any order of the filter.

Theorem 1.3. *Let $H \in (\frac{1}{2}, 1)$. Assume we observe the Hermite process Z of order q with Hurst parameter H. Then $\hat{H}_N$ is strongly consistent, i.e.*

$$\lim_{N\to\infty} \hat{H}_N = H \ \ a.s.$$

In fact, we have more precisely that $\lim_{N\to\infty}\left(H - \hat{H}_N\right)\log N = 0$ a.s.

Remark 1.2. If we look at the above theorem more carefully, we observe that this is a slightly different notion of consistency than the usual one. In the case of the *mle*, for example, we let N tend to infinity which means that the horizon from which we sample goes to infinity. Here, we do have a fixed horizon $[0, 1]$ and by letting $N \to \infty$ we sample infinitely often. If we had convergence in distribution this would not be an issue, since we could rescale the process appropriately by taking advantage of the self-similarity property, but in terms of almost sure or convergence in probability it is not exactly equivalent.

The next step is to determine the asymptotic distribution of $\hat{H}_N$. Obviously, it should depend on the distribution of the underlying process, and in our case, on q and H. We consider the following three cases separately: fBm (Hermite process of order $q = 1$), Rosenblatt process ($q = 2$), and Hermite processes of higher order $q > 2$. We summarize the results in the following theorem, where the limit notation $X_n \overset{L^2(\Omega)}{\to} X$ denotes convergence in the mean square $\lim_{N\to\infty} \mathbf{E}\left[(X_N - X)^2\right] = 0$, and $\overset{\mathcal{D}}{\to}$ continues to denote convergence in distribution.

Theorem 1.4.

(1) Let $H \in (0, 1)$ and B^H be a fractional Brownian motion with Hurst parameter H.

(a) Assume that we use the filter of order 1.

i. If $H \in (0, \frac{3}{4})$, then as $N \to \infty$

$$\sqrt{N} \log N \frac{2}{\sqrt{c_{1,H}}} \left(\hat{H}_N - H \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{1.19}$$

where $c_{1,H} := 2 + \sum_{k=1}^{\infty} \left(2k^{2H} - (k-1)^{2H} - (k+1)^{2H} \right)^2$.

ii. If $H \in (\frac{3}{4}, 1)$, then as $N \to \infty$

$$N^{1-H} \log N \frac{2}{\sqrt{c_{2,H}}} \left(\hat{H}_N - H \right) \overset{L^2(\Omega)}{\rightarrow} Z^{(2,H)}, \tag{1.20}$$

where $c_{2,H} := \frac{2H^2(2H-1)}{4H-3}$.

iii. If $H = \frac{3}{4}$, then as $N \to \infty$

$$\sqrt{N \ \log N} \frac{2}{\sqrt{c_{3,H}}} \left(\hat{H}_N - H \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{1.21}$$

where $c_{3,H} := \frac{9}{16}$.

(b) Now, let α be of any order $p \geq 2$. Then,

$$\sqrt{N} \log N \frac{1}{c_{6,H}} \left(\hat{H}_N - H \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{1.22}$$

where $c_{6,H} = \frac{1}{2} \sum_{i \in \mathbb{Z}} \rho_H^{\alpha}(i)^2$.

(2) Suppose that $H > \frac{1}{2}$ and the observed process $Z^{2,H}$ is a Rosenblatt process with Hurst parameter H.

(a) If α is a filter of order 1, then

$$N^{1-H} \log N \frac{1}{2c_{4,H}} \left(\hat{H}_N - H \right) \overset{L^2(\Omega)}{\rightarrow} Z^{(2,H)}, \tag{1.23}$$

where $c_{4,H} := 16 d(H)^2$.

