Research Article

Asymptotic Normality of a Hurst Parameter Estimator Based on the Modified Allan Variance

Alessandra Bianchi,[1] Massimo Campanino,[2] and Irene Crimaldi[3]

[1] *Department of Pure and Applied Mathematics, University of Padua, Via Trieste 63, 35121 Padova, Italy*

[2] *Department of Mathematics, University of Bologna, Piazza di Porta S. Donato 5, 40126 Bologna, Italy*

[3] *IMT Institute for Advanced Studies, Piazza S. Ponziano 6, 55100 Lucca, Italy*

Correspondence should be addressed to Irene Crimaldi, irene.crimaldi@imtlucca.it

In order to estimate the memory parameter of Internet traffic data, it has been recently proposed a log-regression estimator based on the so-called modified Allan variance (MAVAR). Simulations have shown that this estimator achieves higher accuracy and better confidence when compared with other methods. In this paper we present a rigorous study of the MAVAR log-regression estimator. In particular, under the assumption that the signal process is a fractional Brownian motion, we prove that it is consistent and asymptotically normally distributed. Finally, we discuss its connection with the wavelets estimators.

1. Introduction

It is well known that different kinds of real data (hydrology, telecommunication networks, economics, and biology) display self-similarity and long-range dependence (LRD) on various time scales. By self-similarity we refer to the property that a dilated portion of a realization has the same statistical characterization as the original realization. This can be well represented by a self-similar random process with a given scaling exponent H (*Hurst parameter*). The long-range dependence, also called long memory, emphasizes the long-range time correlation between past and future observations and it is thus commonly equated to an asymptotic power law behaviour of the spectral density at low frequencies or, equivalently, to an asymptotic power-law decrease of the autocovariance function, of a given stationary random process. In this situation, the memory parameter of the process is given by the exponent d characterizing the power law of the spectral density. (For a review of historical and statistical aspects of the self-similarity and the long memory see [1].)

Though a self-similar process cannot be stationary (and thus nor LRD), these two properties are often related in the following sense. Under the hypothesis that a self-similar process has stationary (or weakly stationary) increments, the scaling parameter H enters in the description of the spectral density and covariance function of the increments, providing an asymptotic power law with exponent $d = 2H - 1$. Under this assumption, we can say that the self-similarity of the process reflects on the long-range dependence of its increments. The most paradigmatic example of this connection is provided by the fractional Brownian motion and by its increment process, the fractional Gaussian noise [2].

In this paper we will consider the problem of estimating the Hurst parameter H of a self-similar process with weakly stationary increments. Among the different techniques introduced in the literature, we will focus on a method based on the log-regression of the so-called modified Allan variance (MAVAR). The MAVAR is a well-known time-domain quantity generalizing the classic Allan variance [3, 4], which has been proposed for the first time as a traffic analysis tool in [5]. In a series of paper [5–7], its performance has been evaluated by simulation and comparison with the real IP traffic. These works have pointed out the high accuracy of the method in estimating the parameter H and have shown that it achieves better confidence if compared with the well-established wavelet log-diagram.

The aim of the present work is to substantiate and enrich these results from the theoretical point of view, studying the rate of convergence of the estimator toward the memory parameter. In particular, our goal is to provide the limit properties and the precise asymptotic normality of the MAVAR log-regression estimator in order to compute the related asymptotic confidence intervals. This will be reached under the assumption that the signal process is a fractional Brownian motion. Although this hypothesis may look restrictive (indeed this estimator is designed for a larger class of stochastic processes), the obtained results are a first step toward the mathematical study of the MAVAR log-regression estimator. To our knowledge, there are no similar results in the literature. The present paper also provides the theoretical foundations (mathematical details and proofs) for [8]. Indeed the formulas analytically obtained here have been implemented and numerically tested in [8] for different choices of the regression weights and it has been shown that the numerical evidences are in good agreement with the theoretical results proven here.

For a survey on Hurst parameter estimators of a fractional Brownian motion, we refer to [9, 10]. However we stress once again that the MAVAR-estimator is not specifically a target for the fractional Brownian motion, but it has been thought and successfully used for more general processes.

Our theorems can be viewed as a counterpart of the already established results concerning the asymptotic normality of the wavelet log-regression estimators [11–13]. Indeed, although the MAVAR can be related in some sense to a suitable Haar-type wavelets family (see [14] for the classical Allan variance), the MAVAR and wavelets log-regression estimators do not match as the regression runs on different parameters (see Section 5). Hence, we adopt a different argument which in turns allows us to avoid some technical troubles due to the poor regularity of the Haar-type functions.

The paper is organized as follows. In Section 2 we recall the properties of self-similarity and long-range dependence for stochastic processes and the definition of fractional Brownian motion; in Section 3 we introduce the MAVAR and its estimator, with their main properties; in Section 4 we state and prove the main results concerning the asymptotic normality of the estimator; in Section 5 we make some comments on the link between the MAVAR and the wavelet estimators and on the modified Hadamard variance, which is a generalization of the MAVAR; in the Appendix we recall some results used along the proof.

2. Self-Similarity and Long-Range Dependence

In the sequel we consider a centered real-valued stochastic process $X = \{X(t), t \in \mathbb{R}\}$, that can be interpreted as the signal process. Sometimes it is also useful to consider the τ-increment of the process X, which is defined, for every $\tau > 0$ and $t \in \mathbb{R}$, as

$$Y_\tau(t) = \frac{X(t + \tau) - X(t)}{\tau}. \tag{2.1}$$

In order to reproduce the behavior of the real data, it is commonly assumed that X satisfies one of the two following properties: (i) self-similarity; (ii) long range dependence.

(i) The *self-similarity* of a centered real-valued stochastic process X, starting from zero, refers to the existence of a parameter $H \in (0, 1)$, called *Hurst index or Hurst parameter* of the process, such that, for all $a > 0$, it holds

$$\{X(t), t \in \mathbb{R}\} \stackrel{d}{=} \left\{ a^{-H} X(at), t \in \mathbb{R} \right\}. \tag{2.2}$$

In this case we say that X is an H-self-similar process.

(ii) We first recall that a centered real-valued stochastic process X is *weakly stationary* if it is square integrable and its autocovariance function, $C_X(t, s) := \operatorname{Cov}(X(t), X(s))$, is a translation invariant, namely, if

$$C_X(t, s) = C_X(t + r, s + r), \quad \forall t, s, r \in \mathbb{R}. \tag{2.3}$$

In this case we also set $R_X(t) := C_X(t, 0)$. If X is a weakly stationary process, we say that it displays a *long-range dependence, or long memory*, if there exists $d \in (0, 1)$ such that the spectral density of the process, $f_X(\lambda)$, satisfies the condition

$$f_X(\lambda) \sim c_f |\lambda|^{-d} \quad \text{as } \lambda \longrightarrow 0, \tag{2.4}$$

for some finite constant $c_f \neq 0$, where we write $f(x) \sim g(x)$ as $x \to x_0$, if $\lim_{x \to x_0}(f(x)/g(x)) = 1$. Due to the correspondence between the spectral density and the autocovariance function, given by

$$R_X(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{it\lambda} f_X(\lambda) \, d\lambda, \tag{2.5}$$

the long-range condition (2.4) can be often stated in terms of the autocovariance of the process as

$$R_X(t) \sim c_R |t|^{-\beta} \quad \text{as } |t| \longrightarrow +\infty, \tag{2.6}$$

for $\beta = (1 - d) \in (0, 1)$ and some finite constant $c_R \neq 0$.