(b) If α is a filter of order $p > 1$, then

$$2 c_{7,H}^{-1/2} N^{1-H} \log N \left(\hat{H}_N - H \right) \overset{L^2(\Omega)}{\rightarrow} Z^{(2,H)} \tag{1.24}$$

where

$$c_{7,H} = \frac{64}{c(H)^2} \left(\frac{2H-1}{H \ (H+1)^2} \right) \times$$
$$\left\{ \sum_{q,r=0}^{\ell} b_q b_r \left[|1+q-r|^{2H'} + |1-q+r|^{2H'} - 2|q-r|^{2H'} \right] \right\}^2$$

with $b_q = \sum_{r=0}^{q} \alpha_r$. Here ℓ is the length of the filter, which is related to the order p, see Definition 1.6 and examples following.

(3) *Let* $H \in (\frac{1}{2}, 1)$ *and* $q \in \mathbb{N} \setminus \{0\}$, $q \geq 2$. *Let* $Z^{(q,H)}$ *be a Hermite process of order* q *and self-similarity parameter* H. *Then, for* $H^{'} = 1 + \frac{H-1}{q}$ *and a filter of order 1,*

$$N^{2-2H^{'}} \log N \frac{2}{c_{5,H}} \left(\hat{H}_N(\alpha) - H \right) \overset{L^2(\Omega)}{\to} Z^{(2,2H^{'}-1)}, \tag{1.25}$$

where $c_{5,H} := \frac{4q!d(H)^4(H^{'}(2H^{'}-1))^{2q-2}}{(4H^{'}-3)(4H^{'}-2)}$.

Remark 1.3. In the notation above, $Z^{(2,K)}$ denotes a standard Rosenblatt random variable, which means a random variable that has the same distribution as the Hermite process of order 2 and parameter K at t=1.

Before continuing with a sketch of proof of the theorem, it is important to discuss the theorem's results.

(1) In most of the cases above, we observe that the order of convergence of the estimator depends on H, which is the parameter that we try to estimate. This is not a problem, because it has already been proved, in [6], [7], [27] and [28], that the theorem's convergences still hold when we replace H by $\hat{H}_N$ in the rate of convergence.
(2) The effect of the use of a longer filter is very significant. In the case of fBm, when we use a longer filter, we no longer have the threshold of 3/4 and the suggested estimator is always asymptotically normal. This is important for the following reason: when we start the estimation procedure, we do not know beforehand the value of H and such a threshold would create confusion in choosing the correct rate of convergence in order to scale $\hat{H}_N$ appropriately. Finally, the fact that we have asymptotic normality for all H allows us to construct confidence intervals and perform hypothesis testing and model validation.
(3) Even in the Rosenblatt case the effect of the filter is significant. This is not obvious here, but we will discuss it later in detail. What actually happens is that by filtering the process asymptotic standard error is reduced, i.e. the longer the filter the smaller the standard error.
(4) Finally, one might wonder if the only reason to chose to work with the quadratic variation of the process, instead of a higher order variation (powers higher than 2 in (1.15) and (1.17)), is for the simplicity of calculations. It turns out that there are other, better reasons to do so: Coeurjolly ([8]) proved that the use of higher order variations would lead to higher asymptotic constants and thus to larger standard errors in the case of the fBm. He actually proved that the optimal variation respect to the standard error is the second (*quadratic*).

Proof. [Sketch of proof of Theorem 1.4] We present the key ideas for proving the consistency and asymptotic distribution results. We use a Hermite process of order q and a filter of order 1. However, wherever it is necessary we focus on either fBm or Rosenblatt in order to point out the corresponding differences. The ideas for the proof are very similar in the case of longer filters so the reader should refer to [28] for the details in this approach.

It is convenient to work with the centered normalized quadratic variation defined as

$$V_N := -1 + \frac{1}{N}\sum_{i=0}^{N-1} \frac{\left(Z^{(q,H)}_{\frac{i+1}{N}} - Z^{(q,H)}_{\frac{i}{N}}\right)^2}{N^{-2H}}. \tag{1.26}$$

It is easy to observe that for S_N defined in (1.15),

$$S_N = N^{-2H}\ \left(1 + V_N\right).$$

Using this relation we can see that $\log\left(1+V_N\right) = 2\left(\hat{H}_N - H\right)\log N$, therefore in order to prove consistency it suffices to show that V_N converges to 0 as $N \to \infty$ and the asymptotic distribution of $\hat{H}_N$ depends on the asymptotic behavior of V_N.