Notice that if X is a self-similar process, then it obviously cannot be weakly stationary. On the other hand, assuming that X starts from zero and it is as H-self-similar process with weakly stationary increments, that is, the quantity

$$\mathbb{E}[(X(\tau_2 + s + t) - X(s + t))(X(\tau_1 + s) - X(s))] \tag{2.7}$$

does not depend on s, it turns out that the autocovariance function is given by

$$C_X(s,t) = \frac{\sigma_H^2}{2}\left(|t|^{2H} - |t - s|^{2H} + |s|^{2H}\right), \tag{2.8}$$

with $\sigma_H^2 := \mathbb{E}[X^2(1)]$, which is clearly not a translation invariant. Consequently, denoting by Y_τ its τ-increment process (see (2.1)), the autocovariance function of Y_τ is such that ([15])

$$R_{Y_\tau}(t) \sim \sigma_H^2 H(2H - 1)|t|^{2H-2} \quad \text{as } |t| \longrightarrow +\infty. \tag{2.9}$$

In particular, if $H \in (1/2, 1)$, the process Y_τ displays a long-range dependence in the sense of (2.6) with $\beta = 2 - 2H$. Under this assumption, we thus embrace the two main empirical properties of a wide collection of real data.

A basic example of the connection between self-similarity and long-range dependence is provided by the *fractional Brownian motion* $B_H = \{B_H(t), t \in \mathbb{R}\}$ [2]. This is a centered Gaussian process, starting from zero, with autocovariance function given by (2.8), where [16]

$$\sigma_H^2 = \frac{1}{\Gamma(2H + 1)\sin(\pi H)}. \tag{2.10}$$

It can be proven that B_H is a self-similar process with Hurst index $H \in (0, 1)$, which corresponds, for $H = 1/2$, to the standard Brownian motion. Moreover, its increment process

$$G_{\tau,H}(t) = \frac{B_H(t + \tau) - B_H(t)}{\tau}, \tag{2.11}$$

called fractional Gaussian noise, turns out to be a weakly stationary Gaussian process [2, 17].

In the next sections we will perform the analysis of the modified Allan variance and of the related estimator of the memory parameter.

3. The Modified Allan Variance

In this section we introduce and recall the main properties of the modified Allan variance (MAVAR) and of the log-regression estimator of the memory parameter based on it.

Suppose that X is a centered real-valued stochastic process, starting from zero, with weakly stationary increments. Let $\tau_0 > 0$ be the "sampling period" and define the sequence of times $\{t_k\}_{k \geq 1}$ taking $t_1 \in \mathbb{R}$ and setting $t_i - t_{i-1} = \tau_0$, that is, $t_i = t_1 + \tau_0(i - 1)$.

Definition 3.1. For any fixed integer $p \geq 1$, the modified Allan variance (MAVAR) is defined [3] as

$$\sigma_p^2 = \sigma^2(\tau_0, p) := \frac{1}{2\tau_0^2 p^2} \mathbb{E}\left[\left(\frac{1}{p}\sum_{j=1}^{p} X(t_{j+2p}) - 2X(t_{j+p}) + X(t_j)\right)^2\right]$$

$$= \frac{1}{2\tau^2}\mathbb{E}\left[\left(\frac{1}{p}\sum_{j=1}^{p} X(t_j + 2\tau) - 2X(t_j + \tau) + X(t_j)\right)^2\right],$$ (3.1)

where we set $\tau := \tau_0 p$. For $p = 1$ we recover the well-known Allan variance.

Let us assume that a finite sample $X_1, \ldots, X_n$ of the process X is given, and that the observations are taken at times $t_1, \ldots, t_n$. In other words we set $X_i = X(t_i)$ for $i = 1, \ldots, n$.

A standard estimator for the modified Allan variance (MAVAR estimator) is given by

$$\widehat{\sigma}_p^2(n) = \widehat{\sigma}^2(\tau_0, p, n) := \frac{1}{2\tau_0^2 p^4 n_p} \sum_{r=1}^{n_p}\left(\sum_{j'=r}^{p+r-1}(X_{j'+2p} - 2X_{j'+p} + X_{j'})\right)^2$$

$$= \frac{1}{n_p}\sum_{k=0}^{n_p-1}\left(\frac{1}{\sqrt{2}\tau_0 p^2}\sum_{j=1}^{p}(X_{k+j+2p} - 2X_{k+j+p} + X_{k+j})\right)^2,$$ (3.2)

for $p \in \{1, \ldots, \lfloor n/3 \rfloor\}$ and $n_p := n - 3p + 1$.

For $k \in \mathbb{Z}$, let us set

$$d_{p,k} = d(\tau_0, p, k) := \frac{1}{\sqrt{2}\,\tau_0 p^2}\sum_{j=1}^{p}(X_{k+j+2p} - 2X_{k+j+p} + X_{k+j}),$$ (3.3)

so that the process $\{d_{p,k}\}_k$ turns out to be weakly stationary for each fixed p, with $\mathbb{E}[d_{p,k}] = 0$, and we can write

$$\sigma_p^2 = \mathbb{E}\left[d_{p,0}^2\right] = \mathbb{E}\left[d_{p,k}^2\right],$$

$$\widehat{\sigma}_p^2(n) = \frac{1}{n_p}\sum_{k=0}^{n_p-1} d_{p,k}^2.$$ (3.4)

3.1. Some Properties

Let us further assume that X is an H-self-similar process (see (2.2)), with $H \in (1/2, 1)$. Applying the covariance formula (2.8), it holds

$$\sigma_p^2 = \mathbb{E}\left[d_{p,0}^2\right] = \sigma_H^2 \tau^{2H-2} K(H, p) = \sigma_H^2 \frac{K(H, p)}{\tau^{2-2H}},$$ (3.5)

with $\sigma_H^2 = \mathbb{E}[X(1)^2]$ and

$$K(H,p) := \frac{2}{p}\left(1 - 2^{2H-2}\right) + \frac{1}{2p^2}\sum_{\ell=1}^{p-1}\sum_{h=1}^{\ell}P_{2H}\left(\frac{h}{p}\right), \tag{3.6}$$

where P_{2H} is the polynomial of degree $2H$ given by

$$P_{2H}(x) := \left[-6|x|^{2H} + 4|1+x|^{2H} + 4|1-x|^{2H} - |2+x|^{2H} - |2-x|^{2H}\right]. \tag{3.7}$$

Since we are interested in the limit for $p \to \infty$, we consider the approximation of the two finite sums in (3.6) by the corresponding double integral, namely,

$$\frac{1}{p^2}\sum_{\ell=1}^{p-1}\sum_{h=1}^{\ell}P_{2H}\left(\frac{h}{p}\right) = \int_0^1\int_0^y P_{2H}(x)dx\,dy + O_H\left(p^{-1}\right). \tag{3.8}$$

Computing the integral and inserting the result in (3.6), we get

$$K(H,p) = K(H) + O_H\left(p^{-1}\right), \tag{3.9}$$

where

$$K(H) := \frac{2^{2H+4} + 2^{2H+3} - 3^{2H+2} - 15}{2(2H+1)(2H+2)}. \tag{3.10}$$

From (3.5) and (3.9), we get

$$\left|\sigma_p^2 - \sigma_H^2\tau^{2H-2}K(H)\right| \leq \sigma_H^2\tau^{2H-2}O_H\left(p^{-1}\right). \tag{3.11}$$

Under the above hypotheses on X, one can also prove that the process $\{d_{p,k}\}_{p,k}$ satisfies the stationary condition

$$\text{Cov}(d_{p,k}, d_{p-u,k'}) = \text{Cov}(d_{p,k-k'}, d_{p-u,0}) \quad \text{for } 0 \leq u < p. \tag{3.12}$$

To verify this condition, we write explicitly the covariance as

$$\begin{aligned}
\mathbb{E}[d_{p,k}d_{p-u,k'}] = {} &\frac{1}{2\tau^2(p-u)^2}\sum_{j=1}^{p}\sum_{j'=1}^{p-u}\mathbb{E}\left[(X_{k+j+2p} - X_{k+j+p})(X_{k'+j'+2(p-u)} - X_{k'+j'+(p-u)})\right] \\
&- \mathbb{E}\left[(X_{k+j+2p} - X_{k+j+p})(X_{k'+j'+(p-u)} - X_{k'+j'})\right] \\
&- \mathbb{E}\left[(X_{k+j+p} - X_{k+j})(X_{k'+j'+2(p-u)} - X_{k'+j'+(p-u)})\right] \\
&+ \mathbb{E}\left[(X_{k+j+p} - X_{k+j})(X_{k'+j'+(p-u)} - X_{k'+j'})\right].
\end{aligned} \tag{3.13}$$