By the definition of the Hermite process (1.5), we have that

$$Z^{(q,H)}_{\frac{i+1}{N}} - Z^{(q,H)}_{\frac{i}{N}} = I_q\left(f_{i,N}\right)$$

where we denoted

$$\begin{aligned} f_{i,N}(y_1,\ldots,y_q) &= 1_{[0,\frac{i+1}{N}]}(y_1\vee\ldots\vee y_q)d(H)\int_{y_1\vee\ldots\vee y_q}^{\frac{i+1}{N}} \partial_1 K^{H'}(u,y_1)\ldots\partial_1 K^{H'}(u,y_q)du \\ &\quad - 1_{[0,\frac{i}{N}]}(y_1\vee\ldots\vee y_q)d(H)\int_{y_1\vee\ldots\vee y_q}^{\frac{i}{N}} \partial_1 K^{H'}(u,y_1)\ldots\partial_1 K^{H'}(u,y_q)du. \end{aligned}$$

Now, using the multiplication property (1.11) of multiple Wiener-Itô integrals we can derive a Wiener chaos decomposition of V_N as follows:

$$V_N = T_{2q} + c_{2q-2}T_{2q-2} + \ldots + c_4T_4 + c_2T_2 \tag{1.27}$$

where $c_{2q-2k} := k!\binom{q}{k}^2$ are the combinatorial constants from the product formula for $0 \le k \le q-1$, and

$$T_{2q-2k} := N^{2H-1}I_{2q-2k}\left(\sum_{i=0}^{N-1} f_{i,N}\otimes_k f_{i,N}\right),$$

where $f_{i,N}\otimes_k f_{i,N}$ is the k^{th} contraction of $f_{i,N}$ with itself which is a function of $2q-2k$ parameters.

To determine the magnitude of this Wiener chaos decomposition V_N, we study each of the terms appearing in the decomposition separately. If we compute the L^2

norm of each term, we have

$$\begin{aligned}\mathbf{E}\left[T_{2q-2k}^2\right] &= N^{4H-2}(2q-2k)!\left\|\left(\sum_{i=0}^{N-1} f_{i,N}\otimes_k f_{i,N}\right)^s\right\|^2_{L^2([0,1]^{2q-2k})}\\ &= N^{4H-2}(2q-2k)!\sum_{i,j=0}^{N-1}\langle f_{i,N}\tilde{\otimes}_k f_{i,N}, f_{j,N}\tilde{\otimes}_k f_{j,N}\rangle_{L^2([0,1]^{2q-2k})}\end{aligned}$$

Using properties of the multiple integrals we have the following results

- For $k=q-1$, $\mathbf{E}\left[T_2^2\right]\sim\frac{4d(H)^4(H'(2H'-1))^{2q-2}}{(4H'-3)(4H'-2)}\,N^{2(2H'-2)}$
- For $k=0,\ldots,q-2$

$$\mathbf{E}\left[N^{2(2-2H')}T_{2q-2k}^2\right]=O\left(N^{-2\left(2-2H'\right)2(q-k-1)}\right).$$

Thus, we observe that the term T_2 is the dominant term in the decomposition of the variation statistic V_N. Therefore, with

$$c_{1,1,H}=\frac{4d(H)^4(H'(2H'-1))^{2q-2}}{(4H'-3)(4H'-2)},$$

it holds that

$$\lim_{N\to\infty}\mathbf{E}\left[c_{1,1,H}^{-1}N^{(2-2H')2}c_2^{-2}V_N^2\right]=1.$$

Based on these results we can easily prove that V_N converges to 0 a.s. and then conclude that $\hat{H}_N$ is strongly consistent.