Setting

$$R(t; \tau_1, \tau_2) := \mathbb{E}[(X(\tau_2 + s + t) - X(s + t))(X(\tau_1 + s) - X(s))],$$

$$r(q, h, p, u) := R(\tau_0(q + h + u); \tau_0(p - u), \tau_0 p)$$

$$- R(\tau_0(q + h + p); \tau_0(p - u), \tau_0 p) \qquad (3.14)$$

$$- R(\tau_0(q + h - (p - u)); \tau_0(p - u), \tau_0 p)$$

$$+ R(\tau_0(q + h); \tau_0(p - u), \tau_0 p),$$

we get

$$\mathbb{E}[d_{p,k} d_{p-u,k'}] = \frac{1}{2} \frac{1}{\tau^2} \frac{1}{(p - u)^2} \sum_{j=1}^{p} \sum_{j'=1}^{p-u} r(k - k', j - j', p, u). \qquad (3.15)$$

This immediately provides the stationary condition (3.12).

To better understand the behavior of the covariance $\mathbb{E}[d_{p,k} d_{p-u,k'}]$ as $k - k'$ varies, we apply the covariance formula (2.8) and get

$$R(t; \tau_1, \tau_2) = \frac{\sigma_H^2}{2} \left(|t + \tau_2|^{2H} + |t - \tau_1|^{2H} - |t + \tau_2 - \tau_1|^{2H} - |t|^{2H} \right). \qquad (3.16)$$

Thus, from (3.14),

$$r(q, h, p, u) = \sigma_H^2 \tau^{2H} \sum_{i=1}^{3} g_H^{(i)} \left(\frac{q}{p}, \frac{h}{p}, \frac{u}{p} \right), \qquad (3.17)$$

where

$$g_H^{(1)} \left(\frac{q}{p}, \frac{h}{p}, \frac{u}{p} \right) := \left| \frac{q + h}{p} + \frac{u}{p} + 1 \right|^{2H} + \left| \frac{q + h}{p} + \frac{u}{p} - 1 \right|^{2H} - 2 \left| \frac{q + h}{p} + \frac{u}{p} \right|^{2H},$$

$$g_H^{(2)} \left(\frac{q}{p}, \frac{h}{p}, \frac{u}{p} \right) := -\frac{1}{2} \left(\left| \frac{q + h}{p} + 2 \right|^{2H} + \left| \frac{q + h}{p} \right|^{2H} - 2 \left| \frac{q + h}{p} + 1 \right|^{2H} \right), \qquad (3.18)$$

$$g_H^{(3)} \left(\frac{q}{p}, \frac{h}{p}, \frac{u}{p} \right) := -\frac{1}{2} \left(\left| \frac{q + h}{p} + \frac{2u}{p} \right|^{2H} + \left| \frac{q + h}{p} + \frac{2u}{p} - 2 \right|^{2H} - 2 \left| \frac{q + h}{p} + \frac{2u}{p} - 1 \right|^{2H} \right).$$

Inserting (3.17) in (3.15), we obtain

$$\mathbb{E}[d_{p,k} d_{p-u,k'}] = \sigma_H^2 \tau^{2H-2} G_H \left(\frac{(k - k')}{p}, \frac{u}{p}, p \right), \qquad (3.19)$$

where $\tau = \tau_0 p$ as before, and

$$G_H\left(\frac{q}{p}, \frac{u}{p}, p\right) := \frac{1}{2} \frac{1}{(p-u)^2} \sum_{j=1}^{p} \sum_{j'=1}^{p-u} \sum_{i=1}^{3} g_H^{(i)}\left(\frac{q}{p}, \frac{(j-j')}{p}, \frac{u}{p}\right). \qquad (3.20)$$

Now we set $q = mp + r$ with $m \in \mathbb{Z}$ and r an integer number in $\{0, \ldots, p-1\}$ and we study the asymptotic behavior of $G_H(m + (r/p), u/p, p)$ for $|m| \to +\infty$.

The limit relation

$$|m+1|^{2H} + |m-1|^{2H} - 2|m|^{2H} \sim 2H(2H-1)|m|^{2H-2} \quad \text{as } |m| \longrightarrow +\infty, \qquad (3.21)$$

that is,

$$\lim_{|m| \to +\infty} \frac{|m+1|^{2H} + |m-1|^{2H} - 2|m|^{2H}}{|m|^{2H-2}} = 2H(2H-1), \qquad (3.22)$$

implies that

$$g_H^{(1)}\left(m + \left(\frac{r}{p}\right), \frac{h}{p}, \frac{u}{p}\right) \sim 2H(2H-1)\left|m + \frac{r}{p} + \frac{h}{p} + \frac{u}{p}\right|^{2H-2} \quad \text{as } |m| \longrightarrow +\infty,$$

$$g_H^{(2)}\left(m + \left(\frac{r}{p}\right), \frac{h}{p}, \frac{u}{p}\right) \sim -H(2H-1)\left|m + \frac{r}{p} + \frac{h}{p} + 1\right|^{2H-2} \quad \text{as } |m| \longrightarrow +\infty, \qquad (3.23)$$

$$g_H^{(3)}\left(m + \left(\frac{r}{p}\right), x, \frac{u}{p}\right) \sim -H(2H-1)\left|m + \frac{r}{p} + \frac{h}{p} + \frac{2u}{p} - 1\right|^{2H-2} \quad \text{as } |m| \longrightarrow +\infty,$$

and so, for $|m| \to +\infty$,

$$\sum_{i=1}^{3} g_H^{(i)}\left(m + \left(\frac{r}{p}\right), \frac{h}{p}, \frac{u}{p}\right)$$

$$\sim -H(2H-1)\left(\left|m + \frac{r+h}{p} + 1\right|^{2H-2} + \left|m + \frac{r+h+2u}{p} - 1\right|^{2H-2} - 2\left|m + \frac{r+h+u}{p}\right|^{2H-2}\right)$$

$$= -H(2H-1)$$

$$\times \left(\left|m + \frac{r+u+h}{p} + \frac{p-u}{p}\right|^{2H-2} + \left|m + \frac{r+u+h}{p} - \frac{p-u}{p}\right|^{2H-2} - 2\left|m + \frac{r+u+h}{p}\right|^{2H-2}\right)$$

$$= -H(2H-1)\left(\frac{p-u}{p}\right)^{2H-2}$$

$$\times \left(\left|\frac{mp}{p-u} + \frac{r+u+h}{p-u} + 1\right|^{2H-2} + \left|\frac{mp}{p-u} + \frac{r+u+h}{p-u} - 1\right|^{2H-2} - 2\left|\frac{mp}{p-u} \frac{q+u+h}{p-u}\right|^{2H-2}\right)$$

$$\sim c_H \left(\frac{p-u}{p}\right)^{2H-2} \left|\frac{mp}{p-u} + \frac{r+u+h}{p-u}\right|^{2H-4}$$

$$= c_H \left(\frac{p-u}{p}\right)^2 \left|m + \frac{r+u+h}{p}\right|^{2H-4}$$

$$\sim c_H \left(\frac{p-u}{p}\right)^2 |m|^{2H-4},$$

$$(3.24)$$

where $c_H := (-4!/2)\binom{2H}{4}$. We can conclude that

$$G_H\left(\frac{q}{p}, \frac{u}{p}, p\right) = G_H\left(m + \left(\frac{r}{p}\right), \frac{u}{p}, p\right) \sim \frac{c_H}{2}\left(1 - \frac{u}{p}\right)|m|^{2H-4} \quad \text{as } |m| \longrightarrow +\infty. \qquad (3.25)$$

3.2. The MAVAR Log-Regression Estimator

Let n be the sample size, that is, the number of the observations.

Definition 3.2. Let $\overline{p}, \overline{\ell} \in \mathbb{N}$ such that $1 \le (1 + \overline{\ell})\overline{p} \le p_{\max}(n) = \lfloor n/3 \rfloor$, and let $\underline{w} = (w_0, \dots, w_{\overline{\ell}})$ be a vector of weights satisfying the conditions

$$\sum_{\ell=0}^{\overline{\ell}} w_\ell = 0, \qquad \sum_{\ell=0}^{\overline{\ell}} w_\ell \log(1 + \ell) = 1. \qquad (3.26)$$

The *MAVAR log-regression estimator* associated to the weights $\underline{w}$ is defined as

$$\widehat{\alpha}_n(\overline{p}, \underline{w}) = \widehat{\alpha}_n(\tau_0, \overline{p}, \underline{w}) := \sum_{\ell=0}^{\overline{\ell}} w_\ell \log\left(\widehat{\sigma}^2_{(1+\ell)\overline{p}}(n)\right). \qquad (3.27)$$

Roughly speaking, the idea behind this definition is to use the approximation

$$\left(\widehat{\sigma}^2_{\overline{p}}(n), \dots, \widehat{\sigma}^2_{(1+\overline{\ell})\overline{p}}(n)\right) \cong \left(\sigma^2_{\overline{p}}(n), \dots, \sigma^2_{(1+\overline{\ell})\overline{p}}(n)\right), \qquad (3.28)$$

in order to get, by (3.11) and (3.26),

$$\widehat{\alpha}_n(\overline{p}, \underline{w}) \cong \sum_{\ell=0}^{\overline{\ell}} w_\ell \log\left(\sigma^2_{(1+\ell)\overline{p}}\right) \cong \sum_{\ell=0}^{\overline{\ell}} w_\ell \left[\alpha \log(1 + \ell) + \alpha \log(\tau_0 \overline{p}) + \log\left(\sigma^2_H K(H)\right)\right] = \alpha, \qquad (3.29)$$

where $\alpha := 2H - 2$. Thus, given the data $X_1, \ldots, X_n$, the following procedure is used to estimate H:

(i) compute the modified Allan variance by (3.2), for integer values $(1 + \ell)\overline{p}$, with $1 \leq (1 + \ell)\overline{p} \leq p_{max}(n) = \lfloor n/3 \rfloor$;

(ii) compute the weighted MAVAR log-regression estimator by (3.27) in order to get an estimate $\hat{\alpha}$ of α;

(iii) estimate H by $\widehat{H} = (\hat{\alpha} + 2)/2$.

In the sequel we will give, under suitable assumptions, two convergence results in order to justify these approximations and to get the rate of convergence of $\hat{\alpha}_n(\overline{p}, \underline{w})$ toward $\alpha = 2H - 2$. Obviously, we need to take $\overline{p} = \overline{p}(n) \rightarrow +\infty$ as $n \rightarrow +\infty$ in order to reach jointly the above two approximations.

4. The Asymptotic Normality of the Estimator

Since now on we will always assume that X is a fractional Brownian motion (with $H \in (1/2, 1)$) so that the process $\{d_{p,k}\}_{p,k}$ is also *Gaussian*. Under this assumption, and with the notation introduced before, we can state the following results.

Theorem 4.1. *Let $\overline{p} = \overline{p}(n)$ be a sequence of integers such that*

$$\overline{p}(n) \longrightarrow +\infty, \qquad n\,\overline{p}(n)^{-1} \longrightarrow +\infty. \tag{4.1}$$

Let $\overline{\ell}$ be a given integer, $\widehat{\underline{\sigma}}_n^2(\overline{p}, \overline{\ell})$ the vector $(\widehat{\sigma}_{\overline{p}}^2(n), \widehat{\sigma}_{2\overline{p}}^2(n), \ldots, \widehat{\sigma}_{(1+\overline{\ell})\overline{p}}^2(n))$ and, analogously, set $\underline{\sigma}^2(\overline{p}, \overline{\ell}) = (\sigma_{\overline{p}}^2, \sigma_{2\overline{p}}^2, \ldots, \sigma_{(1+\overline{\ell})\overline{p}}^2)$. Then, it holds

$$\sqrt{\frac{n}{\overline{p}}}\, (\tau_0 \overline{p})^{2-2H} \left(\widehat{\underline{\sigma}}_n^2\left(\overline{p}, \overline{\ell}\right) - \underline{\sigma}^2\left(\overline{p}, \overline{\ell}\right) \right) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, W(H)), \tag{4.2}$$

where $W(H)$ is a suitable symmetric matrix.

From this theorem, as an application of the δ-method, we can state the following result.

Theorem 4.2. *Let $\hat{\alpha}_n(\overline{p}, \underline{w})$ be defined as in (3.27), for some finite integer $\overline{\ell}$ and a weight-vector $\underline{w}$ satisfying (3.26). If $\overline{p} = \overline{p}(n)$ is a sequence of integers such that*

$$\overline{p}(n) \longrightarrow +\infty, \qquad n\overline{p}(n)^{-1} \longrightarrow +\infty, \qquad n\,\overline{p}(n)^{-3} \longrightarrow 0, \tag{4.3}$$

then

$$\sqrt{\frac{n}{\overline{p}}}(\hat{\alpha}_n(\overline{p}, \underline{w}) - \alpha) \xrightarrow[n \to \infty]{d} \mathcal{N}\left(0, \underline{w}_*^T V(H) \underline{w}_*\right), \tag{4.4}$$

where $\alpha = 2H - 2$, the column-vector $\underline{w}_*$ is such that $[\underline{w}_*]_\ell := w_\ell(1 + \ell)^{2-2H}$, and $V(H) = (\sigma_H^2 K(H))^{-2} W(H)$.

Let us stress that from the above result, and due to the condition $n\bar{p}(n)^{-1} \to +\infty$, it follows that the estimator $\hat{\alpha}_n(\bar{p}, \underline{w})$ is consistent.

Before starting the proof of the above theorems, we need the following lemma.

Lemma 4.3. *Let* $\bar{p} = \bar{p}(n)$ *be a sequence of integers such that*

$$\bar{p}(n) \longrightarrow +\infty, \qquad n\bar{p}(n)^{-1} \longrightarrow +\infty. \tag{4.5}$$

For two given integers ℓ', ℓ, *with* $0 \le \ell' \le \ell$, *set* $p_{\ell'} = (\ell' + 1)\bar{p}$ *and* $p_\ell = (\ell + 1)\bar{p}$. *Then*

$$\frac{n}{\bar{p}}(\tau_0\bar{p})^{4-4H} \operatorname{Cov}\left(\hat{\sigma}_{p_\ell}^2(n), \hat{\sigma}_{p_{\ell'}}^2(n)\right) \xrightarrow[n \to +\infty]{} W_{\ell',\ell}(H), \tag{4.6}$$

where $W_{\ell',\ell}(H)$ *is a finite quantity.*

Proof. Since $n/\bar{p} \to +\infty$, without loss of generality we can assume that $(1 + \bar{\ell})\bar{p} \le p_{\max}(n) = \lfloor n/3 \rfloor$ for each n. Recall the notation $n_p = n - 3p + 1$, and set $n_\ell = n_{p_\ell} \sim n$ and $n_{\ell'} = n_{p_{\ell'}} \sim n$. From the definition of the MAVAR estimator and applying the Wick's rule for jointly Gaussian random variables (see the Appendix), we get

$$
\begin{aligned}
\operatorname{Cov}\left(\hat{\sigma}_{p_\ell}^2(n), \hat{\sigma}_{p_{\ell'}}^2(n)\right) &= \frac{1}{n_\ell n_{\ell'}} \sum_{k=0}^{n_\ell-1} \sum_{k'=0}^{n_{\ell'}-1} \operatorname{Cov}\left(d_{p_\ell,k}^2, d_{p_{\ell'},k'}^2\right) \\
&= \frac{2}{n_\ell n_{\ell'}} \sum_{k=0}^{n_\ell-1} \sum_{k'=0}^{n_{\ell'}-1} \operatorname{Cov}\left(d_{p_\ell,k}, d_{p_{\ell'},k'}\right)^2 \\
&= 2\sigma_H^4 (1 + \ell)^{4H-4} \frac{(\tau_0\bar{p})^{4H-4}}{n_\ell n_{\ell'}} \sum_{k=0}^{n_\ell-1} \sum_{k'=0}^{n_{\ell'}-1} G_H^2\left(k - \frac{k'}{p_\ell}, \frac{u_{\ell'\ell}}{p_\ell}, p_\ell\right),
\end{aligned}
\tag{4.7}
$$