Now, in order to understand the asymptotic behavior of the renormalized sequence V_N it suffices to study the limit of the dominant term

$$I_2\left(N^{2H-1}N^{(2-2H')}\sum_{i=0}^{N-1} f_{i,N}\otimes_{q-1} f_{i,N}\right)$$

When $q=1$ (the fBm case), we can use the Nualart–Ortiz-Latorre criterion (Proposition 1.1) in order to prove convergence to Normal distribution. However, in the general case for $q>1$, using the same criterion, we can see that convergence to a Normal law is no longer true. Instead, a direct method can be employed to determine the asymptotic distribution of the above quantity. Let $N^{2H-1}N^{(2-2H')}\sum_{i=0}^{N-1} f_{i,N}\otimes_{q-1} f_{i,N}=f_2^N+r_2$, where r_2 is a remainder term and

$$\begin{aligned}f_2^N(y,z) &:= N^{2H-1}N^{(2-2H')}d(H)^2a(H')^{q-1}\\ &\sum_{i=0}^{N-1}1_{[0,\frac{i}{N}]}(y\vee z)\int_{I_i}\int_{I_i}dvdu\partial_1K(u,y)\partial_1K(v,z)|u-v|^{(2H'-2)(q-1)}.\end{aligned}$$

It can be shown that the term r_2 converges to 0 in $L^2([0,1]^2)$, while f_2^N converges in $L^2([0,1]^2)$ to the kernel of the Rosenblatt process at time 1, which is by definition

$$(H'(2H'-1))^{(q-1)}d(H)^2 N^{2H-1} N^{2-2H'} \sum_{i=0}^{N-1} \int_{I_i}\int_{I_i} |u-v|^{(2H'-2)(q-1)} \partial_1 K^{H'}(u,y) \partial_1 K^{H'}(v,z).$$

This implies, by the isometry property (1.10) between double integrals and $L^2([0,1]^2)$, that the dominant term in V_N, i.e. the second-chaos term T_2, converges in $L^2(\Omega)$ to the Rosenblatt process at time 1. The reader can consult [6], [7], [27] and [28] for all details of the proof. □

1.5. Comparison & Conclusions

In this section, we compare the estimators described in Sections 2 and 4. The performance measure that we adopt is the *asymptotic relative efficiency*, which we now define according to [24]:

Definition 1.7. Let T_n be an estimator of θ for all n and $\{\alpha_n\}$ a sequence of positive numbers such that $\alpha_n \to +\infty$ or $\alpha_n \to \alpha > 0$. Suppose that for some probability law Y with positive and finite second moment,

$$\alpha_n (T_n - \theta) \xrightarrow{\mathcal{D}} Y,$$

(i) The *asymptotic mean square error* of T_n ($amse_{T_n}(\theta)$) is defined to be the asymptotic expectation of $(T_n-\theta)^2$, i.e.

$$amse_{T_n}(\theta) = \frac{EY^2}{\alpha_n}.$$

(ii) Let $T_n^{'}$ be another estimator of θ. The *asymptotic relative efficiency* of $T_n^{'}$ with respect to T_n is defined to be

$$e_{T_n,T_n'}(\theta) = \frac{amse_{T_n}(\theta)}{amse_{T_n'}(\theta)}. \tag{1.28}$$

(iii) T_n is said to be *asymptotically more efficient* than $T_n^{'}$ if and only if

$$\limsup_n e_{T_n,T_n'}(\theta) \le 1, \text{ for all } \theta \text{ and}$$
$$\limsup_n e_{T_n,T_n'}(\theta) < 1, \text{ for some } \theta.$$

Remark 1.4. These definitions are in the most general setup: indeed (i) they are not restricted by the usual assumption that the estimators converge to a Normal distribution; moreover, (ii) the asymptotic distributions of the estimators do not have to be the same. This will be important in our comparison later.

Our comparative analysis focuses on fBm and the Rosenblatt process, since the maximum likelihood and the wavelets methods cannot be applied to higher order Hermite processes.