where $u_{\ell'\ell} = p_\ell - p_{\ell'}$. Since, by (3.20), the function $G_H(q/p_\ell, u_{\ell'\ell}/p_\ell, p_\ell)$ only depends on q/p_ℓ and $u_{\ell'\ell}/p_\ell = (\ell - \ell')/(1 + \ell)$ as $n \to +\infty$, we rewrite the last line as

$$2\sigma_H^4 (1 + \ell)^{4H-4} \frac{(\tau_0\bar{p})^{4H-4}}{n_\ell n_{\ell'}} \sum_{k=0}^{n_\ell-1} \sum_{k'=0}^{n_{\ell'}-1} G_H^2\left(k - \frac{k'}{p_\ell}, \frac{(\ell - \ell')}{1 + \ell}\right). \tag{4.8}$$

This term, multiplied by $(n/\bar{p})(\tau_0\bar{p})^{4-4H}$, is equal to

$$2\sigma_H^4 (1 + \ell)^{4H-4} \frac{1}{\bar{p}} \frac{n}{n_\ell n_{\ell'}} \sum_{k=0}^{n_\ell-1} \sum_{k'=0}^{n_{\ell'}-1} G_H^2\left(k - \frac{k'}{p_\ell}, \frac{(\ell - \ell')}{(1 + \ell)}\right). \tag{4.9}$$

It is easy to see that this quantity converges (to a finite strictly positive limit) if and only if

$$\frac{1}{p_\ell} \sum_{m=0}^{[n/p_\ell]} \sum_{r=0}^{p_\ell-1} G_H^2\left(m + \left(\frac{r}{p_\ell}\right), \frac{(\ell-\ell')}{(1+\ell)}\right) \tag{4.10}$$

converges. From (3.25) it holds

$$G_H^2\left(m + \left(\frac{r}{p_\ell}\right), \frac{(\ell-\ell')}{(1+\ell)}\right) \sim \frac{c_H^2}{4}\left(1 - \frac{\ell-\ell'}{1+\ell}\right)^2 m^{4H-8} \quad \text{as } m \longrightarrow +\infty. \tag{4.11}$$

Thus, the quantity (4.10) is controlled in the limit $n \to \infty$ by the sum $\sum_{m=1}^{\infty} m^{4H-8}$ that is convergent. □

Proof of Theorem 4.1. As before, without loss of generality, we can assume that $(1 + \overline{\ell})\overline{p} \le p_{\max}(n)$ for each n. Moreover, set again $n_\ell = n_{p_\ell}$ and $n_{\ell'} = n_{p_{\ell'}}$. For a given real vector $\underline{v}^T = (v_0, \ldots, v_{\overline{\ell}})$, let us consider the random variable $T_n = T(\overline{p}(n), \overline{\ell}, \underline{v})$ defined as a linear combination of the empirical variances $\widehat{\sigma}_{\overline{p}}^2(n), \ldots, \widehat{\sigma}_{(1+\overline{\ell})\overline{p}}^2(n)$ as follows:

$$T_n := \sum_{\ell=0}^{\overline{\ell}} v_\ell \, \widehat{\sigma}_{(1+\ell)\overline{p}}^2(n) = \sum_{\ell=0}^{\overline{\ell}} \frac{v_\ell}{n_\ell} \sum_{k=0}^{n_\ell-1} d_{(1+\ell)\overline{p},k}^2. \tag{4.12}$$

In order to prove the convergence stated in Theorem 4.1, we have to show that the random variable $\sqrt{n/\overline{p}} \, (\tau_0\overline{p})^{2-2H} \, (T_n - \sum_{\ell=0}^{\overline{\ell}} v_\ell \sigma_{(1+\ell)\overline{p}}^2)$ converges to the normal distribution with zero mean and variance $\underline{v}^T W(H)\underline{v}$. To this purpose, we note that

$$\sqrt{\frac{n}{\overline{p}}}(\tau_0\overline{p})^{2-2H}\left(T_n - \sum_{\ell=0}^{\overline{\ell}} v_\ell\sigma_{(1+\ell)\overline{p}}^2\right) = V_n^T A_n V_n - \mathbb{E}\left[V_n^T A_n V_n\right], \tag{4.13}$$

where V_n is the random vector with entries $d_{(1+\ell)\overline{p},k}$, for $0 \le \ell \le \overline{\ell}$ and $0 \le k \le n_\ell - 1$, and A_n is the diagonal matrix with entries

$$[A_n]_{((1+\ell)\overline{p},k),((1+\ell)\overline{p},k)} = \frac{v_\ell}{n_\ell}\sqrt{\frac{n}{\overline{p}}}(\tau_0\overline{p})^{2-2H} = O\left((\overline{p})^{3/2-2H}n^{-1/2}\right). \tag{4.14}$$

By Lemma 4.3, if $W(H)$ is the symmetric matrix with $[W(H)]_{\ell',\ell} = W_{\ell',\ell}$ when $0 \le \ell' \le \ell \le \overline{\ell}$, it holds

$$\text{Var}\left[V_n^T A_n V_n\right] = \frac{n}{\overline{p}}(\tau_0\overline{p})^{4-4H}\sum_{\ell=0}^{\overline{\ell}}\sum_{\ell'=0}^{\overline{\ell}} v_\ell v_{\ell'}\text{Cov}\left(\widehat{\sigma}_{(1+\ell)\overline{p}}^2(n), \widehat{\sigma}_{\overline{p}(1+\ell')}^2(n)\right)$$

$$\xrightarrow[n\to+\infty]{} \sum_{\ell=0}^{\overline{\ell}}\sum_{\ell'=0}^{\overline{\ell}} v_\ell v_{\ell'} W_{\ell',\ell}(H) = \underline{v}^T W(H)\underline{v}, \tag{4.15}$$

therefore, condition (1) of Lemma A.2 is satisfied.

In order to verify condition (2) of **Lemma A.2**, let C_n denote the covariance matrix of the random vector V_n, and let $\rho[C_n]$ denote its spectral radius. By **Lemma A.3**, we have

$$\rho[C_n] \leq \sum_{\ell=0}^{\overline{\ell}} \rho\big[C_n((1+\ell)\overline{p})\big], \tag{4.16}$$

where $C_n((1+\ell)\overline{p})$ is the covariance matrix of the subvector $(d_{(1+\ell)\overline{p},k})_{k=0,\ldots,n_\ell-1}$. Applying the spectral radius estimate (A.3), and from equality (3.19), we then have

$$\rho\big[C_n((1+\ell)\overline{p})\big] \leq \sigma_H^2 \big[\tau_0(1+\ell)\overline{p}\big]^{2H-2} \left[G_H(0,0,p_\ell) + 2\sum_{q=1}^{n_\ell-1} G_H\left(\frac{q}{p_\ell},0,p_\ell\right) \right]$$
$$= O\big((\overline{p})^{2H-1}\big). \tag{4.17}$$