1.5.1. *Variations estimator vs. mle*

We start with the case of a filter of order 1 for fBm. Since the asymptotic behavior of the variations estimator depends on the true value of H we consider three different cases:

- If $H \in (0, 3/4)$, then

$$e_{\hat{H}_N(\alpha),\ \hat{H}_{mle}}(H) = \frac{\frac{2+\sum_{k=1}^{\infty}(2k^{2H}-(k-1)^{2H}-(k+1)^{2H})^2}{2\sqrt{N}\log N}}{\frac{[2\ D(H)]^{-1}}{\sqrt{N}}} \approx \frac{1}{\log N}.$$

  This implies that

$$\limsup e_{\hat{H}_N(\alpha),\hat{H}_{mle}}(H) = 0,$$

  meaning that $\hat{H}_N(\alpha)$ is asymptotically more efficient than $\hat{H}_{mle}$.
- If $H \in (3/4, 1)$, then

$$e_{\hat{H}_N(\alpha),\ \hat{H}_{mle}}(H) = \frac{\frac{\frac{2H^2(2H-1)}{4H-3}}{4\ N^{1-H}\log N}}{\frac{[2\ D(H)]^{-1}}{\sqrt{N}}} \approx \frac{N^{H-1/2}}{\log N}$$

  This implies that

$$\limsup e_{\hat{H}_N(\alpha),\ \hat{H}_{mle}}(H) = \infty,$$

  meaning that $\hat{H}_{mle}$ is asymptotically more efficient than $\hat{H}_N(\alpha)$.
- If $H = 3/4$, then

$$e_{\hat{H}_N(\alpha),\ \hat{H}_N^{mle}}(H) = \frac{\frac{\frac{16}{9}}{4\sqrt{N}\log N}}{\frac{[2\ D(H)]^{-1}}{\sqrt{N}}} \approx \frac{1}{\sqrt{\log N}}$$

  Similarly, as in the first scenario the variations estimator is asymptotically more efficient than the *mle*.

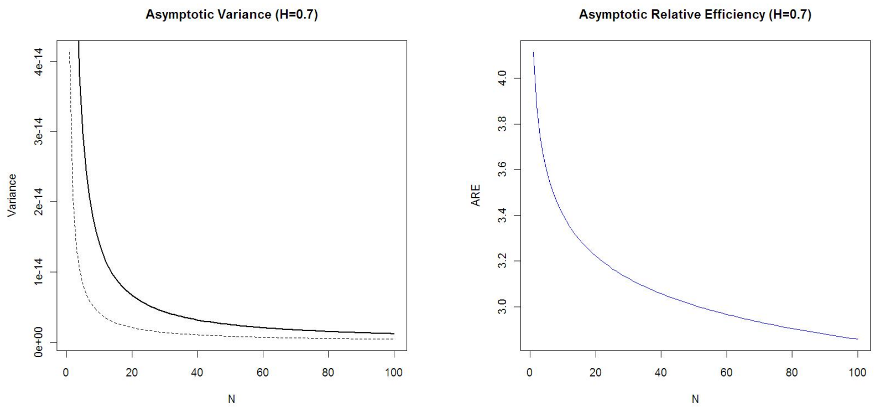
Remark 1.5. By $\hat{H}_{mle}$ we mean either the exact *mle* or the Whittle approximate *mle*, since both have the same asymptotic distribution.

Before discussing the above results let us recall the Cramér-Rao Lower Bound theory (see [24]). Let $X = (X_1, \ldots, X_N)$ be a *sample* (i.e. identically distributed random variables) with common distribution P_H and corresponding density function f_H. If T is an estimator of H such that $E(T) = H$, then

$$Var(T) \geq [I(H)]^{-1} \tag{1.29}$$

where $I(H)$ is the *Fisher information* defined by

$$I(H) := \mathbf{E}\left\{\left[\frac{\partial}{\partial H}\log f_H(X)\right]^2\right\}. \tag{1.30}$$



$\hat{H}_{mle}$ *(dotted)*, $\hat{H}_N$ *(bold)* *Asymptotic Relative Efficiency*

Fig. 1.3. Comparison of the variations' estimator and *mle* for a filter of order p =1.