In order to conclude, it is enough to note that

$$\rho[A_n]\rho[C_n] = O\big((\overline{p})^{3/2-2H} n^{-1/2}\big) O\big((\overline{p})^{2H-1}\big) = O\big((\overline{p})^{1/2} n^{-1/2}\big) \xrightarrow[n\to\infty]{} 0. \tag{4.18}$$

$$\square$$

Proof of Theorem 4.2. By assumptions (4.3) on the sequence $\overline{p} = \overline{p}(n)$, and in particular from the condition $n/(\overline{p})^3 \to 0$, and inequality (3.11), it holds

$$\sqrt{\frac{n}{\overline{p}}}(\tau_0\overline{p})^{2-2H}\Big|\sigma_{(1+\ell)\overline{p}}^2 - \sigma_H^2\big[\tau_0(1+\ell)\overline{p}\big]^{2H-2} K(H)\Big|$$
$$\leq \sigma_H^2(1+\ell)^{2H-2} O\big(n^{1/2}(\overline{p})^{-3/2}\big) \xrightarrow[n\to+\infty]{} 0. \tag{4.19}$$

Thus, from **Theorem 4.1**, we get

$$\sqrt{\frac{n}{\overline{p}}}\Big[(\tau_0\overline{p})^{2-2H}\widehat{\underline{\sigma}}_n^2\big(\overline{p},\overline{\ell}\big) - \underline{\sigma}_*^2\Big] \xrightarrow[n\to\infty]{d} \mathcal{N}(0,W(H)), \tag{4.20}$$

where $\underline{\sigma}_*^2$ is the vector with elements

$$\big[\underline{\sigma}_*^2\big]_\ell := \sigma_H^2(1+\ell)^{2H-2} K(H) \quad \text{for } 0 \leq \ell \leq \overline{\ell}. \tag{4.21}$$

Now we observe that if $f(\underline{x}) := \sum_{\ell=0}^{\overline{\ell}} w_\ell \log(x_\ell)$, then, by (3.26) and (3.27), we have

$$\widehat{\alpha}_n(\overline{p},\underline{w}) = f\big(\widehat{\underline{\sigma}}^2(\overline{p},\overline{\ell})\big) = f\big((\tau_0\overline{p})^{2-2H}\widehat{\underline{\sigma}}^2(\overline{p},\overline{\ell})\big). \tag{4.22}$$

Moreover, $\alpha = f(\underline{\sigma}_*^2)$ and $\nabla f(\underline{\sigma}_*^2) = (\sigma_H^2 K(H))^{-1} \underline{w}_*$. Therefore, by the application of the δ-method, the convergence (4.20) entails

$$\sqrt{\frac{n}{\overline{p}}} \left(\widehat{\alpha}_n(\overline{p}, \underline{w}) - \alpha \right) \xrightarrow[n \to \infty]{d} \mathcal{N}\left(0, \nabla f\left(\underline{\sigma}_*^2\right)^T W(H) \nabla f\left(\underline{\sigma}_*^2\right) \right), \tag{4.23}$$

where

$$\nabla f\left(\underline{\sigma}_*^2\right)^T W(H) \nabla f\left(\underline{\sigma}_*^2\right) = \underline{w}_*^T V(H) \underline{w}_*, \tag{4.24}$$

and thus concludes the proof.   □

Remark 4.4. By Lemma 4.3, and since by definition $V_{\ell',\ell}(H) = (\sigma_H^2 K(H))^{-2} W_{\ell',\ell}(H)$, an estimate of $V_{\ell',\ell}(H)$, with $0 \le \ell' \le \ell \le \overline{\ell}$, is given by

$$\frac{n}{\overline{p}} \frac{(\tau_0 \overline{p})^{4-4H}}{\sigma_H^4 K(H)^2} \mathrm{Cov}\left(\widehat{\sigma}_{p_\ell}^2(n), \widehat{\sigma}_{p_{\ell'}}^2(n) \right). \tag{4.25}$$

Therefore, setting $\rho_n^2(\underline{w}, H)$ equal to the corresponding estimate of $\underline{w}_*^T V(H) \underline{w}_*$ divided by $n/\overline{p}$, that is

$$\rho_n^2(\underline{w}, H) = \frac{2}{K(H)^2} \sum_{\ell=0}^{\overline{\ell}} \sum_{\ell'=0}^{\overline{\ell}} \left(\frac{1 + \ell \wedge \ell'}{1 + \ell \vee \ell'} \right)^{2-2H} \frac{w_\ell w_{\ell'}}{n_\ell n_{\ell'}} \\ \times \sum_{k=0}^{n_{\ell \vee \ell'}-1} \sum_{k'=0}^{n_{\ell \wedge \ell'}-1} G_H^2\left(\frac{k-k'}{\overline{p}(1+\ell \vee \ell')}, \frac{|\ell - \ell'|}{1+\ell \vee \ell'}, \overline{p}(1+\ell \vee \ell') \right), \tag{4.26}$$

and from (4.6), we obtain the convergence

$$\rho_n^{-1}(\underline{w}, H) \left(\widehat{\alpha}_n(\overline{p}, \underline{w}) - \alpha \right) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, 1), \tag{4.27}$$

which can be used to obtain an asymptotic confidence interval for the parameter H.

4.1. An Alternative Representation of the Covariance

It is well known that an FBM (with $X(0) = 0$ and $\mathbb{E}[X^2(1)] = \sigma_H^2$) has the following stochastic integral representation (see [2, 16]):

$$X(t) = \frac{1}{\Gamma(H + 1/2)} \int_{\mathbb{R}} \left\{ [(t-s)^+]^{H-1/2} - (s^-)^{H-1/2} \right\} dW_s. \tag{4.28}$$

Using this representation and Ito's formula, we can obtain another representation of the covariance $\text{Cov}(\hat{\sigma}_p^2(n), \hat{\sigma}_{p-u}^2(n))$ and, consequently, another formula for the coefficient $\rho_n(\underline{w}, H)$.

Indeed, assuming without loss of generality $t_1 = \tau_0$ so that $X_i = X(t_i) = X(i\tau_0)$, by (3.3) and (4.28), we can write for $0 \le u < p$ and $k \in \mathbb{Z}$

$$d_{p-u,k} = \frac{\sigma_H}{\sqrt{2}} (\tau_0 p)^{H-3/2} \frac{p^2}{(p-u)^2} \int_{\mathbb{R}} \gamma_H\left(\frac{k}{p}, p, u, \frac{s}{(\tau_0 p)}\right) dW_s, \tag{4.29}$$

where

$$
\begin{aligned}
&\gamma_H\left(\frac{k}{p}, p, u, \frac{s}{(\tau_0 p)}\right) \\
&:= \left[\Gamma\left(H + \frac{1}{2}\right)\sigma_H\right]^{-1} \\
&\quad \times \frac{1}{p}\sum_{j=1}^{p-u}\left\{\left[\left(\frac{k+j}{p} + \frac{2(p-u)}{p} - \frac{s}{\tau_0 p}\right)^+\right]^{H-1/2}\right. \\
&\quad \left. -2\left[\left(\frac{k+j}{p} + \frac{p-u}{p} - \frac{s}{\tau_0 p}\right)^+\right]^{H-1/2} + \left[\left(\frac{k+j}{p} - \frac{s}{\tau_0 p}\right)^+\right]^{H-1/2}\right\}.
\end{aligned}
\tag{4.30}
$$

By Ito's formula we get

$$d_{p-u,k}^2 = \mathbb{E}\left[d_{p-u,k}^2\right] + \sigma_H^2(\tau_0 p)^{2H-3}\frac{p^4}{(p-u)^4}\int_{\mathbb{R}} Z\left(\frac{k}{p}, p, u, v\right)\gamma_H\left(\frac{k}{p}, p, u, \frac{v}{(\tau_0 p)}\right) dW_v, \tag{4.31}$$

where

$$Z\left(\frac{k}{p}, p, u, v\right) := \int_{-\infty}^{v} \gamma_H\left(\frac{k}{p}, p, u, \frac{s}{(\tau_0 p)}\right) dW_s. \tag{4.32}$$

Inserting this expression in (3.4), we obtain

$$\hat{\sigma}_{p-u}^2(n) - \sigma_{p-u}^2 = \sigma_H^2(\tau_0 p)^{2H-3}\frac{p^4}{(p-u)^4}\int_{\mathbb{R}}\left[\frac{1}{n_{p-u}}\sum_{k=0}^{n_{p-u}-1} Z\left(\frac{k}{p}, p, u, v\right)\gamma_H\left(\frac{k}{p}, p, u, \frac{v}{\tau_0 p}\right)\right] dW_v. \tag{4.33}$$