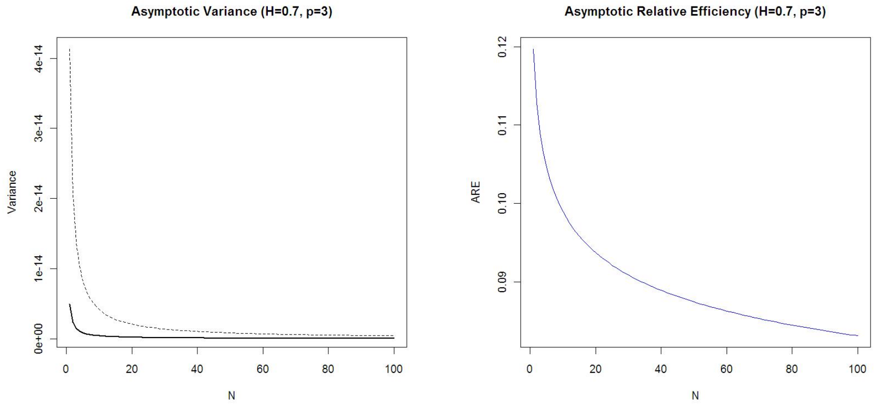
The inverse of the Fisher information is called *Cramér-Rao Lower Bound*.

Remark 1.6. It has been proved by Dahlhaus, [9], that the asymptotic variance of both the approximate and exact *mle* converges to the Cramér-Rao Lower Bound and consequently both estimators are asymptotically efficient according to the Fisher criterion. Thus how can the variations estimator be more efficient in some cases?

The variations-based estimator is computed using data coming from a fixed time horizon and more specifically $[0, 1]$, i.e. data such as $X_a = (X_0, X_{\frac{1}{N}}, \ldots, X_1)$, while the *mle* is computed using data of the form $X_b = (X_0, X_1, \ldots, X_N)$. The time-scaling makes a big difference since the vectors X_a and X_b **do not have the same distribution**. The construction of the Fisher information (and accordingly the asymptotic Cramér-Rao Lower Bound) depends on the underlying distribution of the sample and it is going to be different for X_a and X_b. This implies that the Cramér-Rao Lower Bound attained by the *mle* using X_b is not the same as the Cramér-Rao Lower Bound attained by the *mle* using X_a. By the self-similarity property we can derive that $X_a =^{\mathcal{D}} N^H X_b$, which indicates that if we want to compute the information matrix for the rescaled data, the scaling contains the parameter H and this will alter the information matrix and its rate of convergence.

We begin by observing what happens in practice for a filter of order 1. In the following graphs, we compare the corresponding variations estimator with the mle, in the case of a simulated fractional Brownian motion with $H = 0.65$, by plotting the asymptotic variance against the sample size N.

As we observe in Figure 1.3, the *mle* performs better than the estimator based on the variations of the process with filter order $p = 1$. It has a smaller asymptotic variance and the asymptotic relative efficiency seems to converge to zero extremely slowly, even for a very large sample size N. This is because $\hat{H}_N$ is faster only by



$\hat{H}_{mle}$ *(dotted)*, $\hat{H}_N$ *(bold)* *Asymptotic Relative Efficiency*

Fig. 1.4. Comparison of the variations' estimator and *mle* for a filter of order $p = 10$.

the factor $\log N$ (which is quite slow) and the constants in the case of $\hat{H}_N$ are quite large.

Let us consider now the case of longer filters ($p \geq 2$). Using the results proved in the previous sections (esp. Theorem 1.4 part (1.b)), we have that for all $H \in (0, 1)$

$$e_{\hat{H}_N(\alpha),\, \hat{H}_N^{mle}}(H) \approx \frac{1}{\log N}$$

and from this we conclude that the variations estimator is always asymptotically more efficient than the *mle*. If we do the same plots as before we can see (Figure 1.4) that the constant is now significantly smaller.

1.5.2. *Variations' vs. Wavelet Estimator*

In this subsection we compare the variations and the wavelets estimators for the both the fBm and the Rosenblatt process.

fBm :

(1) Let $0 < H < 3/4$, then for a filter of order $p \geq 1$ in the variations estimator, and for any $Q \geq 1$ in the wavelets estimator, we have

$$e_{\hat{H}_N(\alpha),\, \hat{H}_{wave}}(H) \approx \frac{1}{\sqrt{\alpha(N)}\log N}.$$

Based on the properties of $\alpha(N)$ as stated before (Theorem 1.2), we conclude that

$$\lim_{N \to 0} e_{\hat{H}_N(\alpha),\, \hat{H}_{mle}}(H) = 0,$$

which implies that the variations estimator is asymptotically more efficient than the wavelets estimator.

(2) When $3/4 < H < 1$, then for a filter of order $p = 1$ in the variations estimator, and $Q = 1$ for the wavelets estimator, we have

$$e_{\hat{H}_N(\alpha),\, \hat{H}_{wave}}(H) \approx \frac{N^{2-2H}}{\alpha(N)^{2-2H}\log N}$$

If we choose $\alpha(N)$ to be the *optimal* as suggested by Bardet and Tudor in [1] , i.e. $\alpha(N) = N^{1/2+\delta}$ for δ small, then $e_{\hat{H}_N(\alpha),\, \hat{H}_{wave}}(H) \approx \frac{N^{(1-H)(1-2\delta)}}{\log N}$, which implies that the wavelet estimator performs better.

(3) When $3/4 < H < 1$, then for a filter of order $p \geq 2$ in the variations estimator and $Q = 1$ for the wavelets estimator, using again the optimal choice of $\alpha(N)$ as proposed in [1], we have

$$e_{\hat{H}_N(\alpha),\, \hat{H}_{wave}}(H) \approx \frac{N^{(\frac{1}{2}-H)-2\delta(1-H)}}{\log N},$$

so the variations estimator is asymptotically more efficient than the wavelets one.

Rosenblatt process :

Suppose that $1/2 < H < 1$, then for any filter of any order $p \geq 1$ in the variations estimator, and any $Q \geq 1$ for the wavelets based estimator, we have

$$e_{\hat{H}_N(\alpha),\, \hat{H}_{wave}}(H) \approx \frac{1}{\alpha(N)^{1-H}\log N}.$$

Again, with the behavior of $\alpha(N)$ as stated in Theorem 1.2, we conclude that the variations estimator is asymptotically more efficient than the wavelet estimator.

Overall, it appears that the estimator based on the discrete variations of the process is asymptotically more efficient than the estimator based on wavelets, in most cases. The wavelets estimator does not have the problems of computational time which plague the *mle*: using efficient techniques, such as Mallat's algorithm, the wavelets estimator takes seconds to compute on a standard PC platform. However, the estimator based on variations is much simpler, since it can be constructed by simple transformation of the data.

Summing up, the conclusion is that the heuristic approaches (R/S, variograms, correlograms) are useful for a preliminary analysis to determine whether long memory may be present, due to their simplicity and universality. However, in order to estimate the Hurst parameter it would be preferable to use any of the other techniques. Overall, the estimator based on the discrete variations is asymptotically more efficient than the estimator based on wavelets or the *mle*. Moreover, it can be applied not only when the data come from a fractional Brownian motion, but also when they come from any other non-Gaussian Hermite process of higher order.

Finally, when we apply a longer filter in the estimation procedure, we are able to reduce the asymptotic variance and consequently the standard error significantly.

The benefits of using longer filters needs to be investigated further. It would be interesting to study the choice of different types of filters, such as wavelet-type filters versus finite difference filters. Specifically, the complexity introduced by the construction of the estimator based on a longer filter, which is not as straightforward as in the case of filter of order 1, is something that will be investigated in a subsequent article.

References

[1]J-M. Bardet and C.A. Tudor (2008): *A wavelet analysis of the Rosenblatt process: chaos expansion and estimation of the self-similarity parameter.* Preprint.