It follows, using Ito's formula again and after the change of variables $s = \tau_0(pv + k')$ and $v = \tau_0(pr + k')$,

$$
\begin{aligned}
\operatorname{Cov}\left(\widehat{\sigma}_p^2(n), \widehat{\sigma}_{p-u}^2(n)\right) = {}& \sigma_H^4 (\tau_0 p)^{4H-4} \frac{p^4}{(p-u)^4} \frac{1}{n_p n_{p-u}} \\
& \times \sum_{k=0}^{n_p-1} \sum_{k'=0}^{n_{p-u}-1} \int_{\mathbb{R}} \left[\gamma_H\left(\frac{(k-k')}{p}, p, 0, r\right) \gamma_H(0, p, u, r) \right. \\
& \left. \times \int_{-\infty}^{r} \gamma_H\left(\frac{(k-k')}{p}, p, 0, v\right) \gamma_H(0, p, u, v) dv \right] dr.
\end{aligned}
\tag{4.34}
$$

5. Some Comments

5.1. The Modified Allan Variance and the Wavelet Estimators

Suppose that X is a self-similar process with weakly stationary increments and consider the generalized process $Y = \{Y(t), t \in \mathbb{R}\}$ defined through the set of identities

$$
\langle Y; I_{[t_1, t_2]} \rangle = \int_{t_1}^{t_2} Y(t) dt = X(t_2) - X(t_1), \quad \forall t_1, t_2 \in \mathbb{R}.
\tag{5.1}
$$

In short, we write $Y = \dot{X}$. From this definition and with the notation introduced in Section 3, we can rewrite the MAVAR as

$$
\sigma^2(\tau_0, p) = \frac{1}{2\tau_0^2 p^2} \mathbb{E}\left[\left(\frac{1}{p} \sum_{i=1}^{p} \left(\int_{t_{i+p}}^{t_{i+2p}} Y(t) dt - \int_{t_i}^{t_{i+p}} Y(t) dt \right) \right)^2 \right]
\tag{5.2}
$$

and its related estimator as

$$
\widehat{\sigma}^2(\tau_0, p, n) = \frac{1}{2\tau_0^2 p^2 n_p} \sum_{k=0}^{n_p-1} \left[\frac{1}{p} \sum_{i=1}^{p} \left(\int_{t_{i+k+p}}^{t_{i+k+2p}} Y(t) dt - \int_{t_{i+k}}^{t_{i+k+p}} Y(t) dt \right) \right]^2.
\tag{5.3}
$$

Now we claim that, for p fixed, the quantity

$$
d(\tau_0, p, k) := \frac{1}{\sqrt{2} p^2 \tau_0} \sum_{i=1}^{p} \left(\int_{t_{i+k+p}}^{t_{i+k+2p}} Y(t) dt - \int_{t_{i+k}}^{t_{i+k+p}} Y(t) dt \right)
\tag{5.4}
$$

recalls a family of discrete wavelet transforms of the process Y, indexed by τ_0 and k. To see that, let us fix $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ and set $\tau_0 = 2^j$ and $t_1 = 2^j$, so that $t_{i+k} = 2^j(i+k)$, for all $i \in \mathbb{N}$.

With this choice on the sequence of times, it is not difficult to construct a function $\psi(s)$ such that

$$d_{k,j} := d\left(2^j, p, k\right) = \langle Y; \psi_{k,j} \rangle \quad \text{with } \psi_{k,j}(s) := 2^{-j} \psi\left(2^{-j} s - k\right). \tag{5.5}$$

An easy check shows that the function

$$\psi(s) := \frac{1}{p\sqrt{p}} \sum_{i=1}^{p} \psi^i(s), \qquad \psi^i(s) := \frac{1}{\sqrt{2p}} \left(I_{[i+p,i+2p]}(s) - I_{[i,i+p]}(s)\right), \tag{5.6}$$

satisfies (5.5). Notice also that the components ψ^i, $i = 1, \ldots p$, are suitably translated and renormalized Haar functions.

In the case $p = 1$, corresponding to the classical Allan variance, the function ψ is exactly given by the Haar mother wavelet, as was already pointed out in [14].

Though the wavelet representation could be convenient in many respects, the Haar mother wavelet does not satisfy one of the conditions which are usually required in order to study the convergence of the estimator (see condition (W2) in [18]). Moreover, there is a fundamental difference between the two methods: in the wavelet setting the log-regression is done over the scale parameter τ_0 for p fixed, while the MAVAR log-regression given in (3.27) is performed over p and for τ_0 fixed.

5.2. The Modified Hadamard Variance

Further generalizing the notion of the MAVAR, one can define the modified Hadamard variance (MHVAR). For fixed integers p, Q, and $\tau_0 \in \mathbb{R}$, set

$$\sigma^2(\tau_0, p, Q) := \frac{1}{Q! \tau_0^2 p^2} \mathbb{E}\left[\left(\frac{1}{p} \sum_{i=1}^{p} \sum_{q=0}^{Q} c(Q, q) X(t_{i+qp})\right)^2\right], \tag{5.7}$$

where $c(Q, q) := (-1)^q (Q!/q!(Q-q)!)$. Notice that for $Q = 2$ we recover the modified Allan variance. The MHVAR is again a time-domain quantity which has been introduced in [7] for the analysis of the network traffic. A standard estimator for this variance is given by

$$
\begin{aligned}
\widehat{\sigma}_n^2(\tau_0, p, Q) &:= \frac{1}{Q! \tau_0^2 p^4 n_{Q,p}} \sum_{h=1}^{n_{Q,p}} \left(\sum_{i'=h}^{p+h-1} \sum_{q=0}^{Q} c(Q, q) X_{i'+qp}\right)^2 \\
&= \frac{1}{n_{Q,p}} \sum_{k=0}^{n_{Q,p}-1} \left(\frac{1}{\sqrt{Q!} \tau_0 p^2} \sum_{i=1}^{p} \sum_{q=0}^{Q} c(Q, q) X_{k+i+qp}\right)^2,
\end{aligned}
\tag{5.8}
$$

for $p = 1, \ldots, [n/(Q+1)]$ and $n_{Q,p} := n - (Q+1)p + 1$.

Similarly to the analysis performed for the MAVAR, let us set

$$d_{p,k} = d(\tau_0, p, Q, k) := \frac{1}{\sqrt{Q!}\tau_0 p^2} \sum_{i=1}^{p} \sum_{q=0}^{Q} c(Q,q) X_{k+i+qp}, \tag{5.9}$$

so that we can write

$$\sigma^2(\tau_0, p, Q) = \mathbb{E}\left[d_{p,0}^2\right], \qquad \widehat{\sigma}_n^2(\tau_0, p, Q) = \frac{1}{n_{Q,p}} \sum_{k=0}^{n_{Q,p}-1} d_{p,k}^2. \tag{5.10}$$

This suggests that convergence results, similar to Theorems 4.1 and 4.2, can be achieved also for the MHVAR and its related log-regression estimator.

5.3. The Case of Stationary Processes

In applications, MAVAR and MHVAR are also used in order to estimate the memory parameter of long-range dependent processes. This general case is not included in our analysis (which is restricted to the fractional Brownian motion) and it requires a more involved investigation. To our knowledge, there are no theoretical results along this direction.

Appendix

In this appendix we recall the Wick's rule for jointly Gaussian random variables and some facts used along the proofs.

Wick's Rule

Let us consider a family $\{Z_i\}$ of jointly Gaussian random variables with zero mean. The Wick's rule is a formula that provides an easy way to compute the quantity $\mathbb{E}[Z_\Lambda] := \mathbb{E}[\prod_{i\in\Lambda} Z_i]$, for any index-set Λ (see, e.g., [19]).