[2]J.B. Bassingthwaighte and G.M. Raymond (1994): Evaluating rescaled range analysis for time series, *Annals of Biomedical Engineering*, **22**, 432-444.

[3] J. Beran (1994): *Statistics for Long-Memory Processes.* Chapman and Hall.

[4]J.-C. Breton and I. Nourdin (2008): Error bounds on the non-normal approximation of Hermite power variations of fractional Brownian motion. *Electronic Communications in Probability*, **13**, 482-493.

[5] P. Breuer and P. Major (1983): Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Analysis*, **13** (3), 425-441.

[6] A. Chronopoulou, C.A. Tudor and F. Viens (2009): Application of Malliavin calculus to long-memory parameter estimation for non-Gaussian processes. *Comptes rendus - Mathematique* **347**, 663-666.

[7] A. Chronopoulou, C.A. Tudor and F. Viens (2009): *Variations and Hurst index estimation for a Rosenblatt process using longer filters.* Preprint.

[8] J.F. Coeurjolly (2001): Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Statistical Inference for Stochastic Processes*, **4**, 199-227.

[9] R. Dahlhaus (1989): Efficient parameter estimation for self-similar processes. *Annals of Statistics*,**17**, 1749-1766.

[10] R.L. Dobrushin and P. Major (1979): Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **50**, 27-52.

[11] P. Flandrin (1993): Fractional Brownian motion and wavelets. *Wavelets, Fractals and Fourier transforms.* Clarendon Press, Oxford, 109-122.

[12] R. Fox, M. Taqqu (1985): Non-central limit theorems for quadratic forms in random variables having long-range dependence.*Probab. Th. Rel. Fields*, **13**, 428-446.

[13] Hurst, H. (1951): Long Term Storage Capacity of Reservoirs, *Transactions of the American Society of Civil Engineers*, **116**, 770-799.

[14] A.K. Louis, P. Maass, A. Rieder (1997): *Wavelets: Theory and applications* Pure & Applied Mathematics. Wiley-Interscience series of texts, monographs & tracts.

[15] M. Maejima and C.A. Tudor (2007): Wiener integrals and a Non-Central Limit Theorem for Hermite processes, *Stochastic Analysis and Applications*, **25** (5), 1043-1056.

[16] B.B. Mandelbrot (1975): Limit theorems of the self-normalized range for weakly and strongly dependent processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **31**, 271-285.

[17] I. Nourdin, D. Nualart and C.A Tudor (2007): *Central and Non-Central Limit Theorems for weighted power variations of the fractional Brownian motion.* Preprint.

[18] I. Nourdin, G. Peccati and A. Réveillac (2008): Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. H. Poincaré Probab. Statist.*, 18 pages, to appear.

[19] D. Nualart (2006): *Malliavin Calculus and Related Topics. Second Edition.* Springer.

[20] D. Nualart and G. Peccati (2005): Central limit theorems for sequences of multiple stochastic integrals. *The Annals of Probability*, **33**, 173-193.

[21] D. Nualart and S. Ortiz-Latorre (2008): Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Processes and their Applications,* **118**, 614-628.

[22] G. Peccati and C.A. Tudor (2004): Gaussian limits for vector-valued multiple stochastic integrals. *Séminaire de Probabilités*, **XXXIV**, 247-262.

[23] G. Samorodnitsky and M. Taqqu (1994): *Stable Non-Gaussian random variables.* Chapman and Hall, London.

[24] J. Shao (2007): *Mathematical Statistics.* Springer.

[25] M. Taqqu (1975): Weak convergence to the fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **31**, 287-302.

[26] C.A. Tudor (2008): Analysis of the Rosenblatt process. *ESAIM Probability and Statistics*, **12**, 230-257.

[27] C.A. Tudor and F. Viens (2008): Variations and estimators through Malliavin calculus. *Annals of Probability*, 37 pages, to appear.

[28] C.A. Tudor and F. Viens (2008): Variations of the fractional Brownian motion via Malliavin calculus. *Australian Journal of Mathematics*, 13 pages, to appear.

[29] P. Whittle (1953): Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-434.