Since the Z_i's are zero-mean random variables, if Λ has odd cardinality, we trivially get $\mathbb{E}[Z_\Lambda] = 0$. We then assume that $|\Lambda| = 2k$, for some $k \geq 1$. To recall the Wick's rule, it is convenient to introduce the following graph representation. To the given index-set Λ we associated a vertex-set V indexed by the distinct elements of Λ, and to every vertex $j \in V$ we attached as many half-edges as many times the index j appears in Λ. In particular there is a biunivocal correspondence between the set of half-edges and Λ, while $|V| \leq |\Lambda|$. Gluing together two half-edges attached to vertices i and j, we obtain the edge (i, j). Performing this operation recursively over all remaining half-edges, we end up with a graph G, with vertex set $V(G)$ and edge-set $E(G)$. Let $\mathcal{G}_\Lambda$ denote the set of graphs (possibly not distinguishable) obtained by performing this "gluing procedure" in all possible ways.

With this notation, and for all index-sets Λ with even cardinality, the *Wick's rule* for a family $\{Z_i\}$ of jointly centered Gaussian random variables provides the identity

$$\mathbb{E}[Z_\Lambda] = \sum_{G \in \mathcal{G}_\Lambda} \prod_{(i,j) \in E(G)} \mathbb{E}[Z_i Z_j]. \tag{A.1}$$

Example A.1. Consider the quantity $\mathbb{E}[Z_k^2 Z_j^2]$, for $j \neq k$. By the graphical representation, we take a vertex set $V = \{j, k\}$ with two half-edges attached to each vertex and perform the gluing operation on the half-edges. We then obtain the following three graphs (identified by their edges): $G_1 = (j, k)(j, k)$, $G_2 = (j, k)(j, k)$, and $G_3 = (j, j)(k, k)$. Thus, from the Wick's rule, we get the identity

$$\mathbb{E}\left[Z_k^2 Z_j^2\right] = 2\mathbb{E}[Z_j Z_k]^2 + \mathbb{E}\left[Z_j^2\right]\mathbb{E}\left[Z_k^2\right], \tag{A.2}$$

that has been used throughout the paper.

Now we recall some facts used in the proof of Theorem 4.1.
Denote by $\rho[A]$ the spectral radius of a matrix $A = \{a_{i,j}\}_{1 \leq i,j \leq n}$, then

$$\rho[A] \leq \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{i,j}|. \tag{A.3}$$

Moreover the following lemmas hold.

Lemma A.2 (see [13]). *Let (V_n) be a sequence of centered Gaussian random vectors and denote by C_n the covariance matrix of V_n. Let (A_n) be a sequence of deterministic symmetric matrices such that*

(1) $\lim_{n \to +\infty} \mathrm{Var}[V_n^T A_n V_n] = \lambda^2 \in [0, +\infty)$,

(2) $\lim_{n \to +\infty} \rho[A_n]\rho[C_n] = 0$.

Then $V_n^T A_n V_n - \mathbb{E}[V_n^T A_n V_n]$ converges in distribution to the normal law $\mathcal{N}(0, \lambda^2)$.

Lemma A.3 (see [11]). *Let $m \geq 2$ be an integer and C a $m \times m$ covariance matrix. Let r be an integer such that $1 \leq r \leq m - 1$. Denote by C_1 the top left submatrix with size $r \times r$ and by C_2 the bottom right submatrix with size $(m - r) \times (m - r)$, that is,*

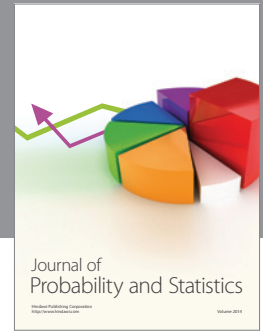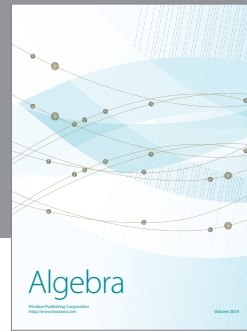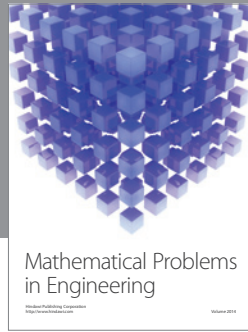$$C_1 = [C_{i,j}]_{1 \leq i,j \leq r}, \qquad C_2 = [C_{i,j}]_{r+1 \leq i,j \leq m}. \tag{A.4}$$

Then $\rho[C] \leq \rho[C_1] + \rho[C_2]$.

Acknowledgments

References

[1] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, London, UK, 1994.

[2] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.

[3] L. G. Bernier, "Theoretical analysis of the modified Allan variance," in *Proceedings of the 41st Annual Frequency Control Symposium*, pp. 161–121, 1987.

[4] S. Bergni, "Characterization and modelling of clocks," in *Synchronization of Digital Telecommu-Nications Networks*, John Wiley & Sons, 2002.

[5] S. Bregni and L. Primerano, "The modified Allan variance as time-domain analysis tool for estimating the hurst parameter of long-rangé dependent traffic," in *IEEE Global Telecommunications Conference (GLOBECOM '04)*, pp. 1406–1410, December 2004.

[6] S. Bregni and W. Erangoli, "Fractional noise in experimental measurements of IP traffic in a metropolitan area network," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '05)*, pp. 781–785, December 2005.

[7] S. Bregni and L. Jmoda, "Accurate estimation of the Hurst parameter of long-range dependent traffic using modified Allan and Hadamard variances," *IEEE Transactions on Communications*, vol. 56, no. 11, pp. 1900–1906, 2008.

[8] A. Bianchi, S. Bregni, I. Crimaldi, and M. FERRARI, "Analysis of a hurst parameter estimator based on the modified Allan variance," in *Proceedings of the IEEE Global Telecommunications Conference and Exhibition (GLOBECOM) and IEEE Xplore*, 2012.

[9] J. F. Coeurjolly, "Simulation and identification of the fractional brownian motion: a bibliographical and comparative study," *Journal of Statistical Software*, vol. 5, pp. 1–53, 2000.

[10] J.-F. Coeurjolly, "Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths," *Statistical Inference for Stochastic Processes*, vol. 4, no. 2, pp. 199–227, 2001.

[11] E. Moulines, F. Roueff, and M. S. Taqqu, "Central limit theorem for the log-regression wavelet estimation of the memory parameter in the gaussian semi-parametric context," *Fractals*, vol. 15, no. 4, pp. 301–313, 2007.

[12] E. Moulines, F. Roueff, and M. S. Taqqu, "On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter," *Journal of Time Series Analysis*, vol. 28, no. 2, pp. 155–187, 2007.

[13] E. Moulines, F. Roueff, and M. S. Taqqu, "A wavelet whittle estimator of the memory parameter of a nonstationary Gaussian time series," *Annals of Statistics*, vol. 36, no. 4, pp. 1925–1956, 2008.

[14] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic patrice abry and darryl veitch," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[15] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds., pp. 39–88, Wiley, New York, NY, USA, 2000.

[16] T. Lindstrom, "A weighted random walk approximation to fractional Brownian motion," Tech. Rep. 11, Department of Mathematics, University of Oslo, 2007, http://arxiv.org/abs/0708.1905.

[17] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*, Springer Series in Statistics, Springer, New York, NY, USA, 1987.

[18] F. Roueff and M. S. Taqqu, "Asymptotic normality of wavelet estimators of the memory parameter for linear processes," *Journal of Time Series Analysis*, vol. 30, no. 5, pp. 534–558, 2009.

[19] J. Glimm and A. Jaffe, *Quantum Physics. A Functional Integral Point of View*, Springer, New York, NY, USA, 2nd edition, 1987.

Advances in
Operations Research

Advances in
Decision Sciences

Mathematical Problems
in Engineering

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International Journal of
Combinatorics

Hindawi

Submit your manuscripts at
http://www.hindawi.com

Advances in
Mathematical Physics

Journal of
Complex Analysis

Journal of
Mathematics

International Journal of
Stochastic Analysis

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

Journal of
Applied Mathematics

Journal of
Optimization